



applied sciences

Future Transportation

Edited by

Giovanni Randazzo, Anselme Muzirafuti,
Dimitrios S. Paraforos and Stefania Lanza

Printed Edition of the Special Issue Published in *Applied Sciences*

Future Transportation

Future Transportation

Editors

Giovanni Randazzo

Anselme Muzirafuti

Dimitrios S. Paraforos

Stefania Lanza

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editors

Giovanni Randazzo
University of Messina
Messina
Italy

Anselme Muzirafuti
University of Messina
Messina
Italy

Dimitrios S. Paraforos
Hochschule Geisenheim
University
Geisenheim
Germany

Stefania Lanza
GeoloGIS s.r.l.
Messina
Italy

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Applied Sciences* (ISSN 2076-3417) (available at: https://www.mdpi.com/journal/applsci/special_issues/Future).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , <i>Volume Number</i> , Page Range.
--

ISBN 978-3-0365-4857-9 (Hbk)

ISBN 978-3-0365-4858-6 (PDF)

Cover image courtesy of Anselme Muzirafuti

© 2022 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editors	vii
Preface to "Future Transportation"	ix
Julieth Stefany García, José D. Morcillo, Johan Manuel Redondo and Mauricio Becerra Fernández Automobile Technological Transition Scenarios Based on Environmental Drivers Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 4593, doi:10.3390/app12094593	1
Efstathios Bouhouras, Socrates Basbas, Stamatia Ftergioti, Evangelos Paschalidis and Harris Siakantaris COVID-19's Pandemic Effects on Bike Sharing Systems: A New Reality for Urban Mobility? Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 1230, doi:10.3390/app12031230	17
Younes Mamouch, Ahmed Attou, Abdelhalim Miftah, Mohammed Ouchchen, Bouchra Dadi, Lahsen Achkouch, Yassine Et-tayea, Abdelhamid Allaoui, Mustapha Boualoul, Giovanni Randazzo, Stefania Lanza and Anselme Muzirafuti Mapping of Hydrothermal Alteration Zones in the Kelâat M'Gouna Region Using Airborne Gamma-Ray Spectrometry and Remote Sensing Data: Mining Implications (Eastern Anti-Atlas, Morocco) Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 957, doi:10.3390/app12030957	39
Nils Olsen and Natalia Kliewer Location Planning of Charging Stations for Electric Buses in Public Transport Considering Vehicle Scheduling: A Variable Neighborhood Search Based Approach Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 3855, doi:10.3390/app12083855	59
Pedro-Luis Sanchez-Gonzalez, David Díaz-Gutiérrez and Luis R. Núñez-Rivas Digitalizing Maritime Containers Shipping Companies: Impacts on Their Processes Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 2532, doi:10.3390/app12052532	85
Carmen Kettwich, Andreas Schrank, Hüseyin Avsar and Michael Oehl A Helping Human Hand: Relevant Scenarios for the Remote Operation of Highly Automated Vehicles in Public Transport Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 4350, doi:10.3390/app12094350	125
René Cabral, Francisco García-Flores and Eduardo Saucedo The Influence of Sentiments of Economic Agents on Pedestrians and Vehicle Crossings along the US–Mexico Border Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 2512, doi:10.3390/app12052512	151
Alberto Dianin, Michael Gidam and Georg Hauger Isolating the Role of the Transport System in Individual Accessibility Differences: A Space-Time Transport Performance Measure Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 3309, doi:10.3390/app12073309	169
Emad A. Alharbi, Layek L. Abdel-Malek, R. John Milne and Arwa M. Wali Analytical Model for Enhancing the Adoptability of Continuous Descent Approach at Airports Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 1506, doi:10.3390/app12031506	189
Matthias Kowald, Margarita Gutjar, Kai Röth, Christian Schiller and Till Dannewald Mode Choice Effects on Bike Sharing Systems Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 4391, doi:10.3390/app12094391	213

Sheeba Razzaq, Amil Roohani Dar, Munam Ali Shah, Hasan Ali Khattak, Ejaz Ahmed, Ahmed M. El-Sherbeeney, Seongkwan Mark Lee, Khaled Alkhaledi and Hafiz Tayyab Rauf Multi-Factor Rear-End Collision Avoidance in Connected Autonomous Vehicles Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 1049, doi:10.3390/app12031049	229
Miziya Keshkar, Raja Muthalagu and Abdul Rajak ROAD Statistics-Based Noise Detection for DME Mitigation in LDACS Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 3774, doi:10.3390/app12083774	247
Salman Rashid, Shukor Abd Razak and Fuad A. Ghaleb IMU: A Content Replacement Policy for CCN, Based on Immature Content Selection Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 344, doi:10.3390/app12010344	277
Yanqun Yang, Danni Yin, Said M. Easa and Jiang Liu Attitudes toward Applying Facial Recognition Technology for Red-Light Running by E-Bikers: A Case Study in Fuzhou, China Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 211, doi:10.3390/app12010211	299
Lijana Maskeliūnaitė and Henrikas Sivilevičius Identifying the Importance of Criteria for Passenger Choice of Sustainable Travel by Train Using ARTIW and IHAMCI Methods Reprinted from: <i>Appl. Sci.</i> 2021 , <i>11</i> , 11503, doi:10.3390/app112311503	317
Klaudiusz Migawa, Sylwester Borowski, Andrzej Neubauer and Agnieszka Sołtysiak Semi-Markov Model of the System of Repairs and Preventive Replacements by Age of City Buses Reprinted from: <i>Appl. Sci.</i> 2021 , <i>11</i> , 10411, doi:10.3390/app112110411	339
Chan-Juan Liu, Jinran Wu, Harshanie Lakshika Jayetileke and Zhi-Hua Hu Long-Range Dependence and Multifractality of Ship Flow Sequences in Container Ports: A Comparison of Shanghai, Singapore, and Rotterdam Reprinted from: <i>Appl. Sci.</i> 2021 , <i>11</i> , 10378, doi:10.3390/app112110378	355
Hasan A. H. Naji, Qingji Xue, Huijun Zhu and Tianfeng Li Forecasting Taxi Demands Using Generative Adversarial Networks with Multi-Source Data Reprinted from: <i>Appl. Sci.</i> 2021 , <i>11</i> , 9675, doi:10.3390/app11209675	371
Mohammed Al-Sarem, Faisal Saeed, Zeyad Ghaleb Al-Mekhlafi, Badiya Abdulkarem Mohammed, Mohammed Hadwan, Tawfik Al-Hadhrani, Mohammad T. Alshammari, Abdulrahman Alreshidi and Talal Sarheed Alshammari An Improved Multiple Features and Machine Learning-Based Approach for Detecting Clickbait News on Social Networks Reprinted from: <i>Appl. Sci.</i> 2021 , <i>11</i> , 9487, doi:10.3390/app11209487	397
Armin Razmjoo, Amirhossein Gandomi, Maral Mahlooji, Davide Astiaso Garcia, Seyedali Mirjalili, Alireza Rezvani, Sahar Ahmadzadeh and Saim Memon An Investigation of the Policies and Crucial Sectors of Smart Cities Based on IoT Application Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 2672, doi:10.3390/app12052672	413
Izhari Izmi Mazali, Zul Hilmi Che Daud, Mohd Kameil Abdul Hamid, Victor Tan, Pakharuddin Mohd Samin, Abdullah Jubair, Khairul Amilin Ibrahim, Mohd Salman Che Kob, Wang Xinrui and Mat Hussin Ab Talib Review of the Methods to Optimize Power Flow in Electric Vehicle Powertrains for Efficiency and Driving Performance Reprinted from: <i>Appl. Sci.</i> 2022 , <i>12</i> , 1735, doi:10.3390/app12031735	427

About the Editors

Giovanni Randazzo

Prof. Dr. Giovanni Randazzo is an associate professor of coastal geomorphology and environmental geology at the University of Messina (Department of Mathematics, Computer Sciences, Physics and Earth Sciences, Via F. Stagno d'Alcontres, 31–98166 Messina, Italy). He holds a PhD in marine environments and resources obtained from the University of Messina (Italy). Since 1987, his research interest has focused on the study of coastal areas, their management and protection. In these last 30 years, he collaborated with the Smithsonian Institution of Washington D.C. in the study of the Nile delta; with the Thai Geological Service in the study of the east coast of the local peninsula; and with ENEA (the Italian National Agency for New Technologies, Energy and Sustainable Economic Development), he participated in the X Italian expedition in Antarctica. He collaborated in the environmental assessment impact of various public works (especially in coastal areas), as well as in the drafting of the Territorial Landscape Plan of the province of Messina (Sicily, Italy). In recent years, he actively participated in the debate on the emergence of waste, writing scientific articles, intervening in local press, and participating in various debates, where he presented a scheme of the management of emergency alternatives to those not acting for the Sicilian region. In 2013, he founded Geologis s.r.l., a branch of the University of Messina, active in the field of territory surveys using aerial and marine drones equipped with RGB cameras, LiDAR sensors, and thermal imaging cameras. On behalf of the European Union, he coordinated several projects at a national level and/or as a local unit related to coastal management and territorial security. Since December 2017, he has been the lead partner of the Pocket Beach Management & Remote Surveillance System - Program Interreg VA Italia Malta 2014–2020. He is the author of more than 120 scientific publications.

Anselme Muzirafuti

Dr. Anselme Muzirafuti, born in Rwanda, is an assistant professor at the University of Messina (Department of Mathematics, Computer Sciences, Physics and Earth Sciences, Via F. Stagno d'Alcontres, 31–98166 Messina, Italy). He holds a master's degree in applied geophysics and geology engineering obtained in 2015 at the University of Moulay Ismail, Meknes (Morocco), and a PhD in hydrogeophysics obtained in 2021 at the same university. Since 2009, he has received multi-excellence scholarships from different governments for his higher education studies, including the Government of Rwanda, the Government of the Kingdom of Morocco, and the European Union. In 2016 and 2018, he participated in major conferences on climate change and sustainability, namely, the 22nd conference of parties held in Marrakech (Morocco) and the 24th International Sustainable Development Research Society Conference (Action for a Sustainable World: from Theory to Practice), held in Messina (Italy). His research interest has focused on the use of structural geology, geomatics and geophysics for the sustainable management of territories. He worked on different projects in Morocco and Italy related to geomorphological mapping and surveys using images acquired with satellites and drones. He recently worked as an analyst of satellite images in the BESS project Pocket Beach Management & Remote Surveillance System - Program Interreg VA Italia Malta 2014–2020. The results of his works have been presented in international conferences and published in international journals.

Dimitrios S. Paraforos

Prof. Dr. Dimitrios S. Paraforos is serving as a deputy professor of agricultural engineering in the Special Crops at the Hochschule Geisenheim University (Department of Technology, Von-Lade-Str. 1, D-65366 Geisenheim, Germany). He holds an MSc from the University of Thessaly (Greece) and, in 2016, obtained a PhD from the University of Hohenheim, both in agricultural engineering. His research focuses on precision farming, digital technologies in agriculture, and, more generally, on control systems, robotics, and automation applied to agriculture; in addition, he has obtained industry experience working as an automation engineer in the food industry. He is the author of more than 53 papers indexed by Scopus, with an important contribution in the field of sensors and ISOBUS technologies for the enhancement of agricultural practices. Currently, he is coordinating the ERA-NET ICT-Agri II European project iFAROS on developing methods for increasing the efficiency and precision of site-specific fertilizer applications.

Stefania Lanza

Dr. Stefania Lanza is an administrator of Geologis s.r.l, an academic branch of the University of Messina (Via F. Stagno d'Alcontres, 31-98166 Messina, Italy). In 2007, she obtained a PhD in geology from the University of Messina (Italy) with a thesis on "The risk assessment of coastal areas: from planning to monitoring". She worked on different projects related to sedimentology and geomorphology mapping. In 2008, she took the final master's course exam with a thesis entitled: "Coastal monitoring of the coast of Badalona (Spain) contribution to the 2007-2008 survey campaign" supervised by Prof. Jordi Serra of the Autonomous University of Barcelona. In 2013, she co-founded Geologis s.r.l., a branch of the University of Messina, active in the field of territory surveys using aerial and marine drones equipped with RGB cameras, LiDAR sensors, and thermal imaging cameras. She recently worked as the coordinator of the geomorphological and sedimentological activities of the BESS project "Pocket Beach Management & Remote Surveillance System - Program Interreg VA Italia Malta 2014-2020". She is currently working as a coordinator of field activities for the "BIOBLU project - Robotic Bioremediation for Coastal Debris in Blue Flag Beach and in a Maritime Protected Area - Interreg V-A Italy - Malta 2014-2020 program".

Preface to “Future Transportation”

This Special Issue on Future Transportation falls in the scope of the current effort to mitigate and adapt to a changing climate. It is launched with the aim of collecting and promoting recent scientific studies proposing and evaluating advances in technologies leading to the sustainable future transportation of people and goods. It mainly addresses the policy makers, entrepreneurs and academicians engaged in the fight against climate change by tackling the main contributors of greenhouse gas (GHG) emissions. A special thanks is addressed to the authors who submitted their manuscripts to contribute to this initiative. We acknowledge the funders who financially assisted the authors to conduct their research, and we also thank the reviewers and editors who contributed to the evaluation of the scientific quality of the submitted manuscripts.

Giovanni Randazzo, Anselme Muzirafuti, Dimitrios S. Paraforos, and Stefania Lanza

Editors

Article

Automobile Technological Transition Scenarios Based on Environmental Drivers

Julieth Stefany García ¹, José D. Morcillo ^{2,*}, Johan Manuel Redondo ³ and Mauricio Becerra-Fernandez ⁴

¹ Facultad de Minas, Universidad Nacional de Colombia, Medellín 050034, Colombia; jugarciaco@unal.edu.co

² Facultad de Ingeniería y Tecnologías, Universidad de Monterrey, Monterrey 66238, Mexico

³ Facultad de Ciencias Económicas y Administrativas, Universidad Católica de Colombia, Bogotá 110231, Colombia; jmredondo@ucatolica.edu.co

⁴ Facultad de Ingeniería, Universidad Católica de Colombia, Bogotá 110231, Colombia; mbecerra@ucatolica.edu.co

* Correspondence: jose.morcillo@udem.edu.mx

Abstract: Different industrial sectors are assuming measures to mitigate their greenhouse gas emissions, facing the imminent materialization of climate change effects. In the transport sector, one of the measures involves the change in energy source of vehicles, leading to a transition from vehicles powered by fossil fuels (conventional) to electric. Nevertheless, electric vehicles have different drivers that promote their purchases. This work only considers the informed buyers' interest in making their decisions using environmental criteria. However, these technologies have a series of impacts, including the generation of hazardous waste such as used batteries, which leads consumers to question the environmental impacts generated by conventional and electric vehicles; consequently, it is uncertain which prospective scenarios will dominate in various nations and what will promote them. Therefore, the proposed model is studied as a dynamical system, with bifurcations of codimension 2, which means that it is possible to represent all possible prospective scenarios of this configuration through a bifurcation diagram. In this way, the analysis allows us to find that four families of technological transitions (trajectories that qualitatively can be identified as being of the same behavior class) emerge from the relationships established in the system, showing similarities to the different transition situations recognized on the planet. This model is an attractive tool to classify automobiles' technological transitions, despite having no other criteria. In fact, although decarbonization is an urgent quest in the transport sector, there are still too many challenges to guarantee environmentally friendly technologies.

Keywords: technological transitions; automobiles; system dynamics; dynamical systems; bifurcations

Citation: García, J.S.; Morcillo, J.D.; Redondo, J.M.; Becerra-Fernandez, M. Automobile Technological Transition Scenarios Based on Environmental Drivers. *Appl. Sci.* **2022**, *12*, 4593. <https://doi.org/10.3390/app12094593>

Academic Editors: Giovanni Randazzo, Dimitrios S. Paraforos, Stefania Lanza, Anselme Muzirafuti and Daniel Villanueva Torres

Received: 25 February 2022

Accepted: 29 April 2022

Published: 1 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In 2015, during the Paris Conference, governments agreed to limit global warming to levels below preindustrial levels [1–3]. To achieve this goal it will be necessary to keep emissions between 420 and 1200 GT of CO₂ by 2100 [4,5], which led to 190 countries making commitments that define voluntary climate actions until 2030, such as Nationally Determined Contributions (NDCs) and action plans to achieve targets of the Sustainable Development Goals [6].

The energy sector is one of the largest contributors to emissions worldwide. For 2021, this sector emitted approximately 33 GT of CO₂, which shows a growth of 4.8% compared to 2020 [7].

For its part, the transportation sector is responsible for 24% of direct emissions of CO₂ from the use of fuels [8,9]. Likewise, road automobiles account for almost three-quarters of the CO₂ emissions from transport and are also the fragment with the greatest opportunity for decarbonization [10]. For this reason, to fulfill the agreed commitments, the need

to achieve an energy transition towards zero and low emission technologies have been identified to counteract the effects of climate change [11].

Rapid urbanization, the dynamism in the populated centers, and the dependence on transportation to improve the quality of life have increased the demand for automobiles, of which a large proportion run on fossil fuels that are strongly related to the emissions of greenhouse gases (GHG), leading to increased air pollution, increased respiratory illnesses, and impaired quality of life [12,13].

Among the alternatives established by nations to mitigate the emissions of CO₂ from the transport sector are automobiles with alternative technologies that use zero and low emission fuels, such as hybrid vehicles (HEV), plug-in hybrid vehicles (PHEVs), and battery electric vehicles (BEVs). These alternatives are valuable for consumers who make their decisions based on the environmental impact generated by these technologies [14–18].

As a consequence, the growth in vehicle sales of these technologies has intensified. According to the International Energy Agency, 2.2 million electric cars were sold in 2019, representing 2.5% of global car sales. During 2020, the automobile market was contracted due to the effects of the pandemic, but sales of electric cars increased to 3 million, representing 4.1% of total automobile sales. During the year 2021, sales of electric vehicles doubled and reached a total of 6.6 million, which represents about 9% of the global car market [19]. According to the data, a total of 16 million electric vehicles travel on roads worldwide [19], with 90% of global sales of these technologies concentrated in China, Europe, and the USA [10].

A Tank to Wheel analysis conducted in [20] found that battery-powered electric cars reduce CO₂ emissions by 32% compared to their conventional counterparts. Likewise, the transition towards these technologies reduces GHG emissions and contributes to the fulfillment of the commitments agreed by countries [9,13,21–24]. However, to achieve decarbonization in the transport sector, it is necessary to encourage generation with non-conventional sources of renewable energy, given that those nations in which fuels and coal have a high share may not perceive the benefits of the electrification of the sector [25–29]. For this reason, countries with high emissions of CO₂ must diversify their energy generation matrix, take advantage of their electric vehicle capacity, and reduce CO₂ emissions [30,31].

Electric vehicles have different drivers that promote their purchase; one of the main drivers is the interest of informed buyers who make their decisions using environmental criteria, such as reducing noise pollution and greenhouse gas emissions and improving energy efficiency, as well as improving the ecological image that the individuals and/or organizations have [32].

However, the zero and low emission vehicles sector still faces significant challenges in the transition towards electric technologies, such as increasing the charging infrastructure, increasing the energy efficiency, increasing the penetration of renewable energies in the energy matrix, and reducing the costs of batteries and charging stations, as well as improving the profitability of the cars [33]. Furthermore, it is necessary to establish incentives or policies that governments can promote to ensure the rapid and efficient adoption of environmentally friendly cars [34].

The transition to electric automobiles also has a series of negative significant environmental impacts related to mineral exploitation and final disposal of the electric and electronic parts used [35]. Examples of this are lithium, cadmium, and nickel, which show a balance between supply and demand in the short and medium-term but a scarcity in the long term due to the expected increase in demand for hybrid and electric vehicles (BEV, PHEV, and HEV) [36] and current recycling rates below 1% [36].

Regarding waste, the components of electric batteries are hazardous waste, causing significant health and environmental problems without proper management [37]. In this sense, it is necessary to establish recycling systems and carry out socialization actions to recover the lithium, nickel, and cadmium and be able to reintegrate them into the value chain, both for vehicle batteries and for other uses [37]. Furthermore, the battery's

performance must be improved to reduce the intensity of exploitation of these minerals in the sector, while increasing other efforts to develop alternative options [36].

This work discusses the transition scenarios of automobile technologies towards zero and low emission alternatives, considering that current technologies generate other impacts related to the disposal of hazardous solid waste. The assumption is that the technological change decision has as its only determinants emissions and hazardous waste reduction; so, the buyers base their decision on environmental responsibility. As a result, the buyers fall into the paradox of choosing between the impact alternatives due to the benefits derived from automobile use.

Different authors have investigated issues in the transition towards zero and low emission technologies [32], and their approaches and methodologies are varied. Some methods to analyze transition are life cycle analysis [15], modeling with different techniques [31,35,38–43], system dynamics [35,44,45], cost analysis [23,29,46], literature reviews [47,48], linear regressions [49], LEAP software models [24], surveys and interviews [50–52], ANSWER MARKAL software models [53], and Well to Wheel analyses [26]. These authors have focused their studies on evaluating and analyzing how technological transitions occur in national transport sectors. However, they only address the current conditions and the incentives implemented in each country.

Conversely, the approach presented in this work aims to be systemic, by identifying the trend behavior derived from consumers whose only criterion for automobile choice is environmental responsibility. In this sense, we present a first approximation that allows identifying different categories of national transition, called prospective transition scenarios, which emerge from an analysis based on bifurcation theory. This approach is novel in the literature, only a first approximation for analyzing capacity scenarios in electric markets was proposed in [54,55].

In this way, this work recognizes and describes the technological transition scenarios involving conventional and electric automobiles based on the environmental drivers of emissions and disposal of hazardous waste.

In summary, the emissions generated by the transport sector are decisive when consumers take action. Among the alternatives to decarbonize the transport sector is the transition from conventional to less polluting technologies. In particular, electric vehicles are an alternative that has gained strength, which has led them to increase their market share. This technology has become more popular among shoppers who make environmentally responsible purchasing decisions. However, EVs generate hazardous waste from their batteries. This causes the consumers to question whether to prioritize the environmental impacts generated by fuel vehicles or EVs. This will impact the growth or reduction in the fleet of both of these technologies. As a consequence, it is uncertain which prospective scenarios will have significant participation and what will promote them. In this research work, only the informed decision of consumers with an environmental commitment was considered a driver.

The document is organized as follows. This introductory section has shown the problem and purpose of this work, along with the state of the art. Section 2 presents the mathematical modeling carried out from system dynamics to obtain the dynamical system from which the results come. This section explains a dynamical system, a bifurcation analysis, how the system equilibrium points were calculated, and how their stabilities were established. Section 3 presents the phase portraits, time series, and the bifurcation diagram obtained for the dynamical system under study. In Section 4, the implications of the results are presented and future work is proposed. Finally, Section 5 presents the conclusions of this research.

2. Materials and Methods

The starting point of the analysis presented in this document is the mathematical model of Equation (4). This system of first-order ordinary differential equations has been obtained from system dynamics methodology (see for example [56]).

The rules of causality are simple: (1) conventional automobiles generate emissions of CO₂ [22,23,57], (2) electric automobiles produce hazardous waste [36,58], (3) emissions and hazardous waste generate nonconformity in vehicle consumers taking the option to change technology to reduce its environmental impact [32], (4) the enthusiasm to acquire one or another technology leads to the entry of automobiles into the system. The hazardous waste of conventional vehicles is out of the scope of this research work, since it is not as important a selection criterion for buyers as the hazardous waste of electric automobiles.

From these causality rules, we built the Forrester diagram shown in Figure 1. Conventional and electric automobiles are the level variables of the system. The equations presented below are an interpretation of the Forrester diagram. The rules used were [56]: (1) the change over time in level variables, represented in boxes, is the difference between the inflow and outflow variables, symbolized as valves; (2) the flow variables are the products between the variables connected with arrows, and (3) the auxiliary variables (represented as circles), which are functions of the variables connected through arrows, are defined according to the relationship between them.

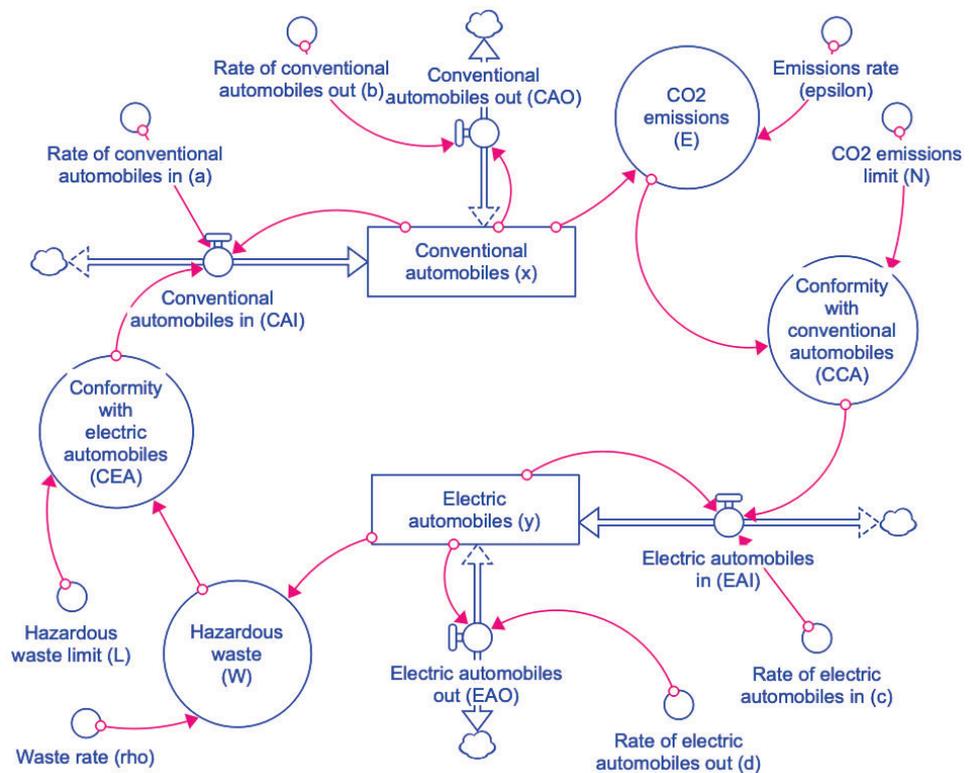


Figure 1. Stocks and Flows diagram of the automobile technology transition from environmental drivers.

The conventional automobiles x change over time (represented by a point on the variable), given in [automobiles], is the difference between automobiles entering the market CAI and those leaving CAO , both given in [automobiles/year]. The electric automobiles y change over time, given in [automobiles], is the difference between the automobiles that enter the market EAI and those leaving EAO , both given in [automobiles/year]. So, the level equations are defined as follows:

$$\begin{aligned} \dot{x} &= CAI - CAO \\ \dot{y} &= EAI - EAO \end{aligned} \tag{1}$$

The change ratio equations presented in Equation (1) show that the increase in the quantities of each automobile category depends on the disagreement that users have with the other category, and the decrease depends on an average depreciation, as follows:

$$\begin{aligned} CAI &= a \cdot x \cdot (1 - CEA) & CAO &= b \cdot x \\ EAI &= c \cdot y \cdot (1 - CCA) & EAO &= d \cdot y \end{aligned} \tag{2}$$

where $a, b, c,$ and d are exchange rates in the range $[0, 1]$, given in $[year^{-1}]$, and the conformities with each category depend on the closeness of the emissions E , given in $[TonCO_2/year]$, or hazardous waste disposal W , given in $[Ton/year]$, with an allowable limit value (emissions limit N , given in $[TonCO_2/year]$, and hazardous waste limit W , given in $[Ton/year]$) that could be agreed through national or international environmental commitments. The equations for emissions and hazardous waste are defined as follows:

$$\begin{aligned} CEA &= W/L & CCA &= E/N \\ E &= \epsilon \cdot x & W &= \rho \cdot y \end{aligned} \tag{3}$$

where $\epsilon \geq 0$ and $\rho \geq 0$ are exchange rates given in $[TonCO_2/(automobile \cdot year)]$ and $[Ton/(automobile \cdot year)]$, respectively.

Finally, by replacing the equations, and considering $\alpha = a - b, \beta = c - d, \delta = a\rho/L,$ and $\lambda = c\epsilon/N$, we obtain the following mathematical model:

$$\begin{aligned} \dot{x} &= x(\alpha - \delta y) \\ \dot{y} &= y(\beta - \lambda x) \end{aligned} \tag{4}$$

Note that if the system of Equation (4) is studied under the variation of α and β , different scenarios of exchange rates will be considered.

The model validation was performed using the tests reported in [59]. In particular, we conducted the empirical structure confirmation test and empirical parameters confirmation test. The theoretical structure confirmation test was performed through the literature review mentioned above to explain the causal rules construction. The theoretical parameters confirmation was carried out during the construction of the model by defining the application ranges, as shown above. The dimensional consistency test appears together with the equations explanation, verifying the units proposed. We also carried out extreme value tests allowing us to recognize the limits and scope of the model. When possible, the bifurcation analysis is better than the system sensitivity tests because the bifurcation analysis can identify all the probable system behaviors, while the sensitivity analysis only allows visualizing the transitory state under the parameter variation in a limited range, using the Monte Carlo method. For this reason, our bifurcation analysis shown below supports the validity of our proposed model. Finally, we carried out phase relationship tests to compare the behavior between the state variables and verify the behavior consistency through phase portraits, as shown in Section 3. We did not conduct behavior validation with real time series, because it is still too early to have enough national data that could be comparable with the behaviors shown by our prospective mathematical model.

The mathematical model of Equation (4) is a continuous dynamical system. For this reason, we analyzed the equilibrium points and their stability and determined that the system has bifurcations in codimension two.

A dynamical system (see [60,61]) is a triple (X, T, φ) , where X is the state space, T is an ordered set representing time, and

$$\begin{aligned} \varphi &: T \times X \longrightarrow X \\ (t, x) &\longmapsto \varphi_t(x) \end{aligned} \tag{5}$$

is a parameterized family of evolution operators that satisfies the following:

- The system state does not change spontaneously

$$\varphi_0(x) = x \quad \text{for all } x \in X \tag{6}$$

- The evolution law of the system does not change in time

$$\varphi_{t+s}(x) = (\varphi_t \circ \varphi_s)(x) = \varphi_t(\varphi_s(x)) \quad \text{for all } x \in X, \text{ and all } t, s > 0. \tag{7}$$

In the model of Equation (4), the state space is defined by the state variables of conventional automobiles x and electric automobiles y ; that is, $X = \{(x, y) \in \mathbb{R}\}$. The set of time is a subset of the real numbers, so the dynamical system will be said to be continuous; that is, $T = \{t \in \mathbb{R} : t \geq 0\}$. Finally, the evolution law is given by the solution of the differential system of Equation (4).

The mathematical perspective from dynamical systems is relevant in a trend behavior analysis because it allows characterizing the qualitative structure of the system’s behavior beyond just understanding the behavior for a specific time, which could be misleading. Roughly speaking, the dynamical systems identify invariant sets and characterize their stability. Invariant sets are states that do not change as the system evolves, i.e., state sets that remain in time. Examples include equilibrium points, periodic orbits, and strange attractors. The stability can characterize the system trajectories as attracting, repelling, fixing (centers), or combining these three options.

Since the equilibrium points occur when the system does not change, it must be satisfied that \dot{x} and \dot{y} are zero [60,61], making the differential problem an algebraic issue, as follows:

$$\begin{aligned} \dot{x} = 0 & \implies x(\alpha - \delta y) = 0 \\ \dot{y} = 0 & \implies y(\beta - \lambda x) = 0 \end{aligned} \tag{8}$$

The stability of the equilibrium points is calculated with the eigenvalues of the evolution matrix of the differential system around the equilibrium points [60,61], i.e., consider Equation (4) around the equilibrium points in the form $\dot{X} = JX$, where $X = (x, y)^T$ is the state vector, and J is the Jacobian matrix defined as:

$$J(x, y) = \begin{pmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial y} \\ \frac{\partial \dot{y}}{\partial x} & \frac{\partial \dot{y}}{\partial y} \end{pmatrix} = \begin{pmatrix} \alpha - \delta y & -\delta y \\ -\lambda y & \beta - \lambda x \end{pmatrix} \tag{9}$$

By the Hartman-Grobman theorem [60,61], the nonlinear behavior so close to an equilibrium point is the linearized behavior given by the Jacobian matrix in the equilibrium point. So, the evolution matrix around an equilibrium point (\bar{x}, \bar{y}) is $J(\bar{x}, \bar{y})$.

The eigenvalues χ are defined as the roots of the characteristic polynomial $\chi^2 - tr(J)\chi + det(J)$, where $tr(J)$ is the trace of the Jacobian matrix J obtained by summing the diagonal entries of J , and $det(J)$ is the determinant of the Jacobian matrix.

A more robust analysis, called bifurcation analysis, also seeks to identify how the system parameters’ variation can give rise to the appearance/disappearance of invariant sets or change their stability. This bifurcation analysis consists in finding system perturbations that transform its phase portraits (system trajectory structure) [60,61]. These perturbations are represented by the variation of the system parameters in such a way that if the variation of n parameters qualitatively changes the phase portraits, it is said to be a bifurcation of codimension n [60,61].

3. Results

The results presented in this section are based on the dynamical system analysis of Equation (4). These ordinary differential equations were programmed and solved using the ODE45 in the MATLAB® software. Then, and in order to keep the sense of the application, four transition families in the nonnegative space of the state variables were found. To clarify, from a qualitative perspective, a transitions family is a set of similar behavior

curves (topologically equivalent trajectories in the states space). For example, two orbits that converge to the same equilibrium point showing a sigmoidal behavior are said to be topologically equivalent, although the route of their trajectories is different (quantitatively different). Usually, these behaviors are easily visualized in the phase portrait, which is a representation of all possible solutions or trajectories of a system of differential equations for any initial condition and a specific set of parameter values. As shown in Figures 2 and 3, each of these transition families is displayed as a *different* phase portrait.

By solving Equation (8), it is found that Equation (4) has the equilibrium $Eq_1(0, 0)$ and $Eq_2(\beta/\lambda, \alpha/\delta)$, each of which has a stability that depends on the values of the α and β parameters. For the equilibrium Eq_1 , the eigenvalues are α and β ; while for the equilibrium Eq_2 , the eigenvalues are $\pm\sqrt{\alpha\beta}$. This leads to four cases called transition families; each one is presented in Figures 2 and 3.

The first family transition corresponds to two unstable nodes (see the phase portrait in Figure 2a). The first is located at the origin (source type), while the second is in the nonnegative states space (saddle type). Because it is possible to have different trajectories in the phase portrait depending on the initial condition, this family transition has four distinct trend behaviors (which is why we call the transition scenarios families).

For the first case, see Figure 2a,b, a behavior is visualized in which the two technologies grow over time until they reach a particular point, in which electric automobiles tend to disappear, while the conventional vehicles become dominant. For the second case, see Figure 2c,d, the behavior is similar, but the disappearance now occurs for conventional automobiles, while electric vehicles become dominant. In the third case, see Figure 2e,f, there is a drop in the technologies used that, after a specific time, ends and causes electric automobiles to recover and become the dominant technology. Finally, in the fourth case, see Figure 2g,h, there is again an abandonment of the two technologies, but the conventional technology recovers in the long term.

The second family transition corresponds to an unstable node at the origin (saddle type), see the phase portrait in Figure 3a. Another equilibrium point is outside the nonnegative states space (center type), but its analysis is irrelevant for the application. This second family transition is the one that looks like the most expected transition: conventional automobiles diminish until they disappear, while electric automobiles gradually increase their dominance of the market, until they are the only type of vehicle (see Figure 3b).

The third family transition corresponds to an unstable node at the origin (saddle type), see the phase portrait in Figure 3c. Another equilibrium point is outside the nonnegative states space (center type), but its analysis is irrelevant for the application. This scenario appears as a theoretical possibility, and does not represent the interests of those who talk about the transition of vehicle technologies; it shows the strengthening of conventional automobiles, while electric automobiles disappear (see Figure 3b).

The last family transition corresponds to a stable node at the origin (sink type), see Figure 3e. Another equilibrium point is outside the nonnegative states space (saddle type), but its analysis is irrelevant for the application. In this scenario, the absence of interest of consumers in the two types of technologies leads to their disappearance as transport alternatives (see Figure 3f).

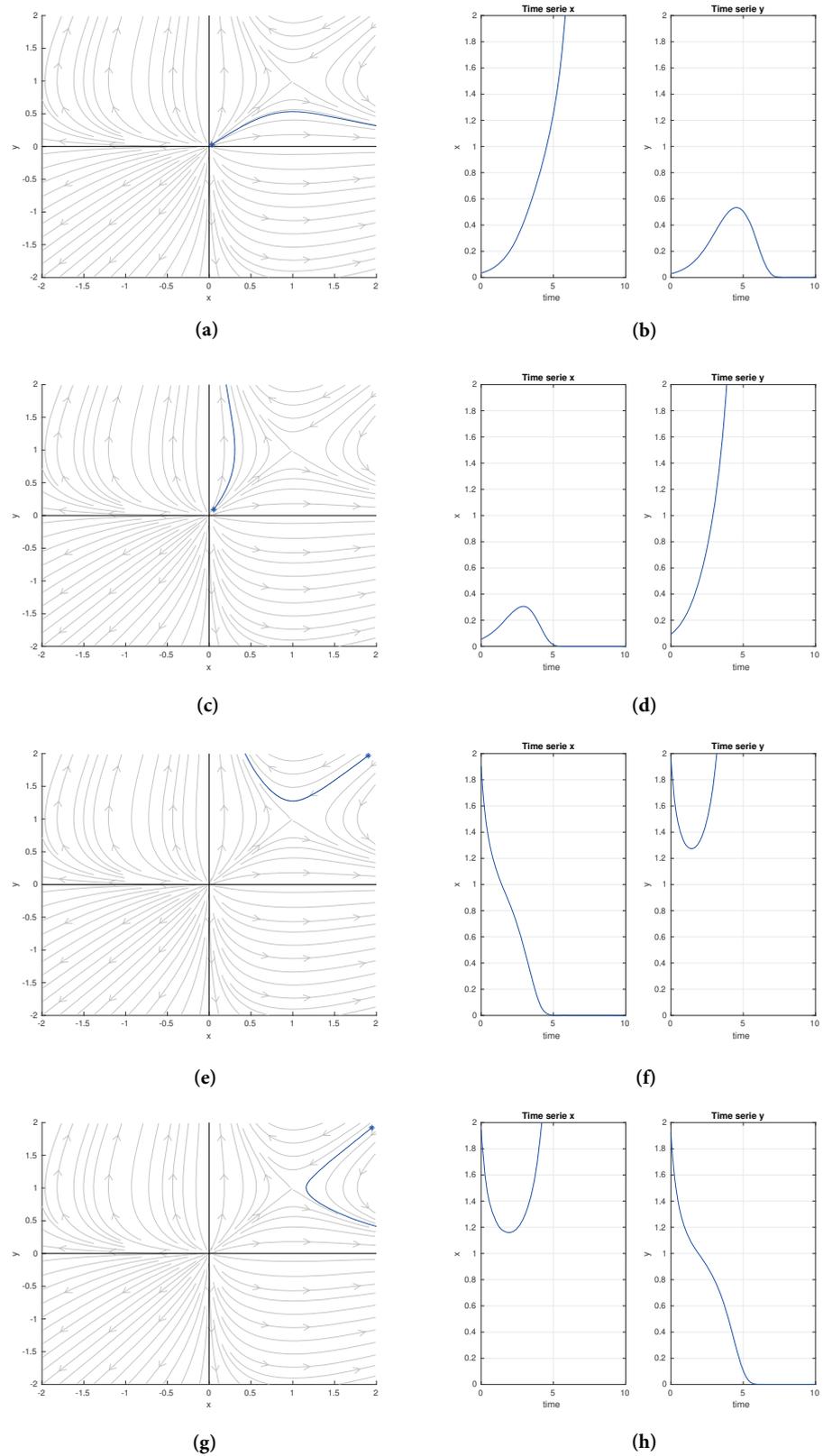


Figure 2. Case 1 simulations. This case occurs when vehicle entry rates are higher than exit rates in both categories. In (a,b), both technologies grow; however, finally, the conventional one dominates, while in (c,d), the opposite occurs. In (e,f), both technologies decrease, but the electric one manages to recover; while in (g,h), the opposite occurs.

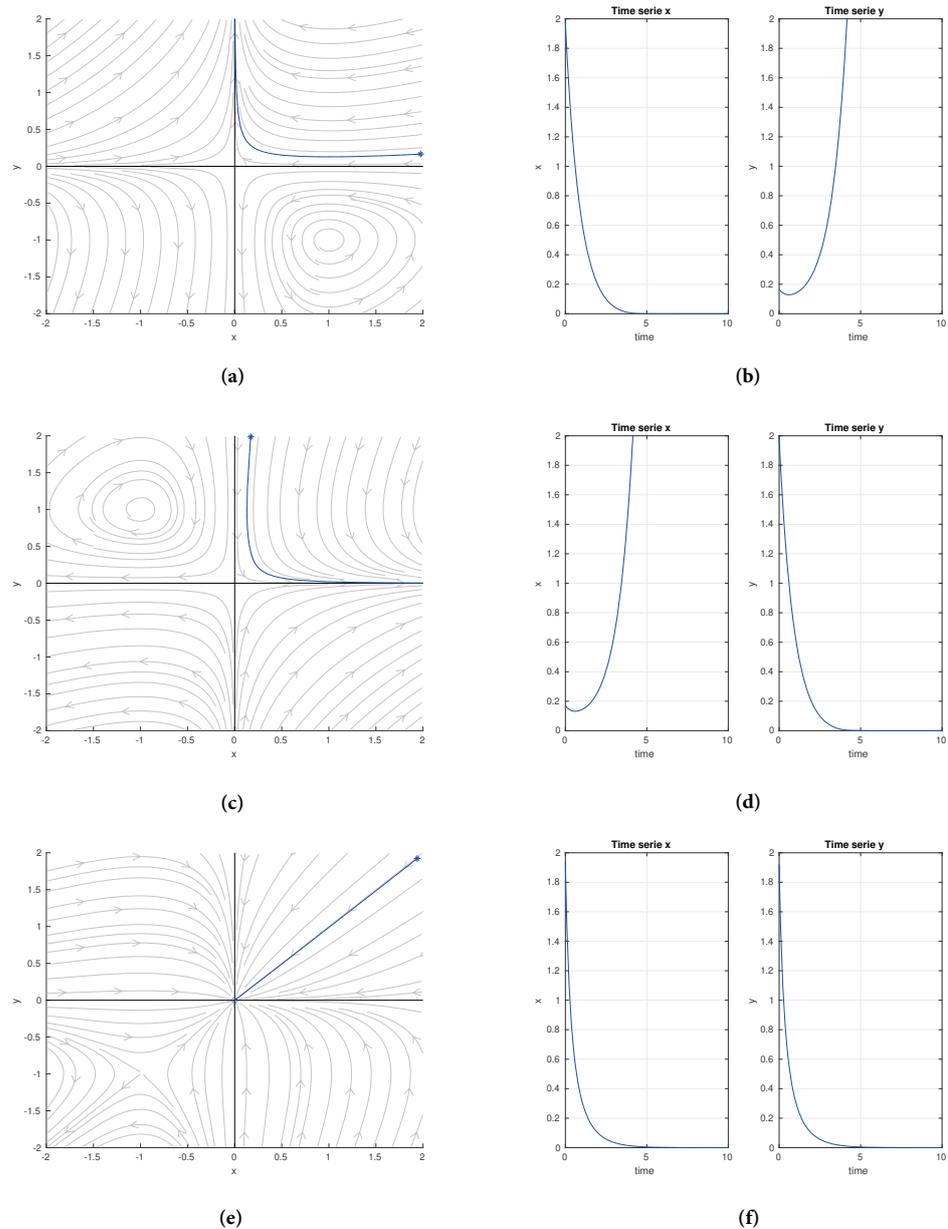


Figure 3. Simulations for cases 2, 3 and 4. This figure corresponds to another case of entrance rates growth being higher than exit rates in both automobile categories. In (a,b), the scrapping of conventional automobiles is greater than the entry into the fleet (sales), while in (c,d), sales of conventional automobiles are greater than their operating output (scrap), which increases the number of automobiles in the fleet. Finally, in (e,f), the lack of interest in the two technologies can be seen, causing their disappearance.

Two-Dimensional Bifurcation Analysis

In principle, one-dimensional bifurcation diagrams are computed when it is necessary to study the stability or behavior of a dynamical system under the variation of any of its parameters. On the other hand, when it is necessary to study the stability or behavior of a dynamical system under the simultaneous variation of two important parameters, then a two-dimensional bifurcation diagram is required [60]. In essence, two-dimensional bifurcation diagrams show different colors, each of which represents a precise combination of the two parameters leading to a specific system behavior.

To obtain these bifurcation diagrams, it is necessary to program a behavior detector, so that when a specific behavior is detected, the program stores the parameters leading to such dynamics, and an identification number will represent a color in the diagram.

Since the equilibrium points' stability of the dynamical system in Equation (4) depends on the bifurcation parameters α and β , it is possible to obtain a two-dimensional bifurcation diagram, as depicted in Figure 4. This diagram summarizes the behaviors displayed by the system for the different values of the two bifurcation parameters. Notably, the parameter array with the lowest probability of occurrence is for the first family of transitions (in the figure, this family appears as case 1 for graphical simplicity). The parameter arrays to obtain transition families 2, 3, and 4 have been denoted as case 2, case 3, and case 4, respectively. The probability areas of these last cases are broader than case 1; so, in theory, their occurrence chances must be greater than case 1.

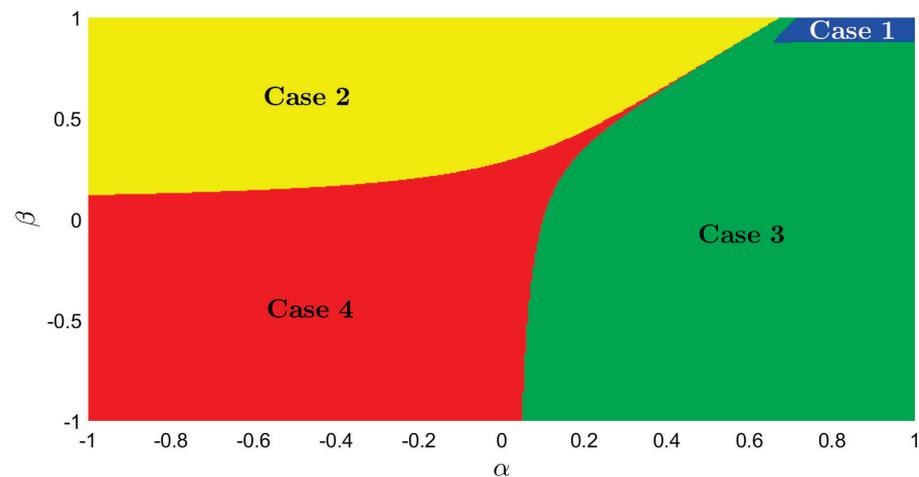


Figure 4. Two-dimensional bifurcation diagram varying α and β . Every color represents the system behavior for each case previously analyzed.

4. Discussion

Below is a discussion of the behaviors found in each transition family of vehicle technology from environmental drivers. Each of these transition families would allow the recognition of a different type of transition worldwide, allowing their classification, as will be shown next. Furthermore, we present the possible future work of this research.

4.1. Family Transition 1

From the application perspective, transition 1 is interesting, because the stable manifold and the unstable manifold of the saddle node define four key regions whose behaviors are different.

The trend behaviors presented in Figure 2b,d can be identified with the current behavior, in which the number of conventional and electric automobiles continues growing globally. The long-term outcome is the one that can change due to the final dominance of one of the two technologies. Many nations would be classified in this state today. However, in developed countries, the behavior could be similar to the one shown in Figure 2d, due to the greater investment capacity and awareness of climate change that its inhabitants have; while in underdeveloped countries, it is easier to interpret their trend behavior as that shown in Figure 2b, due to the survival situation of their inhabitants, in addition to the massive confluence of the population in urban centers, where restrictive transport measures and the inability to purchase an electric vehicle due to other criteria than environmental ones (such as the high price of these technologies) determine the purchase of a second conventional vehicle.

On the other hand, the trend behaviors presented in Figure 2f,h would correspond to the assumption that the number of conventional and electric automobiles is so high that

there is a general rejection of these automobiles use because of their environmental impacts over climate change and land health (related with hazardous waste). This leads in the long term to the disappearance of one of the two technologies, while the other one becomes dominant. This behavior can be considered in nations in which vehicle use is abandoned in favor of migrating towards public transport and active mobility.

Despite the fact that the behavior shown in this transitional family appears to be the least probable in the bifurcation diagram of Figure 4, the values for which it is satisfied are not utopian. Nor it is utopian to think that the engine of this de-escalation of conventional and electric automobiles could occur due to the impacts they generate on the environment.

4.2. Family Transition 2

The trending behavior presented in Figure 3b can be pointed out as the most natural and expected within the transitions. The trajectories show a smooth transition from conventional to electric technologies, which is how a handover would occur.

It is the most desired transition, led by environmental awareness about climate change and its consequences on global socioecological systems. In this transition, the issue of hazardous waste is an issue on which we can still wait.

4.3. Family Transition 3

The trending behavior presented in Figure 3b shows that automobiles with electric technologies are not convincing due to their effects on the environment related to the management of their hazardous waste.

It should be remembered that the model does not consider the mining activities and their social and ecological implications in the nations where the raw materials are obtained. This should be included in future work for analyzing these transitions. It is intuited that its inclusion in the model could generate new transition behaviors, but it must be tested.

The question would be: how could this scenario be viable when fossil fuel reserves will be scarce so soon?

4.4. Family Transition 4

Finally, the trending behavior presented in Figure 3f shows the lack of interest in the two technologies, causing their disappearance. In this case, no negative environmental effects are prioritized, showing losing interest in the technologies based on conventional or electric mobility and migrating to alternatives such as active mobility or the use of public transport.

4.5. Future Research

The research group recognizes that the only transition drivers are not those related to a full awareness of the environmental implications of each of the vehicle technologies presented. We can also list those corresponding to the price of the vehicle ([50,62]), the charging infrastructure [47,63], the energy value [64], the incentives offered by the government [18,44,51], and the useful life of the technology [38,65].

In this way, the future work must include other criteria in the modeling, with the corresponding analysis of the trend behavior that each new criterion defines on the system behavior.

However, the related issue that makes this inclusion nontrivial lies in the bifurcation analysis, through which all possible prospective scenarios for the system configuration could be visualized [66].

We also expect to obtain management recommendations that at least guarantee adequate management of the related environmental issues to each family of transitions.

5. Conclusions

This work allows the classification of the different types of technological transitions that are taking place around the world, from conventional automobiles to electric auto-

mobiles, motivated by a single decision criterion for consumers based on environmental awareness.

The transition scenario definition has not been defined a priori but has emerged from systemic modeling and its corresponding bifurcation analysis.

Due to the bifurcation analysis, it can be assured that the trend behaviors for the configuration and assumptions made to the system are completely bounded to the four families of transitions listed in this work. In this way, specific recommendations for the four transition families can be generated rather than general ones that end up ignoring the complexities of each of the configurations.

Promoting environmental awareness and disseminating the environmental benefits of electric vehicles can contribute to the growth of the fleet, increasing its popularity among a segment of buyers whose decision criterion is protecting the environment. For this popularity not to diminish, it is necessary to establish actions and programs that allow the creation of a relevant value chain for the waste generated by the use of EVs, such as lithium or nickel. This will allow the reduction in mining exploitation. These are public policy actions that could promote the exit of the circulation of conventional technologies and contribute to the circular economy commitments of the countries.

On the other hand, if the two technologies are not of interest to users, it is relevant to strengthen public transport and active mobility in populated centers, guaranteeing road infrastructure, safety, quality, and equitable access. This will imply the development of public policies and government investments.

Finally, the energy transition of the transport sector is a priority step in the decarbonization of anthropogenic activities to tackle climate change. However, this work shows that there are still many challenges for reducing or eliminating the impacts that energy systems have on nature. This raises questions about energy sources, the materials used, processes' efficiency, etc. In this way, this work reinforces the evidence that the present energy solutions are short-term and present a challenge to physicists and engineers to propose solutions that satisfy long-term multidimensional needs.

Author Contributions: Conceptualization, J.S.G., J.D.M., J.M.R. and M.B.-F.; methodology, J.M.R. and J.D.M.; investigation, J.S.G. and M.B.-F.; validation, J.M.R. and J.D.M.; formal analysis, J.M.R. and J.D.M.; writing—original draft preparation, J.S.G., J.D.M. and J.M.R.; writing—review and editing, J.S.G., J.D.M. and J.M.R.; visualization, J.D.M. and J.M.R.; project administration, J.S.G.; funding acquisition, J.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad de Monterrey and Universidad Católica de Colombia under the project CON0000497 “Dynamic model for renewable energy supply in Colombia”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors want to acknowledge Universidad de Monterrey and Universidad Católica de Colombia for their support of this research work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rhodes, C.J. The 2015 Paris Climate Change Conference: COP21. *Sci. Prog.* **2016**, *99*, 97–104. [[CrossRef](#)]
2. Guan, D.; Meng, J.; Reiner, D.M.; Zhang, N.; Shan, Y.; Mi, Z.; Shao, S.; Liu, Z.; Zhang, Q.; Davis, S.J. Structural decline in China's CO₂ emissions through transitions in industry and energy systems. *Nat. Geosci.* **2018**, *11*, 551–555. [[CrossRef](#)]
3. Zeng, S.; Li, G.; Wu, S.; Dong, Z. The Impact of Green Technology Innovation on Carbon Emissions in the Context of Carbon Neutrality in China: Evidence from Spatial Spillover and Nonlinear Effect Analysis. *Int. J. Environ. Res. Public Health* **2022**, *19*, 730. [[CrossRef](#)]
4. Goodwin, P.; Katavouta, A.; Roussenov, V.M.; Foster, G.L.; Rohling, E.J.; Williams, R.G. Pathways to 1.5 C and 2 C warming based on observational and geological constraints. *Nat. Geosci.* **2018**, *11*, 102–107. [[CrossRef](#)]

5. The Intergovernmental Panel on Climate Change. Mitigation of climate change. In *Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2014; Volume 1454, p. 147.
6. Winkler, H.; Boyd, A.; Torres Gunfaus, M.; Raubenheimer, S. Reconsidering development by reflecting on climate change. *Int. Environ. Agreem. Politics Law Econ.* **2015**, *15*, 369–385. [[CrossRef](#)]
7. Agency, I.E. *Global Energy Review 2021*; International Energy Agency: Paris, France, 2021.
8. Agency, I.E. *World Energy Outlook 2020*; International Energy Agency: Paris, France, 2020.
9. Gil-García, I.C.; García-Cascales, M.S.; Dagher, H.; Molina-García, A. Electric vehicle and renewable energy sources: Motor fusion in the energy transition from a multi-indicator perspective. *Sustainability* **2021**, *13*, 3430. [[CrossRef](#)]
10. Agency, I.E. *Global EV Outlook 2020*; International Energy Agency: Paris, France, 2020.
11. Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; Gomis, M.; et al. Climate Change 2021: The Physical Science Basis. In *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2021.
12. Nimesh, V.; Kumari, R.; Soni, N.; Goswami, A.K.; Mahendra Reddy, V. Implication viability assessment of electric vehicles for different regions: An approach of life cycle assessment considering exergy analysis and battery degradation. *Energy Convers. Manag.* **2021**, *237*, 114104. [[CrossRef](#)]
13. Zeng, D.; Dong, Y.; Cao, H.; Li, Y.; Wang, J.; Li, Z.; Hauschild, M.Z. Are the electric vehicles more sustainable than the conventional ones? Influences of the assumptions and modeling approaches in the case of typical cars in China. *Resour. Conserv. Recycl.* **2021**, *167*, 105210. [[CrossRef](#)]
14. Sengupta, S.; Cohan, D.S. Fuel cycle emissions and life cycle costs of alternative fuel vehicle policy options for the City of Houston municipal fleet. *Transp. Res. Part Transp. Environ.* **2017**, *54*, 160–171. [[CrossRef](#)]
15. Bauer, G. The impact of battery electric vehicles on vehicle purchase and driving behavior in Norway. *Transp. Res. Part Transp. Environ.* **2018**, *58*, 239–258. [[CrossRef](#)]
16. Bakker, S.; Dematera Contreras, K.; Kappiantari, M.; Tuan, N.A.; Guillen, M.D.; Gunthawong, G.; Zuidgeest, M.; Liefferink, D.; Van Maarseveen, M. Low-carbon transport policy in four ASEAN countries: Developments in Indonesia, the Philippines, Thailand and Vietnam. *Sustainability* **2017**, *9*, 1217. [[CrossRef](#)]
17. Bergman, N.; Schwanen, T.; Sovacool, B.K. Imagined people, behaviour and future mobility: Insights from visions of electric vehicles and car clubs in the United Kingdom. *Transp. Policy* **2017**, *59*, 165–173. [[CrossRef](#)]
18. Sovacool, B.K.; Kester, J.; Noel, L.; de Rubens, G.Z. Are electric vehicles masculinized? Gender, identity, and environmental values in Nordic transport practices and vehicle-to-grid (V2G) preferences. *Transp. Res. Part Transp. Environ.* **2019**, *72*, 187–202. [[CrossRef](#)]
19. Agency, I.E. *Electric Cars Fend off Supply Challenges to more than Double Global Sales*; International Energy Agency: Paris, France, 2022.
20. Ke, W.; Zhang, S.; He, X.; Wu, Y.; Hao, J. Well-to-wheels energy consumption and emissions of electric vehicles: Mid-term implications from real-world features and air pollution control progress. *Appl. Energy* **2017**, *188*, 367–377. [[CrossRef](#)]
21. Gopal, A.R.; Park, W.Y.; Witt, M.; Phadke, A. Hybrid- and battery-electric vehicles offer low-cost climate benefits in China. *Transp. Res. Part Transp. Environ.* **2018**, *62*, 362–371. [[CrossRef](#)]
22. Duan, H.; Hu, M.; Zuo, J.; Zhu, J.; Mao, R.; Huang, Q. Assessing the carbon footprint of the transport sector in mega cities via streamlined life cycle assessment: A case study of Shenzhen, South China. *Int. J. Life Cycle Assess.* **2017**, *22*, 683–693. [[CrossRef](#)]
23. Liao, F.; Molin, E.; van Wee, B. Consumer preferences for electric vehicles: A literature review. *Transp. Rev.* **2017**, *37*, 252–275. [[CrossRef](#)]
24. Simsek, Y.; Sahin, H.; Lorca, Á.; Santika, W.G.; Urmee, T.; Escobar, R. Comparison of energy scenario alternatives for Chile: Towards low-carbon energy transition by 2030. *Energy* **2020**, *206*, 118021. [[CrossRef](#)]
25. Bauer, C.; Hofer, J.; Althaus, H.J.; Del Duce, A.; Simons, A. The environmental performance of current and future passenger vehicles: Life cycle assessment based on a novel scenario analysis framework. *Appl. Energy* **2015**, *157*, 871–883. [[CrossRef](#)]
26. Onn, C.C.; Mohd, N.S.; Yuen, C.W.; Loo, S.C.; Koting, S.; Abd Rashid, A.F.; Karim, M.R.; Yusoff, S. Greenhouse gas emissions associated with electric vehicle charging: The impact of electricity generation mix in a developing country. *Transp. Res. Part Transp. Environ.* **2018**, *64*, 15–22. [[CrossRef](#)]
27. Rahman, I.; Vasant, P.M.; Singh, B.S.M.; Abdullah-Al-Wadud, M.; Adnan, N. Review of recent trends in optimization techniques for plug-in hybrid, and electric vehicle charging infrastructures. *Renew. Sustain. Energy Rev.* **2016**, *58*, 1039–1047. [[CrossRef](#)]
28. Singh, R.; Kumar, A.; Singh, A.R.; Naidoo, R.; Bansal, R.C.; Kumar, P. Environmental feasibility of incorporation of electric taxis in South Africa. *J. Eng.* **2019**, *2019*, 5078–5084. [[CrossRef](#)]
29. Taljegard, M.; Göransson, L.; Odenberger, M.; Johnsson, F. Impacts of electric vehicles on the electricity generation portfolio—A Scandinavian-German case study. *Appl. Energy* **2019**, *235*, 1637–1650. [[CrossRef](#)]
30. Doucette, R.T.; McCulloch, M.D. Modeling the prospects of plug-in hybrid electric vehicles to reduce CO₂ emissions. *Appl. Energy* **2011**, *88*, 2315–2323. [[CrossRef](#)]
31. Jochem, P.; Babrowski, S.; Fichtner, W. Assessing CO₂ emissions of electric vehicles in Germany in 2030. *Transp. Res. Part Policy Pract.* **2015**, *78*, 68–83. [[CrossRef](#)]

32. Asghari, M.; MirzapourAl-e hashem, S.M.J. New Advances in Vehicle Routing Problems: A Literature Review to Explore the Future. In *Green Transportation and New Advances in Vehicle Routing Problems*; Springer Nature: Berlin, Germany, 2020; pp. 1–42. [\[CrossRef\]](#)
33. Hao, H.; Cheng, X.; Liu, Z.; Zhao, F. Electric vehicles for greenhouse gas reduction in China: A cost-effectiveness analysis. *Transp. Res. Part Transp. Environ.* **2017**, *56*, 68–84. [\[CrossRef\]](#)
34. Liu, X.; Sun, X.; Zheng, H.; Huang, D. Do policy incentives drive electric vehicle adoption? Evidence from China. *Transp. Res. Part Policy Pract.* **2021**, *150*, 49–62. [\[CrossRef\]](#)
35. Fishman, T.; Myers, R.J.; Rios, O.; Graedel, T. Implications of emerging vehicle technologies on rare earth supply and demand in the United States. *Resources* **2018**, *7*, 9. [\[CrossRef\]](#)
36. Greim, P.; Solomon, A.; Breyer, C. Assessment of lithium criticality in the global energy transition and addressing policy gaps in transportation. *Nat. Commun.* **2020**, *11*, 4570. [\[CrossRef\]](#)
37. Christensen, P.A.; Anderson, P.A.; Harper, G.D.; Lambert, S.M.; Mrozik, W.; Rajaeifar, M.A.; Wise, M.S.; Heidrich, O. Risk management over the life cycle of lithium-ion batteries in electric vehicles. *Renew. Sustain. Energy Rev.* **2021**, *148*, 111240. [\[CrossRef\]](#)
38. Kieckhäfer, K.; Wachter, K.; Spengler, T.S. Analyzing manufacturers' impact on green products' market diffusion—The case of electric vehicles. *J. Clean. Prod.* **2017**, *162*, S11–S25. [\[CrossRef\]](#)
39. Onat, N.C.; Kucukvar, M.; Tatari, O.; Egilmez, G. Integration of system dynamics approach toward deepening and broadening the life cycle sustainability assessment framework: A case for electric vehicles. *Int. J. Life Cycle Assess.* **2016**, *21*, 1009–1034. [\[CrossRef\]](#)
40. Oshiro, K.; Masui, T. Diffusion of low emission vehicles and their impact on CO₂ emission reduction in Japan. *Energy Policy* **2015**, *81*, 215–225. [\[CrossRef\]](#)
41. Zhou, X.; Kuosmanen, T. What drives decarbonization of new passenger cars? *Eur. J. Oper. Res.* **2020**, *284*, 1043–1057. [\[CrossRef\]](#)
42. Zeng, S.; Nan, X.; Liu, C.; Chen, J. The response of the Beijing carbon emissions allowance price (BJC) to macroeconomic and energy price indices. *Energy Policy* **2017**, *106*, 111–121. [\[CrossRef\]](#)
43. Zeng, S.; Jia, J.; Su, B.; Jiang, C.; Zeng, G. The volatility spillover effect of the European Union (EU) carbon financial market. *J. Clean. Prod.* **2021**, *282*, 124394. [\[CrossRef\]](#)
44. Shafiei, E.; Davidsdottir, B.; Leaver, J.; Stefansson, H.; Asgeirsson, E.I. Comparative analysis of hydrogen, biofuels and electricity transitional pathways to sustainable transport in a renewable-based energy system. *Energy* **2015**, *83*, 614–627. [\[CrossRef\]](#)
45. Jochem, P.; Gómez Vilchez, J.J.; Ensslen, A.; Schäuble, J.; Fichtner, W. Methods for forecasting the market penetration of electric drivetrains in the passenger car market. *Transp. Rev.* **2018**, *38*, 322–348. [\[CrossRef\]](#)
46. García-Olivares, A.; Solé, J.; Osychenko, O. Transportation in a 100% renewable energy system. *Energy Convers. Manag.* **2018**, *158*, 266–285.
47. Coffman, M.; Bernstein, P.; Wee, S. Electric vehicles revisited: A review of factors that affect adoption. *Transp. Rev.* **2017**, *37*, 79–93. [\[CrossRef\]](#)
48. Paiho, S.; Saastamoinen, H.; Hakkarainen, E.; Similä, L.; Pasonen, R.; Ikäheimo, J.; Rämä, M.; Tuovinen, M.; Horsmanheimo, S. Increasing flexibility of Finnish energy systems—A review of potential technologies and means. *Sustain. Cities Soc.* **2018**, *43*, 509–523. [\[CrossRef\]](#)
49. Mersky, A.C.; Sprei, F.; Samaras, C.; Qian, Z.S. Effectiveness of incentives on electric vehicle adoption in Norway. *Transp. Res. Part Transp. Environ.* **2016**, *46*, 56–68. [\[CrossRef\]](#)
50. Bjerkan, K.Y.; Nørbech, T.E.; Nordtømme, M.E. Incentives for promoting battery electric vehicle (BEV) adoption in Norway. *Transp. Res. Part Transp. Environ.* **2016**, *43*, 169–180. [\[CrossRef\]](#)
51. Kester, J.; Noel, L.; de Rubens, G.Z.; Sovacool, B.K. Promoting Vehicle to Grid (V2G) in the Nordic region: Expert advice on policy mechanisms for accelerated diffusion. *Energy Policy* **2018**, *116*, 422–432. [\[CrossRef\]](#)
52. Lin, X.; Sovacool, B.K. Inter-niche competition on ice? Socio-technical drivers, benefits and barriers of the electric vehicle transition in Iceland. *Environ. Innov. Soc. Transit.* **2020**, *35*, 1–20. [\[CrossRef\]](#)
53. Dhar, S.; Pathak, M.; Shukla, P.R. Electric vehicles and India's low carbon passenger transport: A long-term co-benefits assessment. *J. Clean. Prod.* **2017**, *146*, 139–148. [\[CrossRef\]](#)
54. Morcillo, J.D.; Franco, C.J.; Angulo, F. Simulation of demand growth scenarios in the Colombian electricity market: An integration of system dynamics and dynamic systems. *Appl. Energy* **2018**, *216*, 504–520. [\[CrossRef\]](#)
55. Morcillo, J.D.; Angulo, F.; Franco, C.J. Simulation and Analysis of Renewable and Nonrenewable Capacity Scenarios under Hybrid Modeling: A Case Study. *Mathematics* **2021**, *9*, 1560. [\[CrossRef\]](#)
56. Sterman, J. *Business Dynamics*; McGraw-Hill, Inc.: New York, NY, USA, 2000.
57. Bektaş, T.; Laporte, G. The pollution-routing problem. *Transp. Res. Part Methodol.* **2011**, *45*, 1232–1250. [\[CrossRef\]](#)
58. Watari, T.; Nansai, K.; Nakajima, K.; McLellan, B.C.; Dominish, E.; Giurco, D. Integrating circular economy strategies with low-carbon scenarios: Lithium use in electric vehicles. *Environ. Sci. Technol.* **2019**, *53*, 11657–11665. [\[CrossRef\]](#)
59. Barlas, Y. Formal aspects of model validity and validation in system dynamics. *Syst. Dyn. Rev. J. Syst. Dyn. Soc.* **1996**, *12*, 183–210. [\[CrossRef\]](#)
60. Kuznetsov, Y.A. *Elements of Applied Bifurcation Theory*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 112.
61. Wiggins, S.; Wiggins, S.; Golubitsky, M. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2.

62. Dong, X.; Zhang, B.; Wang, B.; Wang, Z. Urban households' purchase intentions for pure electric vehicles under subsidy contexts in China: Do cost factors matter? *Transp. Res. Part Policy Pract.* **2020**, *135*, 183–197. [[CrossRef](#)]
63. Graabak, I.; Wu, Q.; Warland, L.; Liu, Z. Optimal planning of the Nordic transmission system with 100% electric vehicle penetration of passenger cars by 2050. *Energy* **2016**, *107*, 648–660. [[CrossRef](#)]
64. Bhuvandas, D.; Gundimeda, H. Welfare impacts of transport fuel price changes on Indian households: An application of LA-AIDS model. *Energy Policy* **2020**, *144*, 111583. [[CrossRef](#)]
65. Kanger, L.; Geels, F.W.; Sovacool, B.; Schot, J. Technological diffusion as a process of societal embedding: Lessons from historical automobile transitions for future electric mobility. *Transp. Res. Part Transp. Environ.* **2019**, *71*, 47–66. [[CrossRef](#)]
66. Redondo, J.M.; Olivar, G.; Ibarra-Vega, D.; Dyner, I. Modeling for the regional integration of electricity markets. *Energy Sustain. Dev.* **2018**, *43*, 100–113. [[CrossRef](#)]

Article

COVID-19's Pandemic Effects on Bike Sharing Systems: A New Reality for Urban Mobility?

Efstathios Bouhouras ¹, Socrates Basbas ^{1,*}, Stamatia Ftergioti ², Evangelos Paschalidis ³
and Harris Siakantaris ⁴

¹ Faculty of Engineering, School of Rural & Surveying Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; stbouh@auth.gr

² Department of Economics, University of Ioannina, P.O. Box 1186, 45110 Ioannina, Greece; Ftergioti.st@gmail.com

³ Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK; E.Paschalidis@leeds.ac.uk

⁴ Cyclopolis, 81 Ifigenias & Velikopoulou St., 14231 Athens, Greece; hs@cyclopolis.gr

* Correspondence: smpasmpa@auth.gr

Abstract: On 11 March 2020, the World Health Organization made the assessment that a new disease (COVID-19) caused by a novel coronavirus (SARS-CoV-2) could be characterized as a pandemic. From that point, a chain reaction of events and difficult decisions requiring action was launched. National governments all over the world announced partial or total quarantine lockdowns in an effort to control the virus' spreading in order to save as many lives as possible. The effects of the pandemic were multifaceted and transport was not excluded. The current paper examines data regarding the level of usage (provided by the administrator) of bike-sharing systems in three case studies/medium-sized Greek cities (Igoumenitsa, Chania, and Rhodes) and through a statistical analysis identifies if there is a correlation between the implemented measures and the modal choice of the residents. The main results and conclusions of this analysis reveal that the level of usage of these specific bike-sharing systems was significantly increased during the lockdown period compared to the situation before the lockdown and the pandemic in general.

Keywords: COVID-19; bike sharing system; urban mobility; regression analysis

Citation: Bouhouras, E.; Basbas, S.; Ftergioti, S.; Paschalidis, E.; Siakantaris, H. COVID-19's Pandemic Effects on Bike Sharing Systems: A New Reality for Urban Mobility? *Appl. Sci.* **2022**, *12*, 1230. <https://doi.org/10.3390/app12031230>

Academic Editors: Jianbo Gao, Giovanni Randazzo, Anselme Muzirafuti and Dimitrios S. Paraforos

Received: 30 November 2021

Accepted: 17 January 2022

Published: 25 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The first case of COVID-19 in Greece was recorded on 26 February 2020. From that day and in less than a month, the Greek government began the implementation of restrictions which led to a general quarantine lockdown, affecting people's mobility, thus creating a new reality for urban mobility (modal shift, mobility restrictions, etc.). It is rather clear that COVID-19 has affected, and continues to affect, people's transport mode choice. According to surveys "consumers are already showing an increased cautiousness about health, meaning that safety is likely to be a more important factor for consumers when deciding if, how and when to move" [1].

Three major European cities during the first months of the pandemic announced measures that allowed bikers and pedestrians to comply with social requirements. Specifically, according to the press, authorities in Milan, Italy, decided to reallocate street space from private cars to cycling and walking, in response to COVID-19. Over 35 km of streets were to be transformed over the summer of 2020, through a rapid experimental citywide expansion of cycling and walking space in order to protect residents. The Strade Aperte plan states: "included low-cost temporary cycle lanes, new and widened pavements, 30kph speed limits and pedestrian and cyclist priority streets" [2].

Another European city adopting measures aiming to promote cycling and walking is Paris, France. According to the press, an investment of over 300 million EUR is underway in the Ile-de-France area for the construction of an extensive network of separated

broad cycle paths and highways. The local authorities highlighted the fact that cycling is one of those transport means that allows users to maintain necessary and appropriate social distancing and therefore should be supported in order to play its role, not only during the COVID-19 crisis but also after its end [3]. In addition, the World Health Organization has urged cities to make full use of cycling and walking. Specifically, on a brochure published by the WHO entitled “Moving around during the COVID-19 outbreak”, it is mentioned that whenever feasible, cycling and walking should be selected since they provide physical distancing while helping people meet the minimum requirement for daily physical activity [4].

Finally, authorities in the city of Brussels, Belgium, decided that the area of Pentagon (city center), the district within the Kleine Ring, will become a residential area at least during the COVID-19 crisis. The measure endured for 3 months, from 11 May 2020, and during that period (a) pedestrians were allowed to use the full width of the public road and not only the pedestrian paths but (b) vehicles were also allowed inside that area with a maximum speed of 20 km/h to ensure pedestrians’ priority and cyclists’ safety and (c) existing pedestrian zones retained their status [5].

Another press release regarding the performance of bike-sharing systems in the United States of America during the first months of the COVID-19 pandemic, highlights the fact that the bike-sharing systems in Boston, Chicago, and New York (all administrated by the same company) were free for healthcare or public transportation workers and first responders. Moreover, the bike-sharing system in Washington D.C. was free of charge to essential workers while the respective system in Minnesota was free for healthcare workers. The Houston bike-sharing system saw increased usage, even after COVID-19 precautions closed some stations to prevent gatherings in parks. Many systems—including Kansas City, Detroit, and Memphis—were temporarily offering unlimited free rides for all users. Another aspect of the COVID-19 pandemic effects on the bike-sharing systems in the United States concerns the fact that “*some systems had developed partnerships with local restaurants, allowing them to use their bikes for food delivery. Of the more than 189 cities with bike sharing systems in the U.S., only a few had shut down during the time of the pandemic*”. However, there were many cases of bike-sharing systems that presented a significant decrease in their level of usage due to the fear of contagion among other reasons (Santa Monica, Central Los Angeles, West Los Angeles, and North Hollywood) [6].

A questionnaire-based survey conducted by the Laboratory of Renewable and Sustainable Energy Systems of the Environmental Engineering Department of Technical University of Crete, Greece, under the framework of the Horizon 2020 CIVITAS DESTINATIONS research project, addressed to the residents of the cities of Chania and Rethemno, Crete, during the period 16–22 March 2020, revealed that the percentage of choosing bicycles as their transport mode during the previously mentioned period compared to the period of January–February 2020, increased by almost 18% [7], embossing the dynamic of the bicycle as a primary transport mode.

The effect of the COVID-19 pandemic on bike-sharing systems is still being monitored and recorded by researchers worldwide. Surveys have been implemented focusing on collecting information regarding individuals’ immediate responses to the travel restrictions during the pandemic in specific areas [8,9] while others have covered larger areas [10,11]. Researchers are focused on four major topics: (a) environmental quality [12–14], (b) socio-economic impacts [15], (c) management and governance, and d) transportation and urban design [16–19] without excluding urban road freight transport [20]. Many of the above-mentioned research activities are focused on the impacts of implemented restriction measures concerning urban mobility in an effort to control the COVID-19 outbreak [8,21–23] by introducing, for example, terms such as “responsible transport” [24] or by applying Big Data from mobile spatial statistics attempting to estimate population density patterns in order to customize applied measures to the characteristics of urban areas [25].

Based on the information and knowledge gained through a literature review on this topic (COVID-19 pandemic and mobility restriction measures), it was decided to develop

a paper regarding BS systems in Greece. The present paper examines the performance of the bike-sharing system in three Greek cities (Igoumenitsa, Rhodes, and Chania) comparing the periods before, during, and after the general lockdown, based on data provided by the operator (Cyclopolis) of these bike-sharing systems. An attempt has been made to identify changes in travel behavior related to the usage of BS bicycles, aiming to assist policymakers and policy takers to better design and implement appropriate and efficient customized policies. Moreover, the authors' intention was not to develop a forecasting model or to explain horizontally how restriction measures affected the performance of bike-sharing systems, but through the findings of this paper to provide a small amount of knowledge to the researchers in the specific field and help them understand and compare travel behaviors and impacts in other cities/countries.

Analysis of the provided data was performed by developing a regression model with aggregated data and thus conclusions could be extracted concerning the performance of the bike-sharing systems across three Greek cities. Based on the first level analysis and its findings, it was decided to develop a regression model using aggregated data in an effort to identify correlations among the parameters describing the BS systems' performance. Since such data were not available at the time the research took place, we were unable to deal with the modal shift in the framework of the present paper. The second section (Materials and Methods) describes the methodology followed for the data collection and analysis while in the third section (Results), the key findings of the descriptive and in-depth statistical analysis are described. Finally, in the Discussion and Conclusions section, the limitations of the study are presented, along with the main conclusions and future research.

2. Materials and Methods

The methodology followed in the current paper was defined mostly by the availability of the collected (by the operator of the selected BS systems) relevant data (number of trips, number of registered users, time period of recorded data, etc.) as well as the examination of the available literature (similar cases, methodologies developed for evaluating the BS systems' performance during the pandemic). Concerning the available data, the contribution of the operator (Cyclopolis) of several BS systems in Greek cities was crucial for the implementation of the research. The examination of the relevant literature/references revealed that the majority of the information came from press publications and that there were not many scientific publications on the issue during the first six months of the pandemic in Europe.

The methodology followed for the development of the current paper can be summarized in the following steps:

1. Examining and analyzing the available literature/references.
2. Communication with the operator of the BS systems (Cyclopolis) requesting available data as well as ensuring the necessary licenses by the authorities of the three Greek municipalities.
3. Examining the provided data in order to decide the analysis methodology, a process which led to the conclusion that based on the characteristics of the provided data the development of a regression model would be the most appropriate. Furthermore, it was decided that in order for the data to refer to the same time period it was necessary to aggregate the data to a weekly time period.
4. Analysis of the provided data. The analysis was implemented in two levels: (a) first-level analysis focusing on those attributes that could better describe the performance of the BS systems in terms of a transport mode and (b) an in-depth regression analysis in an effort to identify if and in which level the COVID-19 pandemic affected the performance of the BS systems. Regression analysis was implemented due to its ability to infer the relationship between independent and dependent variables. As the primary objective of the current paper was to present the effect of the pandemic on BS systems, it was necessary to analyze the provided data in two different time periods.

The first concerned the lockdown period and the second regarded the periods before and after the lockdown. During these periods, independent and dependent variables were set and then examined in order to identify which was the effect of the pandemic.

To evaluate the effect of general quarantine lockdown on the performance of bike-sharing systems in Greek cities, we used data provided by the operator of the bike-sharing systems (Cyclopolis). It must be mentioned that in the cities of Rhodes and Chania, public transport is operational throughout the entire year while in the city of Igoumenitsa it is operational only during summertime (June–August) and for specific routes (mainly connecting the urban areas with the beaches located near the city). Daily data were collected for three Greek cities (Chania, Igoumenitsa, and Rhodes) for a rather long time. Specifically, for the cities of Chania and Rhodes, collected data were available from May 2019 to May 2020 while for the city of Igoumenitsa data were available for the time period July 2019–July 2020. However, in order to analyze the data in the same time period, the number of trips, avoided cars (the number of private cars that will not be used due to the usage of BS bicycles), and avoided CO₂ emissions were examined for the period May 2019–May 2020 by using simple equations. To proxy the performance of the bike-sharing systems, we used the number of rides per week in a given city (a ride refers to the process of an individual renting a bicycle and following a route of their choice before returning it to a rental station, aka a trip). Thus, our dependent variable was the weekly number of rides (Rides) recorded in the city's bike-sharing system.

Due to limitations in data availability, we used a reduced number of explanatory variables. The main variable of interest was a dummy variable (Lockdown 23 March–4 May 2020) which captures the general quarantine lockdown and all the restrictions imposed due to the COVID-19 pandemic in Greece (it should be noted that on 4 May 2020 the Greek government announced the gradual de-escalation of restrictions that lasted until July 2020, including the partial lifting of movement restrictions and not the total lifting of restrictions). This variable takes the value of 1 during the quarantine lockdown period, i.e., the weeks between 23 March and 4 May 2020 and zero otherwise. Additionally, we included an alternative dummy variable (Lockdown 23 March–forward), which takes the value of 1 during the quarantine lockdown period and beyond, i.e., all the weeks after 23 March 2020 even after the easing of restrictions on 4 May, and zero all the weeks before that date. We assumed that quarantine lockdown was positively related to the performance of bike-sharing systems, increasing the number of rides during the quarantine period and beyond. Therefore, we expected a positive sign for both variables, implying that the quarantine lockdown led to an increased use of rental bicycles. Through this approach, an increase in the BS system would be proven to be a conscious choice of the users and not a random incident.

In order to examine the trips' duration, we include the average duration of the rides (Duration) per week, expressed in minutes. Although the trips' duration could be associated with the number of rides, we do not have a priori expectation on the sign of this variable. Finally, taking into account the form of bike-sharing systems and renting infrastructure within each city, the number of rental stations (Stations) was included in the set of regressors. The automatic bike-sharing system of the municipality of Igoumenitsa consists of three rental stations, the municipality of Rhodes has seven rental stations while the municipality of Chania had four rental stations until January 2020 which were reduced to three from February 2020 onwards. Since BSS bicycles were rented and returned by the users to a rental station, we expected a positive effect of the number of available rental stations on the use of BSS bicycles.

Our analysis consists of panel data that contains observations collected at a regular frequency across similar units (i.e., cities in Greece). The use of one panel data model instead of three separate time series models has the advantage of containing more information, variability, and more efficiency than pure time series data. This occurs because panel data models allow for heterogeneity across units (i.e., cities), namely, allowing for any or all the

model parameters to vary across cities whereas time series models assume that the model parameters are common across cities.

Furthermore, the idea of developing three discrete (each for a city examined in the framework of this paper) models was examined but not followed as the restriction measures applied were the same for all of Greece and for the same time period and concerned similar BS systems under the same administrator; thus, it was decided that one model should be developed covering all three cities. Table 1 presents the descriptive statistics and specifically the values of the observations as provided by the administrator. These values were examined and the marginal values were excluded by the model's developing process, although they represent real cases (for example, the minimum value regarding the average duration in minutes is recorded in one of the three BS systems and refers to a case in which the system failed to properly record the duration of the user's trip).

Table 1. Descriptive statistics.

Variable Name	Description	Obs.	Mean	St. Dev.	Min	Max
Number of rides	Weekly number of rides recorded in the bike-sharing system in each city	156	60.54	91.34	1	699
Average duration (min)	Average duration of the rides per week (in minutes)	156	65.64	38.55	0.05	368.67
Rental Stations	Number of rental stations in each city	156	4.51	1.74	3	7
Lockdown (23 March–4 May 2020)	Dummy for general quarantine lockdown, taking a value of one for the weeks between 23 March and 4 May 2020, and zero otherwise	156	0.13	0.34	0	1
Lockdown (23 March–forward)	Dummy for general quarantine lockdown, taking a value of one for all the weeks after 23 March 2020, and zero all the weeks before the 23 March 2020.	156	0.26	0.44	0	1

The final step of the followed methodology concerns the development of the paper and presents the results of the data analysis by most of the effects of the pandemic on the performance of the BS systems based on the users' choices made before, during, and after the lockdown period in the above-mentioned Greek cities.

The bike-sharing system in Igoumenitsa was initially installed and operated in the framework of the ADRIMOB (Sustainable coast MOBility in the ADRIatic area) Project by the Regional Unit of Thesprotia. The project was implemented under the European Territorial Cooperation Programme IPA ADRIATIC 2007–2013 [26]. This is considered the first phase of the bike-sharing system in Igoumenitsa, during which the system was free of charge. The system allowed access to a bike in an automated manner by calculating the actual time of the use of the bicycles. It could accommodate people who were subscribers and made use of the special electronic card EasyBike. The bike-sharing system was operative 24/7 but the working hours of the two contact points were Monday to Friday, 08:00–15:00. The number of registered users was rather small. In 2015, 22 EasyBike cards were subscribed, while in 2016 the number was increased to 25 cards and in 2017 only 1 new card was subscribed. There were two rental stations available equipped with 10 bicycles in total. In the city of Igoumenitsa, there are two bicycle roads constructed and operating. The first one (bicycle and pedestrian road simultaneously, mixed usage) begins near the facilities of the Technological Institute of Epirus at the 28th Oktovriou Str. The bicycle–pedestrian road was constructed in 1998 (total length = 2.8 km and width = 4 m, two-way road) and ends at the road heading to Drepano Beach, which is located 1.5 km from the end of the bicycle–pedestrian road. The users, if they want to continue their trip to Drepano Beach, must use the road constructed for all traffic, which is very dangerous, especially during summertime, when the traffic volumes from and to Drepano Beach are high. The second bicycle road is an urban bicycle road along Leoforos 49 Martiron (location: Ladohori–Igoumenitsa). The total length is 1.1 km, the width of the surface used by the

bicycles is approximately 2 m with pavement on both sides. The road crosses roundabouts and the existing horizontal and vertical signs are considered to be adequate [27].

The second phase of the bike-sharing system in Igoumenitsa concerned the time period during which the SUMPORT project was implemented. Aiming to further promote sustainable mobility, a new rental station was installed near the facilities of the Technological Educational Institute of Epirus. The new station of the new and currently active BS system in Igoumenitsa is presented in Figure 1 [28]. Through this map, the user is informed not only about the location of the rental stations, but by choosing any of the stations, is informed in real time about the capacity of the station (each station can host up to 6 bikes) as well as the number of available bikes and the number of free docking slots.

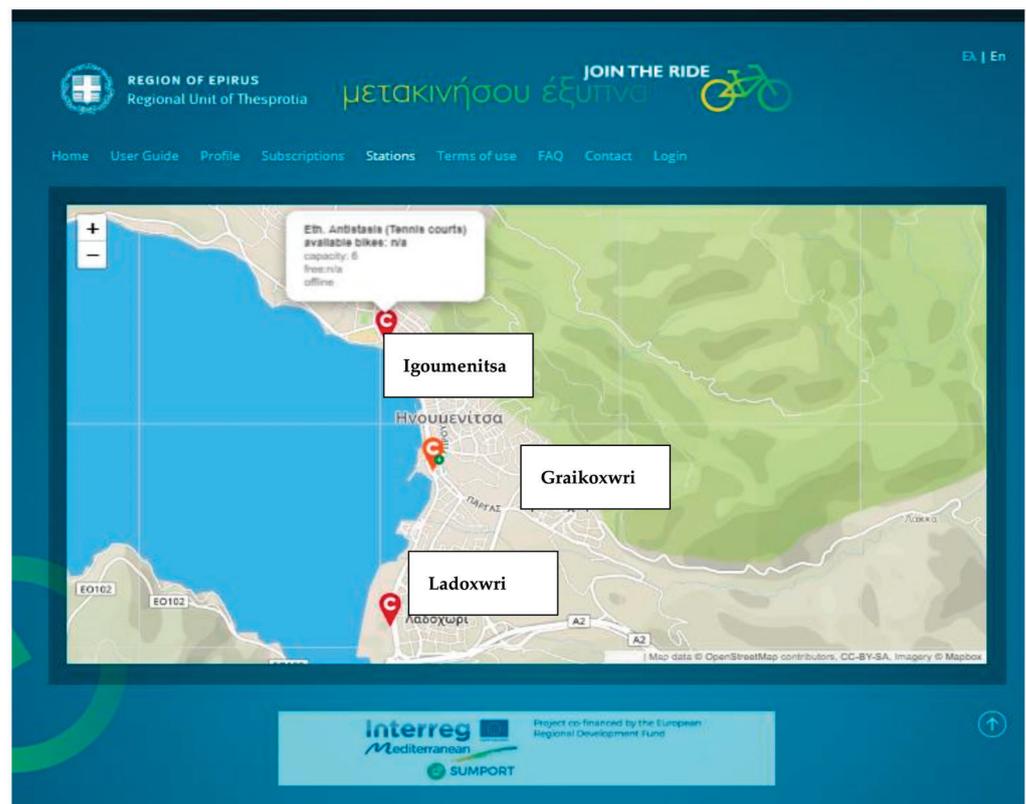


Figure 1. Interactive and real time informative map of the new bike-sharing system in the city of Igoumenitsa [29].

A comparison of the previous and current forms of the bike-sharing system in the city of Igoumenitsa is presented in Table 2 regarding their basic characteristics. The specific comparison highlights that the current system in the city of Igoumenitsa is more organized and user friendly than the previous, encouraging the residents and visitors to use it in an easier way and thus avoiding significant bureaucratic problems. The current system became fully operational just a few months before the COVID-19 pandemic.

Upon securing the assent of the Intermediate Managing Authority of the South Aegean Region for the tendering of the project titled ‘Procurement and installation of bicycle stations in the city of Rhodes’ within the context of the ‘Crete and the Aegean Islands 2017–2013’ Operational Programme, the Municipality of Rhodes conducted an open tender for the procurement and installation of public-use bicycle stations (60 bicycles, 5 bicycling stations, and an operating system) [30]. Subsequently, a relevant contract was concluded between the Municipality of Rhodes and the company Cyclopolis Ltd., London, UK, concerning seven renting stations in total (Aquarium, Marina, Mitropoli, Ag. Nikolaos Square, Eleftheria Square, Symis Square, Ag. Athanasios—Ag. Frangkiskos Gateways). According to the bike-sharing system operating rules, BSS is defined as the automated bike-sharing system

implemented by the municipality of Rhodes, permitting the short-term rental of bicycles. The operation and management of the BSS will be undertaken by the competent services of the municipality and include maintenance of the bicycles and consumables, redistribution of bicycles among stations, promotion of the system, commercial exploitation of the system, and the general operation and management of the system.

Table 2. Comparison of the previous and current forms of the bike-sharing system in the city of Igoumenitsa.

	Previous BSS (ADRI MOB Project)	Current BSS (SUMPORT Project)
Registration process	Required physical presence Free	Automated process through the internet Very low paid once fee (EUR 1)
Operational	A card was acquired to rent a bicycle	A pin is acquired to rent a bicycle
Renting Infrastructure	2 stations	3 stations
Equipment	10 bicycles	18 bicycles
Maximum permitted duration of usage	2 h	7 h
Noncompliance charges/Penalties	Only in case of bicycle's loss	Depends (from 7 days ban to EUR 300 charge in case of bicycle's loss)

The automatic bike-sharing system of Chania's municipality consists of four renting stations (Defkalionos Square, Katehaki Square, Talo Square, and Markopoulou Square) and 70 bicycles in total of which 50 are available for use. The stations consist of info kiosks where the users can borrow a bicycle by using a touch screen [31].

3. Results

3.1. Timeline of COVID-19 Related Events

The timeline of the events related to the COVID-19 pandemic in Greece that led to a general quarantine lockdown is presented in Table 3. The lockdown was announced on 23 March 2020 and endured until 4 May 2020, when the general government announced the beginning of easing restrictions.

Table 3. Timeline and milestones of COVID-19 pandemic in Greece for the period February 2020–May 2020.

Date	Event
26 February	First case of COVID-19 in Greece
27 February	Carnival events in Greece cancelled
10 March	Suspension of educational institutions' operation of all levels nationwide
13 March	Close down cafes, bars, museums, shopping centers, sport facilities and restaurants
16 March	Retail shops closed
22 March	Restrictions on all non-essential movement throughout the country
23 March	Movements were permitted only for 6 categories of reasons
4 April	Extension of restrictions until 27 April
23 April	Extension of restrictions until 4 May
4 May	Begin of easing restrictions
11 May	Remaining retail shops were reopened (except shopping malls)
18 May	All movement restrictions across country were lifted

In order to fully understand the imposed restrictions concerning the mobility of the Greek people, the central government on 23 March 2020 announced that all non-essential movements throughout the country were restricted. Since that date, Greek citizens were allowed to move outside their houses only for the following cases: (i) moving to or from one's workplace during work hours, (ii) going to the pharmacy or visiting a doctor, (iii) going to a food store, (iv) going to the bank for services not possible online, (v) assisting a person in need of help and (vi) going to a major ritual (funeral, marriage, baptism) or movement, for divorced parents, which is essential for contact with their children, and (vii) moving outdoors for exercising or taking one's pet outside, individually or in pairs.

3.2. Exploratory Data Analysis

The analysis of the data provided by the administrator of the BS system in Igoumenitsa for the period July 2019–April 2020/May 2020 (depending on the parameter) is presented in Figures 2–4. Specifically, Figure 2 presents the evolution of the number of yearly subscriptions per month.

The increase in yearly subscriptions is important, especially for the period January 2020–April 2020, which is 68%, while the increase for the period February 2020–April 2020 is 38%. In order to better understand the importance of this increase not only in relation to COVID-19 but also in relation to the weather conditions, it must be noted that the average temperature in Igoumenitsa for January was 14 °C, for February was 15 °C, for March was 16 °C, and finally for April was 19 °C [32]. The average rain precipitation based on historical data for the city of Igoumenitsa is the following: January = 120 mm, February = 125 mm, March = 100 mm, and April = 60 mm [33]. Figure 3 presents the evolution of the total number of registered users for the period July 2019–April 2020.

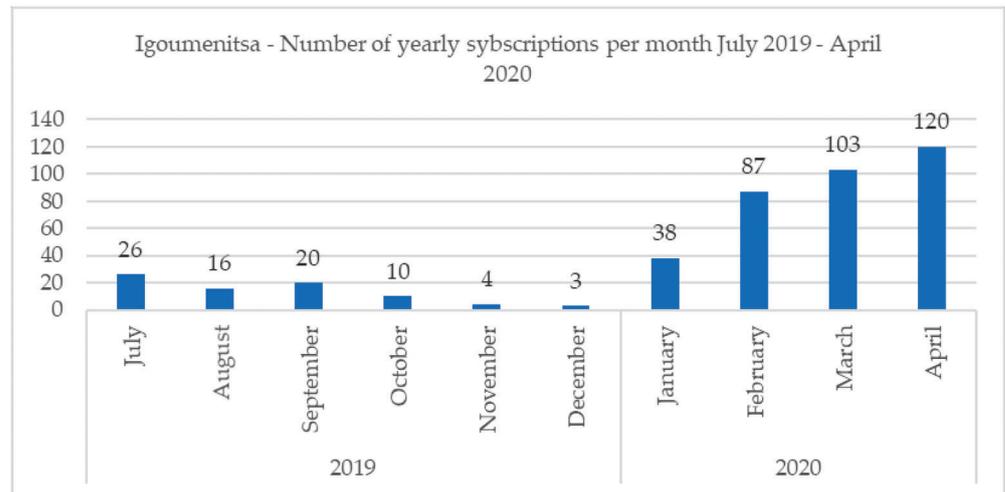


Figure 2. Evolution of the number of yearly subscriptions per month for the time period July 2019–April 2020.

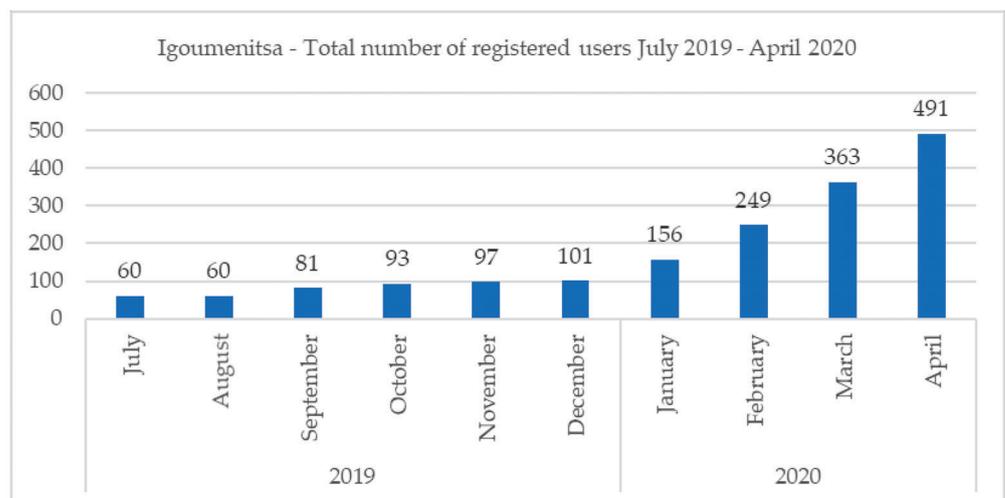


Figure 3. Evolution of the total number of the bike-sharing system’s registered users in the city of Igoumenitsa for the time period July 2019–April 2020.

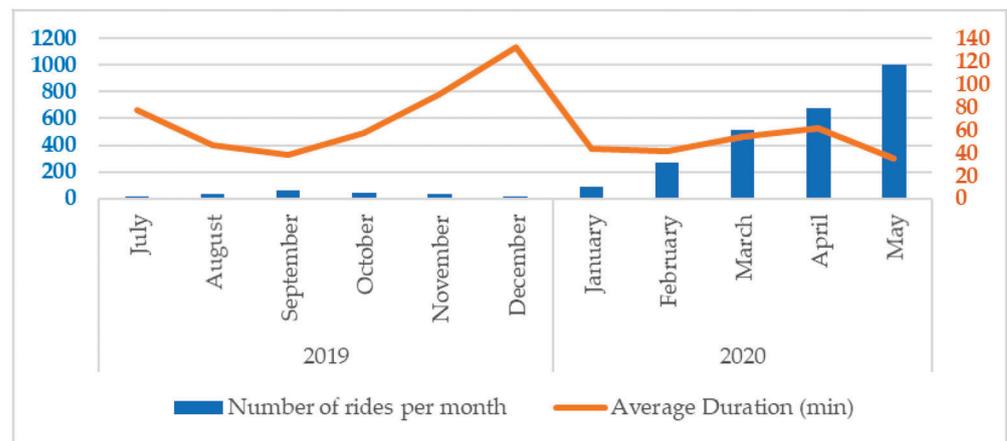


Figure 4. Evolution of Igoumenitsa’s bike-sharing system—number of rides per month and average duration.

It is clear that after the imposition of restriction measures concerning people’s mobility, the number of registered users in the bike-sharing system in Igoumenitsa was significantly increased. Specifically, in the period February 2020–April 2020, the increase was 97% while the respective increase between March 2020 and April 2020 was 35%. Although restrictions were in force on 23 March 2020 concerning the mobility of people, alongside extended police controls, the number of registered users was significantly increased. Moreover, as presented in Figure 4, the number of rides per month doubled during the period February 2020–May 2020. Specifically, the increase (for the above-mentioned time period) was 269%, while between March 2020 and May 2020 was 93%. Figure 4 also presents the average duration of the rides per month. Since the restrictions were terminated, the number of rides during June 2020 was decreased to half compared to May 2020.

Another interesting finding regarding the characteristics of the system’s performance in the above-mentioned three cities regard the summed up (total) number of rides per day, before (1–23 March 2020), and during the lockdown (24 March 2020–4 May 2020) (see Figure 5 and Table 4). The usage of BS bicycles was significantly increased during the lockdown in all cases. It is estimated that the residents of these cities used the BS bicycles not only for leisure trips but also in order to move for their daily needs.

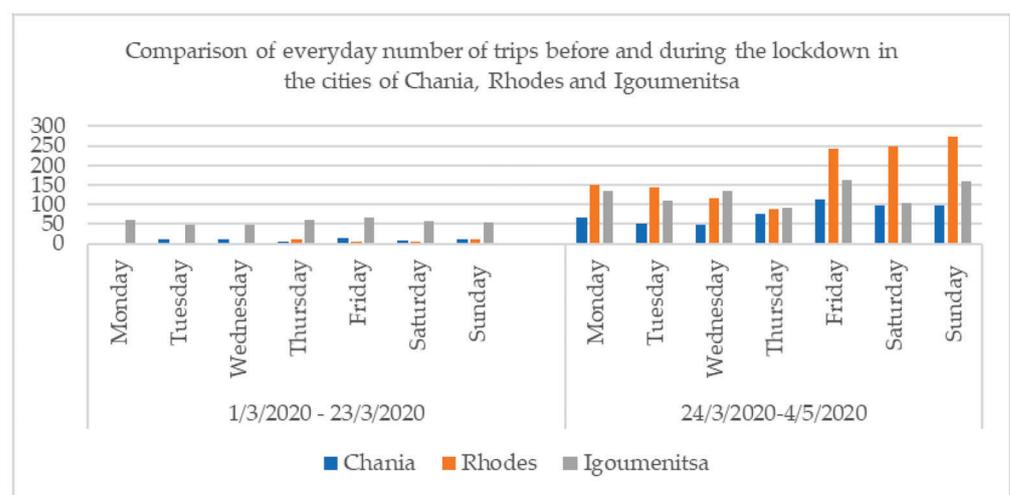


Figure 5. Comparison of summarized (total) number of rides per day before and during the lockdown in the cities of Chania, Rhodes, and Igoumenitsa.

Table 4. Number of rides for Igoumenitsa, Chania, and Rhodes during time periods before and during the lockdown.

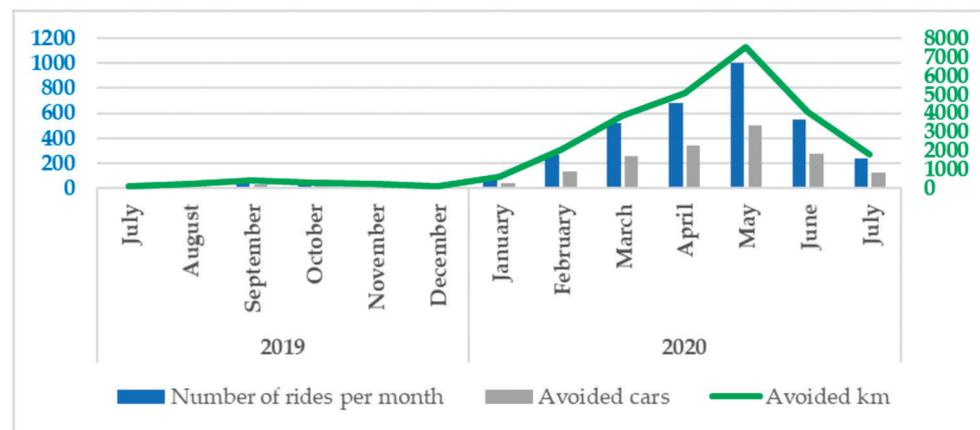
City	Number of Rides						
	March 2020	April 2020	May 2020	1 March 2020–23 March 2020	24 March 2020–30 April 2020	1 May 2020–4 May 2020	
Chania	84	440	507	60	464		86
Rhodes	48	820	1595	38	830		434
Igoumenitsa	517	680	1001	396	801		92

Furthermore, the analysis of the collected data revealed that the majority of the rides during February, March, and April 2020 potentially concerned urban trips as well as leisure purposes, as the bicycles were parked at stations different than those rented [34]. This assumption was concluded by taking into consideration the locations of the stations and the most important landmarks in the examined cities (retail market, touristic and cultural sightseeing, coastal line, etc.). During the implementation of the current bike-sharing system in the city of Igoumenitsa, the operator estimated that over 400 L of fuel was saved due to the modal shift from private cars to bicycles, preventing almost 700 kg of CO₂ from being emitted into the atmosphere.

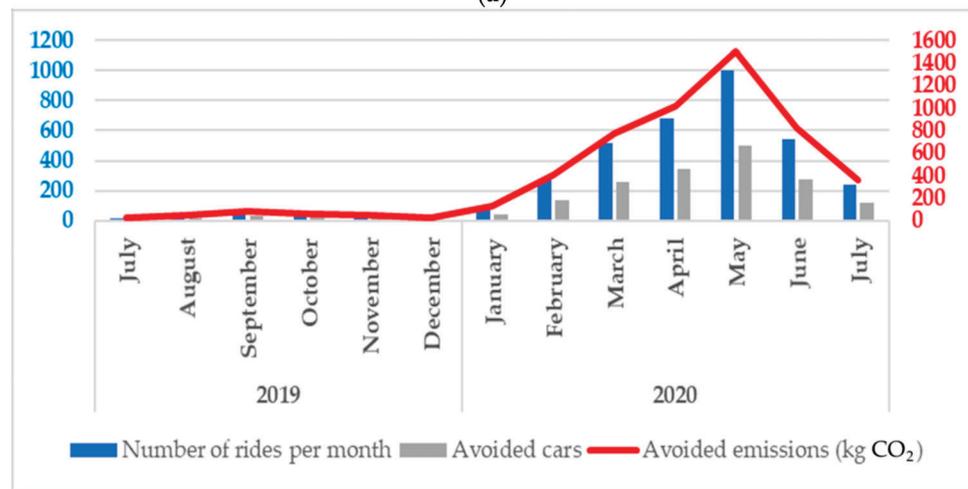
In the framework of the SUMPORT project, key performance indicators (KPIs) were developed in order to be used for monitoring and evaluating the performance of the bike-sharing (and not only) systems [35]. The indicators used for monitoring and evaluating the performance of Igoumenitsa's BS system were the following: (a) The number of rented bikes (provided by the operator), (b) km made by the bicycles (provided by the operator), (c) the number of avoided cars (calculated in-house based on assumptions. Specifically, it is assumed that two rides would be equal to one avoided car), (d) the number of avoided km due to the usage of the BS system (calculated in-house based on assumptions and specifically the average distance covered by a car is equal to 15 km) and (e) the total amount of saved CO₂ (calculated in-house based on assumptions and specifically that each car produces 200 g CO₂ per km). The comparison of the performance of the BS systems in Igoumenitsa, Rhodes, and Chania is based on the above-mentioned indicators. Table 5 presents the performance of Igoumenitsa's BS system using the above-mentioned indicators, while Figure 6a,b presents the performance of Igoumenitsa's system using combo charts for two different sets of variables: (a) The number of rides per month and the number of avoided cars—avoided km and (b) the number of rides per month and the number of avoided cars—avoided emissions (CO₂).

Table 5. Performance of Igoumenitsa's BS system using the SUMPORT key performance indicators (own process).

Year	Month	Total Number of Rides	Average Duration (min)	Km Made by Bikes = (3) × 2	Avoided Cars = (3)/2	Avoided km = (6) × 15	Avoided Emissions (kg CO ₂) = (7) × 0.2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2019	July	15	77	30	8	113	23
	August	30	47	60	15	225	45
	September	58	38	116	29	435	87
	October	41	57	82	21	308	62
	November	34	91	68	17	255	51
	December	16	132	32	8	120	24
2020	January	84	44	168	42	630	126
	February	271	41	542	136	2033	407
	March	517	54	1034	259	3878	776
	April	680	61	1360	340	5100	1020
	May	1001	35	2002	501	7508	1502



(a)



(b)

Figure 6. (a) Combo chart for Igoumenitsa’s BS system relating number of rides per month and avoided cars to number of avoided km. (b) Combo chart for Igoumenitsa’s BS system relating number of rides per month and avoided cars to number of avoided emissions (kg of CO₂).

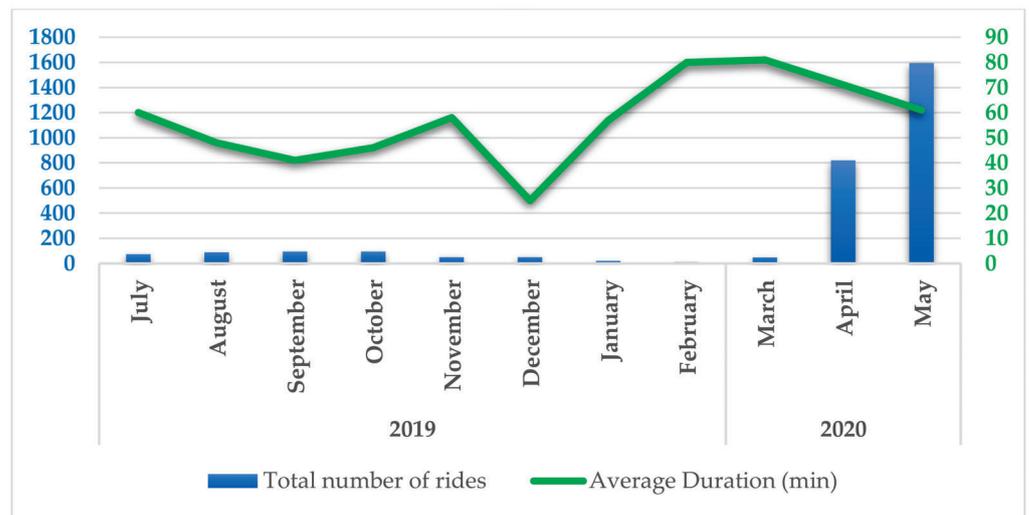
The same indicators were applied for the cases of the cities of Rhodes and Chania. The results are presented in Tables 6 and 7 (see also Figure 7a,b for the city of Rhodes and Figure 8a,b for the city of Chania), respectively.

Table 6. Performance of Rhode’s BS system using the SUMPORT KPIs (own process).

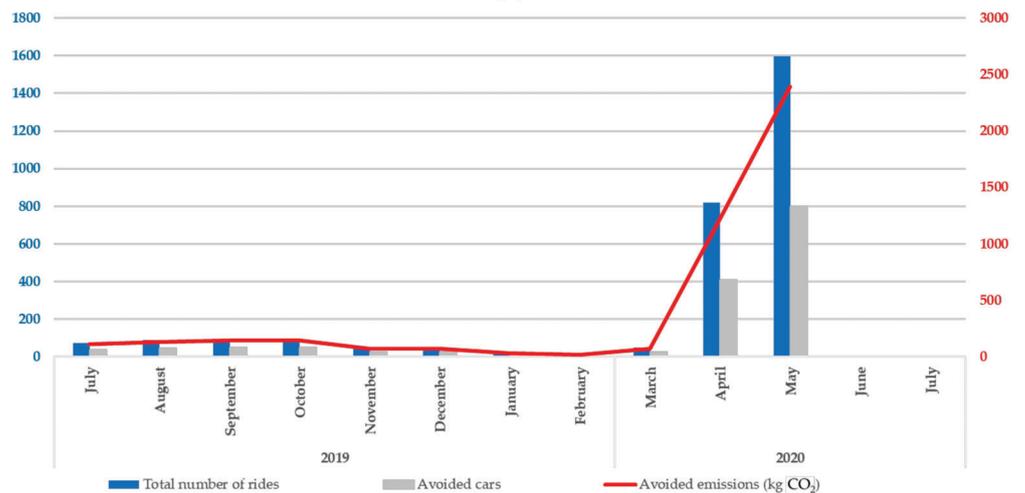
Year	Month	Total Number of Rides	Average Duration (min)	Km made by Bikes = (3) × 2	Avoided Cars = (3)/2	Avoided km = (6) × 15	Avoided Emissions (kg CO ₂) = (7) × 0.2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2019	July	74	60	148	37	555	111
	August	89	48	178	45	668	134
	September	95	41	190	48	713	143
	October	95	46	190	48	713	143
	November	50	58	100	25	375	75
	December	50	25	100	25	375	75
2020	January	21	57	42	11	158	32
	February	12	80	24	6	90	18
	March	48	81	96	24	360	72
	April	820	46	1640	410	6150	1230
	May	1595	41	3190	798	11,963	2393

Table 7. Performance of Chania’s BS system using the SUMPORT KPIs (own process).

Year	Month	Total Number of Rides	Average Duration (min)	Km Made by Bikes = (3) × 2	Avoided Cars = (3)/2	Avoided km = (6) × 15	Avoided emissions (kg CO ₂) = (7) × 0.2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2019	July	289	83	578	145	2168	434
	August	302	85	604	151	2265	453
	September	287	58	574	144	2153	431
	October	191	66	382	96	1433	287
	November	206	75	412	103	1545	309
	December	232	54	464	116	1740	348
2020	January	188	47	376	94	1410	282
	February	89	54	178	45	668	134
	March	84	62	168	42	630	126
	April	440	62	880	220	3300	660
	May	507	68	1014	254	3803	761

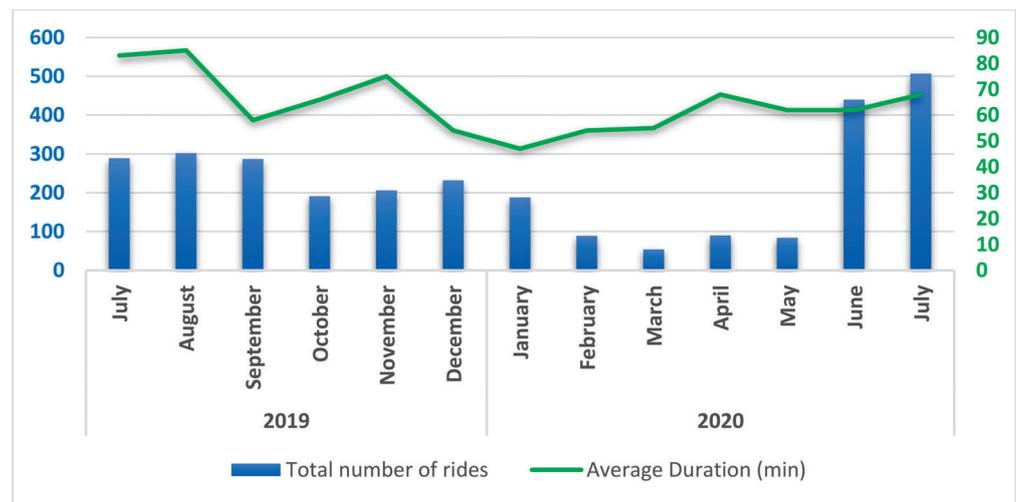


(a)

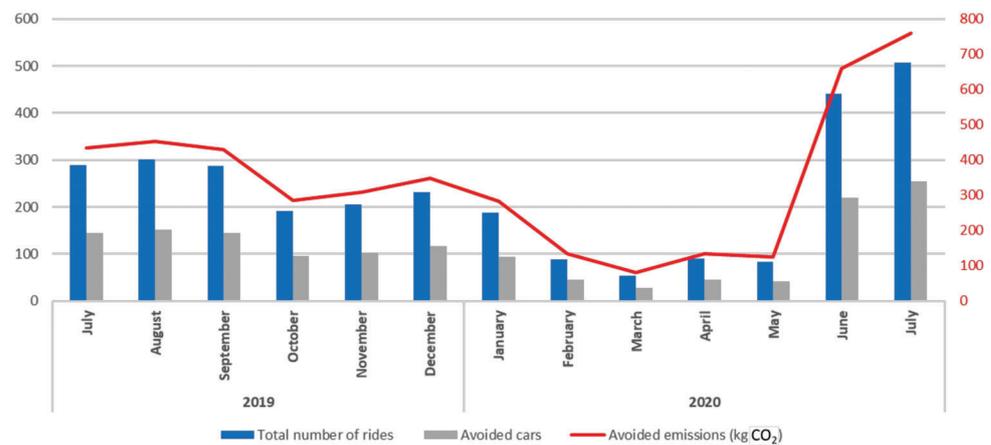


(b)

Figure 7. (a) Combo chart for Rhode’s BS system relating number of rides per month and avoided cars to number of avoided km. (b) Combo chart for Rhode’s BS system relating number of rides per month and avoided cars to number of avoided emissions (kg of CO₂).



(a)



(b)

Figure 8. (a) Combo chart for Chania’s BS system relating number of rides per month and avoided cars to number of avoided km. (b) Combo chart for Chania’s BS system relating number of rides per month and avoided cars to number of avoided emissions (kg of CO₂).

A comparison of the evolution of the total number of rides as well as the average trip’s duration for the period July 2019–May 2020 for the above-mentioned three cities is presented in Figures 9 and 10, respectively. The black dotted frame represents the time period during which the restricting measures were applied.

Another interesting finding from the first-level analysis of the provided data (the second-level analysis or in-depth analysis of the data concerned the exploration of the existence or not of correlations among the parameters of the bike-sharing systems based on the choices of the users) concerns the choice of the users to return or not return the bicycles to the renting stations from which they rented them. It can be assumed that this users’ choice can assist in building an origin–destination matrix. An origin point is considered the renting station from which the user rents the bicycle and a destination point refers to the renting station to which the user returns the bicycle. The analysis of the collected data in an aggregated manner is presented in Figure 11. Specifically, for the city of Chania, two approaches were followed and are presented regarding the Talo renting station, as it was operational only for the year 2019. Therefore, the first approach takes into consideration this fact while in the second the station is totally ignored and excluded from the analysis.

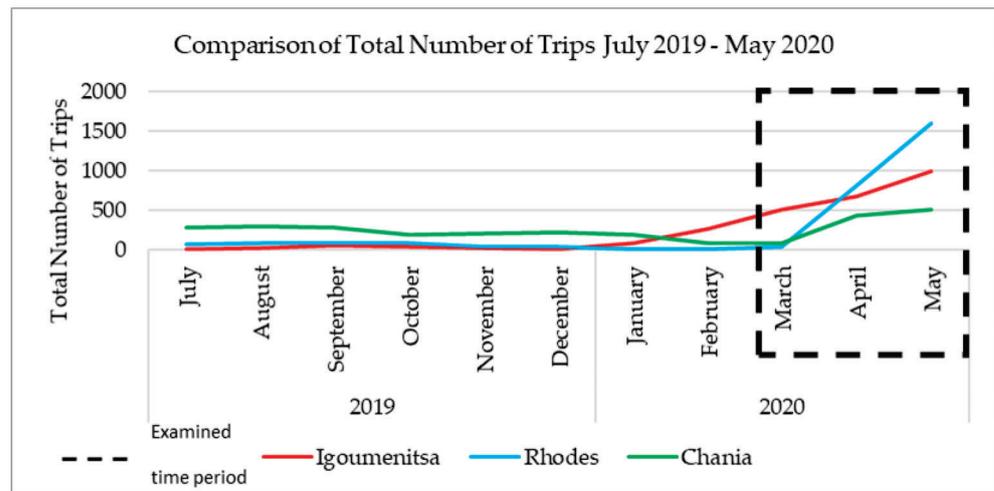


Figure 9. Comparison of the total number of trips for the cities of Igoumenitsa, Rhodes, and Chania during the period July 2019–May 2020.

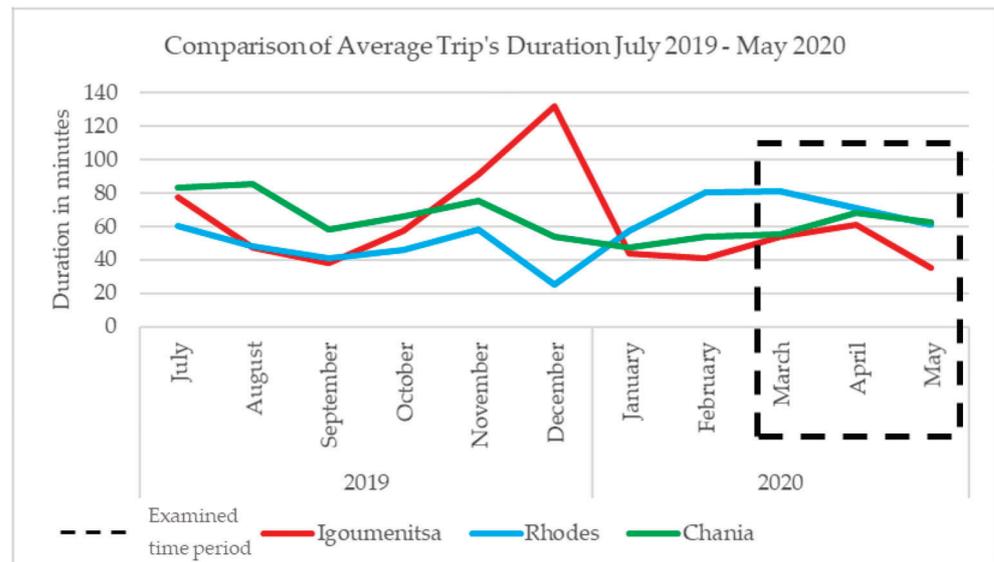


Figure 10. Comparison of the average trip's duration for the cities of Igoumenitsa, Rhodes, and Chania during the period July 2019–May 2020.

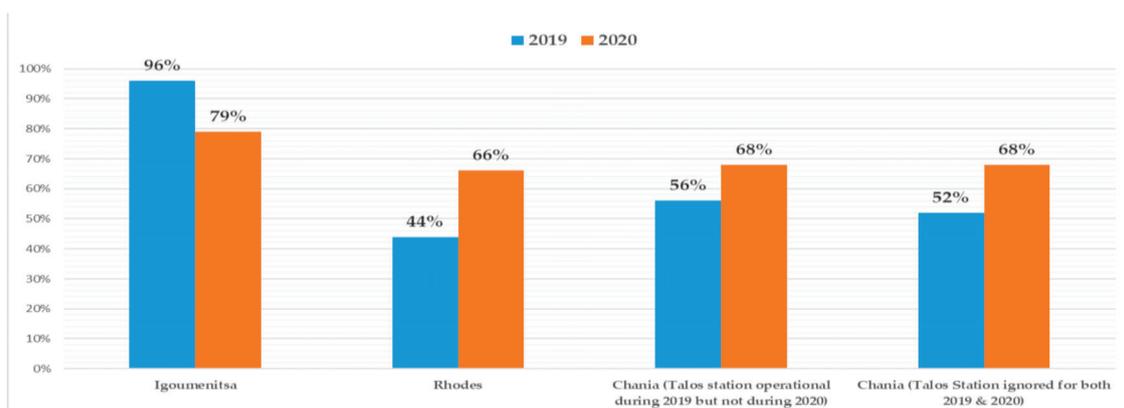


Figure 11. Percentage of BSS bicycles returned by the users to their initially rented station for the cities of Igoumenitsa, Rhodes, and Chania for the period July 2019–May 2020.

3.3. Number of Rides Regression Model

For the purpose of our analysis, daily data provided by the operator (Cyclopolis) of the BS systems were converted into weekly data, which were then used to construct an unbalanced panel dataset of the bike-sharing system in three Greek cities (Igoumenitsa, Rhodes, and Chania) from July 2019 to May 2020.

To analyze the performance of bike-sharing systems we used as a dependent variable the number of rides recorded per week. This variable measures the level of usage of bike-sharing systems in each of the aforementioned Greek cities. In addition, we used a set of independent variables that can be associated with the number of rides. In many cases, the development of a model is highly affected and dependent on the attributes of the collected data. The model developed in the framework of the present research activity aimed to identify whether the lockdown period and the restriction measures applied affected the performance of the BS systems in three Greek cities which can be measured by how ridership was changed before, during, and after the lockdown. For this to be achieved, the number of rides, their duration, and whether the users returned the bicycles to the stations from which they rented them or not, based on the following assumptions, were used:

- The duration could reveal if the users rented the bicycles for pleasure or (a) in order to move to/from their working locations and (b) other than the previous reasons.
- The analysis of the data concerning the renting and returning stations of the bicycles was another factor that could reveal the purpose of using the bicycle as most likely the users that rented a bicycle for pleasure would return it to the same station; while the users renting it in order to move for other reasons would return to a station which is located nearest to their destination.

Firstly, to control the effect of quarantine lockdown, we include a dummy variable that takes the value of 1 during the quarantine period, i.e., the weeks between 23 March and 4 May 2020 and zero otherwise. Moreover, we include an alternative dummy variable that takes the value of 1 during the quarantine period and beyond, i.e., all the weeks after 23 March 2020 and zero otherwise. We do so, in order to control whether the effect of quarantine lockdown on the performance of bike-sharing systems remains after the easing of restrictions on 4 May. Secondly, we include the average duration of the rides per week in each city. Finally, taking into account the availability of bike-sharing systems, the number of rental stations in each city is also included.

Accordingly, the empirical model is formulated as follows:

$$Y_{it} = b_0 + b_1x_{1it} + b_2x_{2it} + b_3x_{3it} + u_{it} \text{ with } u_{it} = \gamma_i + \varepsilon_{it}$$

where,

- x_{1it} : is the dummy for the general quarantine lockdown in Greece;
- x_{2it} : is the weekly average ride duration in each city;
- x_{3it} : is the number of rental stations;
- u_{it} : is the composite error term;
- γ_i : is a random effect, i.e., time-invariant characteristics of a city;
- ε_{it} : is the error term;
- b_0, b_1, b_2, b_3 : parameters to be estimated.

In random-effects models, the random effect is a component of the composite error term, i.e., $u_{it} = \gamma_i + \varepsilon_{it}$, and is not correlated with any regressor. Both random effect and errors are independently identically distributed, $\gamma_i \sim IID(0, \sigma_\gamma^2)$ and $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$. Our random-effects model is estimated by ordinary least square (OLS) regressions with robust clustered standard errors in order to control for both heteroscedasticity and correlation of the error terms [36]. To decide between the use of fixed effects or random effects, we applied the standard Hausman test which showed that the appropriate specification is the random-effects model.

The main results are presented in Table 8. In the first column, we estimate a random-effects model including only our main variable of interest, namely Lockdown (x_1). Then, in columns (2) and (3), we estimate random-effects models including the remaining independent variables (x_1, x_2). In all three columns, the estimated coefficient of lockdown is positive and statistically significant, indicating a positive effect of quarantine lockdown on the number of rides. Specifically, the estimated coefficient suggests that lockdown leads to an increase in the number of rides via bike-sharing systems, during the the lockdown period. It must be noted that for these calculations, rides with a duration of less than 5 min were excluded.

Table 8. The effect of lockdown on number of rides, comparing lockdown period and non-lockdown period, excluding rides with duration less than 5 min.

	(1)	(2)	(3)	(4)
	Random Effects	Random Effects	Random Effects	Fixed Effects
Lockdown (23 March–4 May 2020)	101.653 ** (35.095)	101.653 ** (35.07)	101.62 ** (35.035)	108.034 * (33.22)
Average duration (min)		0.031 (0.18)	0.03 (0.187)	0.027 (0.181)
Rental stations			−0.175 (2.63)	25.152 (13.695)
Constant	47.157 ** (4.759)	45.121 ** (16.222)	45.946 * (27.697)	−68.587 (66.629)
Number of cities	3	3	3	3
Observations	155	155	155	155
R-squared (overall)	0.146	0.146	0.146	0.057

Notes: All regressions include random effects and are estimated with robust clustered standard errors. Lockdown dummy takes the value of 1 during quarantine period, i.e., the weeks between 23 March and 4 May 2020 and zero otherwise. Standard errors in parentheses. *, ** denote statistical significance at 10% and 1%.

Concerning the rest of the independent variables, the coefficient of the weekly average duration of rides is insignificant in all regressions, implying that the duration of rides does not affect the number of rides. Similarly, the rental stations variable exerts an insignificant coefficient, indicating that the number of rental stations is not associated with the weekly number of rides. In column (4), we estimate the baseline model using fixed effects instead of random effects in order to see whether the estimated results depend on model specification. As can be seen, estimated results for the main independent variable, i.e., Lockdown (x_1), hold across both random and fixed effects models.

In Table 9, we re-estimated all regressions of Table 7 using an alternative dummy for lockdown which measures the effect of lockdown during and after the lockdown period. Yet again, rides with a duration of less than 5 min were excluded by these calculations. As can be seen, the results remain the same since the estimated coefficient of Lockdown is still positive and statistically significant in all regressions. Hence, our results can have an alternative interpretation: Lockdown leads to an increase in the number of rides both during the lockdown period and after the easing of restrictions. Regarding the rest independent variables, estimated results for the average duration of rides remain unchanged. On the contrary, the number of rental stations exerts a positive and statistically significant coefficient, indicating that it is positively associated with the number of rides.

Table 9. The effect of lockdown on number of rides, comparing period before and after lockdown, excluding rides with duration less than 5 min.

	(1)	(2)	(3)	(4)
	Random Effects	Random Effects	Random Effects	Fixed Effects
Lockdown (23 March–forward)	120.21 ** (40.521)	120.482 ** (40.477)	122.703 ** (40.355)	140.415 * (36.033)
Average duration (min)		0.083 (0.152)	0.095 (0.162)	0.084 (0.140)
Rental stations			3.787 ** (0.802)	71.298 * (23.009)
Constant	29.132 ** (9.169)	23.562 * (13.05)	5.142 (18.932)	−302.387 (118.304)
Number of cities	3	3	3	3
Observations	155	155	155	155
R-squared (overall)	0.338	0.339	0.344	0.069

Notes: All regressions include random effects and are estimated with robust clustered standard errors. Lockdown dummy corresponds to the alternative dummy which takes the value of 1 during quarantine period and beyond, i.e., all the weeks after the 23 March 2020 and zero otherwise. Standard errors in parentheses. *, ** denote statistical significance at 10% and 1%.

4. Discussion

The statistical analysis of the data provided by the BS systems' operator revealed that the lockdown affected the modal choice of the residents of the three above-mentioned and analyzed Greek cities. It seems that the users of the BS systems in Igoumenitsa, Rhodes, and Chania acknowledged the benefits of using bike-sharing bicycles and as a result continued to use the bicycles even after the end of the lockdown period, although no campaigns were implemented or measures were taken as those in other European cities aiming to promote the usage of BS systems. Even though a decrease was recorded after the end of the restriction measures, the BS systems continued to have an increased level of usage six or nine months ago. The examination of qualitative and quantitative data for a longer time period, regarding the performance of the BS systems, could reveal if the users developed a new pattern concerning their habits and modal choice or if the effect of the lockdown period leading to an increased level of BS systems usage, fades out as time passes. Toward this direction, our team has addressed an official request to the BS systems operator in order to provide us with data covering a longer time period mainly after the end of the first lockdown period.

Similar results to our research were found and conclusions were extracted for the cities of New York [37,38], Budapest [39], Thessaloniki [40,41], Nanjing [42], and the region of Sicily in Italy [43], regarding the level of ridership before, during, and after the imposition of restricting measures concerning urban mobility and specifically the usage of bicycles, the rides' duration [44], the people's perception towards BS systems due to the COVID-19 pandemic, and to bicycles in general [45,46]. Furthermore, during the early stages of the pandemic, in some cases, bike sharing expanded their memberships or improved accessibility for specific working groups [47]. However, there are several cases in which despite an initial increase in using BS systems during the first months of the pandemic, ridership declined for several months in 2020 [48–53]. There are also systems recovering from the effects of the pandemic and mainly lockdown measures applied in many countries; for example, in the UK, the latest data show a significant recovery [54]. A study estimating the effect of the pandemic on the London BS system over the period March–December 2020 indicated that although a reduction in cycle hires was recorded in March and April 2020, the demand increased after May 2020 and even more during April, May, and June 2020 the bikes were hired for longer trips (perhaps as the authors mention due to a shift from public transit) [54]. It is clear that people have reacted differently worldwide concerning the usage of BS systems during the pandemic, making each case unique. For example, in Seoul, the level of usage of the BS system was affected negatively by the number of daily new COVID-19 cases and positively as the necessity for social distancing was becoming a reality for the residents [34]. Moreover, a study regarding the impact of

COVID-19 on the BS system in Slovakia, showed that during the lockdown period, the level of usage was decreased. However, after restrictions were relieved, a slight increase was recorded [55]. In a recently published study regarding the performance of BS systems in the United States [56], one of the main conclusions (five out of eleven BS systems provided feedback) was that bike-sharing moderate-frequency riders (1–2 times per month) may increase after the coronavirus restrictions are lifted.

Based on the above-described analysis, we can state with relative safety that the lockdown period affected the performance of the BS systems in the examined cities since during the lockdown period the level of usage of BS systems increased. This evolution can be capitalized by the local authorities by promoting and advertising the benefits of the BS systems, even those that were not initially (and most likely could not be) conceived, but ultimately were revealed during the COVID-19 pandemic [57–59]. Specially designed campaigns could be organized by the local authorities in order to inform the people of the benefits of using bicycles, by highlighting that BS systems' level of usage was increased during the lockdown period providing them a way out from the restriction measures while at the same time maintaining social distancing. Lastly, those campaigns could promote the usage of BS systems as a significant tool to achieve sustainable mobility [56].

Our study has some limitations. The data provided cover a time period of many months before the applied first lockdown in Greece and only a few months after the easiness of the applied measures. Although the change in the level of BS bicycles' usage was recorded, analyzed, and examined, a longer time period would provide more solid conclusions and results. Furthermore, our analysis is based on quantitative data for the BS systems and at the same time, we do not have access to data concerning other available transport modes in the examined cities. As a result, no comparison among the available transport modes can be made as, for example, public transport and BS bicycles. Finally, the absence of qualitative data regarding the reasons for users' choice to use more (or at least at the same level) BS bicycles during and after the lockdown period compared to the pre-lockdown period, does not allow us to perform an in-depth analysis.

As the pandemic continues, it has been proposed that BS systems continue to be constantly monitored by the operators and despite their level of usage, the users' choices (either these lead to an increase or decrease in their usage) should be investigated through rolling questionnaire-based surveys. As health protocols demand social distancing, it is difficult to perform in situ surveys and therefore it is recommended to exploit digital services such as the internet and smartphones in order to implement these surveys. Furthermore, it is crucial for the sustainability of the BS systems that operators fully comply with the health protocols applied which require decontamination of bicycles and all surfaces in the rental stations ensuring that BS bicycles do not become contagion sources.

If the values presented in Table 10 and specifically the number of rides per 1000 residents are expressed on a daily basis (30 days per month), then for the case of Igoumenitsa, the respective values are equal to 1.88 for March 2020, 2.48 for April 2020, and 3.65 for May 2020. For the case of Chania, the respective values are equal to 0.03 for March 2020, 0.17 for April 2020, and 0.20 for May 2020 while for the case of Rhodes are equal to 0.03 for March 2020, 0.55 for April 2020, and 1.07 for May 2020. When comparing the highest values of the number of rides per 1000 residents (May 2020) with the respective values for other BS systems, Igoumenitsa outperformed the BS system of Mexico City, Chania's BS system performed similarly to Seattle's BS system, and Rhodes' BS system performed similarly to London's BS system based on the findings of a report published in 2018 [60].

Table 10. Effect of lockdown on number of rides per 1000 residents in the cases of Igoumenitsa, Chania, and Rhodes for the period March 2020–May 2020.

City	Number of BS Rental Stations	Number of BS Bicycles for Usage	Population (2011 National Census)	Number of BS Bicycles Available per 1000 Residents	Number of Rides per 1000 Resident		
					March 2020	April 2020	May 2020
Igoumenitsa	3	18	9145 [61]	2.0	56.53	74.36	109.46
Chania	4 (3)	50	84,527 [62]	0.6	0.99	5.20	5.99
Rhodes	7	60	49,500 [63]	1.2	0.97	16.56	32.22

5. Conclusions

Based on the presented data, during the first lockdown period in Greece, the residents of the above-mentioned Greek cities, for unknown (up to present) but assumable reasons, decided to use the installed and operational BS systems in their home cities. Although the increase had started a short time period before the lockdown period, it came to a peak during the lockdown. After the end of the lockdown, the residents tended to return to their prior moving patterns with an increased share for the BS systems compared to the pre-lockdown period. The average duration of the BS bicycles usage during the lockdown period was increased compared to the months before it (Igoumenitsa, Chania) and the months after it (Igoumenitsa, Chania, Rhodes) in 2020, despite the fact that during the lockdown the residents were allowed to move only for a period of one hour during daytime if permission was granted to them and for a small number of specific reasons.

It is necessary to better understand the reasons for the (recorded and presented in this paper) performance of the BS systems in the three Greek cities before, during, and right after the first lockdown period in Greece, to perform a questionnaire-based survey addressed to the users asking them for the reasons that led them to use the BS systems during the lockdown. Furthermore, a request has been planned for submission to the operator of the examined BS systems to provide our team with more data covering mainly the time period after the lockdown period, which will allow us to compare the evolution of the BS systems' performance for a longer time period.

An interesting finding of the above-presented analysis concerns the evolution of the total number of rides after the lockdown. While in the city of Igoumenitsa the total number decreased (but yet remained higher with the respective numbers before the implementation of the restriction measures during the lockdown), in the cities of Rhodes and Chania the total number of rides kept rising. However, based on Table 10, it is clear that the average daily rides in Igoumenitsa in May 2020, which was the most significant among the three cities, were about 30. Considering there is a population of over 9000 in Igoumenitsa, this amounted to about 0.03% of people who opted to use the bikes in the BS systems in the city. While the effect was statistically significant, in reality, the effect was considered very minor.

The airport at Chania served flights from 1 July 2020 and beyond due to the restriction measures applied worldwide [64]; therefore, the increase was not caused by visitors. In the city of Rhodes, the decrease in arrived flights at the island's airport compared to those in 2019 for the period April 2020–June 2020 was 100% [65]. Moreover, in this case, the increase in the BS system was not caused by visitors. In order to understand the reasons for this phenomenon, as mentioned above, a questionnaire-based survey addressed to the users of the BS systems operating in the currently studied cities is under development.

The current situation (COVID-19 pandemic and the measures applied worldwide) is unique for modern human history in terms of the geographical area and the duration of the applied restricting measures/lockdown. Estimations and assumptions can be made in order to explain the phenomenon and identify the effects (if identified) of the lockdown on the users' modal choice and modal shift. However, during the first lockdown period in Greece, the residents of three Greek cities in which BS systems were installed and are currently operating decided to use them more than the previous time period. This fact is

indisputable; however, the reasons that led to this evolution can be only assumed based on the availability of qualitative data for that time period.

The first and more solid assumption is that the increase in the demand for the BS systems was a countermeasure to the restriction measures implemented due to the COVID-19 pandemic as people chose BS bicycles as a safe transport mode ensuring social distancing, freedom of their movements, and environmental benefits. Specifically, the possible benefits identified by the usage of BS bicycles should be capitalized by local decision-makers based on the fact that sustainable mobility has been, is, and will keep being, one of the most important objectives of the European Commission [66–69]. Moreover, there is a high possibility that users that did not own a bicycle and therefore could not exercise (one of the reasons that the residents could go outside for a period of one hour was to exercise), chose the bike-sharing systems as an excuse to go outside and exercise in an effort to decrease the pressure (mainly psychological) created by the lockdown (such a choice was given by the Greek government using a specially developed mobile service) or a transport mode in order to move across those cities covering and serving their needs.

Author Contributions: E.B. Conceptualization, methodology, validation, investigation, and data curation, writing—original draft preparation, writing—review and editing; S.B. validation, writing—review and editing; S.F. formal analysis, writing—original draft preparation; E.P. validation, writing—review and editing; H.S. validation, investigation and data curation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the authorities of the Regional Unit of Thesprotia for providing the data, as well as the permission for using them, related to the SUMPORT Project, the authorities of the municipalities of Rhodes and Chania for the permission of using the bike-sharing systems' data and finally Cyclopolis for providing the necessary data regarding the performance of the bike-sharing systems in the Greek cities of Igoumenitsa, Rhodes, and Chania.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Desomer, E.; Walker, A. Future of Mobility COVID-19: Mobility in Belgium, Article on Deloitte's Brochure. 2020. Available online: https://www2.deloitte.com/content/dam/Deloitte/be/Documents/strategy/Mobility_Covid%20report_0805.pdf (accessed on 26 May 2020).
- GMG The Guardian, ©2020 Guardian News & Media Limited, International Edition. Available online: <https://www.theguardian.com/world/2020/apr/21/milan-seeks-to-prevent-post-crisis-return-of-traffic-pollution> (accessed on 26 May 2020).
- DPG Media, NV. Available online: <https://www.hln.be/nieuws/buitenland/parijse-regio-investeert-versneld-300-miljoen-euro-voor-650-kilometer-aan-bredere-coronafietspaden~{ad972cd1/> (accessed on 26 May 2020).
- World Health Organization. Regional Office for Europe. Available online: <https://who.canto.global/v/coronavirus/s/MFSQ0?viewIndex=1> (accessed on 26 May 2020).
- City of Brussels. Available online: <https://www.brussels.be/residential-area> (accessed on 26 May 2020).
- Streetblog, LA. Available online: <https://la.streetsblog.org/2020/04/14/1-a-county-bike-share-systems-varied-responses-to-covid-19/> (accessed on 26 May 2020).
- Hellenic Institute of Transportation Engineers (HITE). Special Issue 212, December 2019–February 2020. pp. 10–15. Available online: <http://www.ses.gr/wp-content/uploads/2020/05/SES-212.pdf> (accessed on 26 May 2020).
- Fatmi, M.R. COVID-19 impact on urban mobility. *J. Urban Manag.* **2020**, *9*, 270–275. [CrossRef]
- Aloi, A.; Alonso, B.; Benavente, J.; Cordera, R.; Echaniz, E.; González, F.; Ladisa, C.; Lezama-Romanelli, R.L.; Lopez-Parra, A.; Mazzei, V.; et al. Effects of the COVID-19 Lockdown on Urban Mobility: Empirical Evidence from the City of Santander (Spain). *Sustainability* **2020**, *12*, 3870. [CrossRef]
- Lozzi, G.; Rodrigues, M.; Marcucci, E.; Teoh, T.; Gatta, V.; Pacelli, V. Research for TRAN Committee—COVID-19 and Urban Mobility: Impacts and Perspectives; PE625.213; European Parliament, Policy Department for Structural and Cohesion Policies, Brussels, European Union. 2020. Available online: [https://www.europarl.europa.eu/thinktank/en/document/IPOL_IDA\(2020\)652213](https://www.europarl.europa.eu/thinktank/en/document/IPOL_IDA(2020)652213) (accessed on 18 January 2022).
- Guan, L.; Prieur, C.; Zhang, L.; Prieur, C.; Georges, D.; Bellemain, P. Transport effect of COVID-19 pandemic in France. *Annu. Rev. Control* **2020**, *50*, 394–408. [CrossRef]
- Chen, Z.; Hao, X.; Zhang, X.; Chen, F. Have traffic restrictions improved air quality? A shock from COVID-19. *J. Clean. Prod.* **2021**, *279*, 123622. [CrossRef]

13. Ropkins, K.; Tate, J.E. Early observations on the impact of the COVID-19 lockdown on air quality trends across the UK. *Sci. Total Environ.* **2021**, *754*, 142374. [CrossRef] [PubMed]
14. Otmani, A.; Benchrif, A.; Tahri, M.; Bounakhla, M.; Chakir, E.M.; Bouch, M.E.; Krombi, M. Impact of COVID-19 lockdown on PM10, SO2 and NO2 concentrations in Salé City (Morocco). *Sci. Total Environ.* **2020**, *735*, 139541. [CrossRef]
15. Politis, I.; Georgiadis, G.; Nikolaidou, A.; Kopsacheilis, A.; Fyrogenis, I.; Sdoukopoulos, A.; Verani, E.; Papadopoulos, E. Mapping travel behavior changes during the COVID-19 lock-down: A socioeconomic analysis in Greece. *Eur. Transp. Res. Rev.* **2021**, *13*, 1–19. Available online: <https://etrr.springeropen.com/articles/10.1186/s12544-021-00481-7> (accessed on 12 September 2021). [CrossRef]
16. Sharifi, A.; Khavarian-Garmsir, A.R. The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. *Sci. Total Environ.* **2020**, *749*, 142391. [CrossRef]
17. Kanda, W.; Kivimaa, P. What opportunities could the COVID-19 outbreak offer for sustainability transitions research on electricity and mobility. *Energy Res. Soc. Sci.* **2020**, *68*, 101666. [CrossRef]
18. Shamsheeripour, A.; Rahimi, E.; Shabanpour, R.; Mohammadian, A. How is COVID-19 reshaping activity-travel behavior? Evidence from a comprehensive survey in Chicago. *Transp. Res. Interdiscip. Perspect.* **2020**, *7*, 100216. [CrossRef]
19. Mogaji, E. Impact of COVID-19 on transportation in Lagos, Nigeria. *Transp. Res. Interdiscip. Perspect.* **2020**, *6*, 100154. [CrossRef]
20. Loske, D. The impact of COVID-19 on transport volume and freight capacity dynamics: An empirical analysis in German food retail logistics. *Transp. Res. Interdiscip. Perspect.* **2020**, *6*, 100165. [CrossRef]
21. Zhou, Y.; Xu, R.; Hu, D.; Yue, Y.; Li, Q.; Xia, J. Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: A modelling study using mobile phone data. *Lancet Digit. Health* **2020**, *2*, 417–424. [CrossRef]
22. Beck, M.J.; Hensher, D.A.; Wei, E. Slowly coming out of COVID-19 restrictions in Australia: Implications for working from home and commuting trips by car and public transport. *J. Transp. Geogr.* **2020**, *88*, 102846. [CrossRef]
23. Beck, M.J.; Hensher, D.A. Insights into the impact of COVID-19 on household travel and activities in Australia—The early days of easing restrictions. *Transp. Policy* **2020**, *99*, 95–119. [CrossRef]
24. Budd, L.; Ison, S. Responsible Transport: A post-COVID agenda for transport policy and practice. *Transp. Res. Interdiscip. Perspect.* **2020**, *6*, 100151. [CrossRef] [PubMed]
25. Arimura, M.; Ha, T.V.; Okumura, K.; Asada, T. Changes in urban mobility in Sapporo city, Japan due to the COVID-19 emergency declarations. *Transp. Res. Interdiscip. Perspect.* **2020**, *7*, 100212. [CrossRef]
26. ADRIMOB Project. Available online: <http://igoumenitsa.easybike.gr/en/adrimob-project> (accessed on 25 May 2020).
27. Sustainable Urban Mobility in MED PORT Cities—SUMPORT Project, D3.11.2 Final Testing Report on the Extension of Bikesharing System in Igoumenitsa, Regional Unit of Thesprotia, Region of Epirus. 2019. Available online: <https://sumport.interreg-med.eu/pilot-actions/extension-of-bike-sharing-system/> (accessed on 18 January 2022).
28. New Bike Sharing System in the City of Igoumenitsa. Available online: <https://igoumenitsa.cyclopolis.gr/index.php/en/> (accessed on 25 May 2020).
29. Cyclopolis Bike Sharing System of the Municipality of Rhodes. Available online: <https://rhodes.cyclopolis.gr/index.php/en/stations> (accessed on 11 August 2020).
30. Cyclopolis Bike Sharing System of the Municipality of Chania. Available online: <https://chania.cyclopolis.gr/index.php/el/component/content/?view=featured> (accessed on 11 August 2020).
31. AccuWeather, Inc. ©2020, 385 Science Park Road, State College, Pennsylvania 16803, a Weather Forecast Online Service. Available online: <https://www.accuweather.com/en/gr/igoumenitsa/186239/april-weather/186239?year=2020> (accessed on 25 May 2020).
32. World Wide Organization from Amsterdam, ©2010–2020 World Weather & Climate Information. Available online: <https://weather-and-climate.com/average-monthly-precipitation-Rainfall,igoumenitsa-epirus-gr,Greece> (accessed on 25 May 2020).
33. Sustainable Urban Mobility in MED PORT Cities—SUMPORT Project, D3.15.1 Evaluation Methodology, Insitute for Transport and Logisitcs Foundation-ITL. 2017. Available online: <https://sumport.interreg-med.eu/what-we-achieve/final-publication/> (accessed on 18 January 2022).
34. Kim, K.O. Impact of COVID-19 in Usage Patterns of a Bike-Sharing System: Case Study of Seoul. *J. Transp. Eng. Part A Syst.* **2021**, *147*, 05021006. [CrossRef]
35. Beck, N.; Katz, J. What to do (and not to do) with time-series cross-section data. *Am. Political Sci. Rev.* **1995**, *89*, 634–647. [CrossRef]
36. Teixeira, J.F.; Lopes, M. The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York’s Citi Bike. *Transp. Res. Interdiscip. Perspect.* **2020**, *6*, 100166. [CrossRef]
37. Wang, H.; Noland, R.B. Bikeshare and subway ridership changes during the COVID-19 pandemic in New York City. *Transp. Policy* **2021**, *106*, 262–270. [CrossRef]
38. Bucsky, P. Modal share changes due to COVID-19: The case of Budapest. *Transp. Res. Interdiscip. Perspect.* **2020**, *8*, 100141. [CrossRef]
39. Nikiporiadis, A.; Ayfantopoulou, G.; Stamelou, A. Assessing the Impact of COVID-19 on Bike Sharing Usage: The Case of Thessaloniki, Greece. *Sustainability* **2020**, *12*, 8215. [CrossRef]
40. Politis, I.; Georgiadis, G.; Papadopoulos, E.; Fyrogenis, I.; Nikolaidou, A.; Kopsacheilis, A.; Sdoukopoulos, A.; Verani, E. COVID-19 lockdown measures and travel behavior: The case of Thessaloniki, Greece. *Transp. Res. Interdiscip. Perspect.* **2021**, *10*, 100345. Available online: <https://www.sciencedirect.com/science/article/pii/S259019822100052X> (accessed on 22 November 2021). [CrossRef]
41. Hua, M.; Chen, X.; Cheng, L.; Chen, J. Should bike sharing continue operating during the COVID-19 pandemic? Empirical findings from Nanjing China. *arXiv* **2020**, arXiv:2012.02946. [CrossRef]

42. Padmanabhan, V.; Penmetsa, P.; Li, X.; Dhondia, F.; Dhondia, S.; Parrish, A. COVID-19 effects on shared-biking in New York, Boston, and Chicago. *Transp. Res. Interdiscip. Perspect.* **2021**, *9*, 100282. [CrossRef]
43. Campisi, T.; Basbas, S.; Skoufas, A.; Akgun, N.; Ticali, D.; Tesoriere, G. The Impact of COVID-19 Pandemic on the Resilience of Sustainable Mobility in Sicily. *Sustainability* **2020**, *12*, 8829. [CrossRef]
44. Molloy, J.; Schatzmann, T.; Schoeman, B.; Tchervenkov, C.; Hintermann, B.; Axhausen, K.W. Observed impacts of the COVID-19 first wave on travel behaviour in Switzerland based on a large GPS panel. *Transp. Policy* **2021**, *104*, 43–51. [CrossRef]
45. König, A.; Drebler, A. A mixed-methods analysis of mobility behavior changes in the COVID-19 era in a rural case study. *Eur. Transp. Res. Rev.* **2021**, *13*, 15. [CrossRef]
46. Freeman, S. *Bike Sharing during COVID-19*; National Collaborating Center for Environmental Health: Vancouver, BC, Canada, 16 June 2020. Available online: <https://nceh.ca/content/blog/bike-sharing-during-covid-19> (accessed on 13 December 2021).
47. Bureau of Transportation Statistics, United States Department of Transportation. COVID-19 Crushes Bikeshare and E-scooter Ridership, Closes Systems Permanently in Some Cities. 2021. Available online: <https://www.bts.gov/data-spotlight/covid-19-crushes-bikeshare-e-scooter-ridership> (accessed on 17 May 2021).
48. Shang, W.L.; Chen, J.; Bi, H.; Sui, Y.; Chen, Y.; Yu, H. Impacts of COVID-19 pandemic on user behaviors and environmental benefits of bike sharing: A big-data analysis. *Appl. Energy* **2021**, *285*, 11642. [CrossRef]
49. Tokey, A.I. Change of Bike-share Usage in Five Cities of United States during COVID-19. *Findings* **2020**. [CrossRef]
50. Eisenmann, C.; Nobis, C.; Kolarova, V.; Lenz, B.; Winkler, C. Transport mode use during the COVID-19 lockdown period in Germany: The car became more important, public transport lost ground. *Transp. Policy* **2021**, *103*, 60–67. [CrossRef]
51. Shakibaei, S.; De Jong, G.C.; Alpkokin, P.; Rashidi, T.H. Impact of the COVID-19 pandemic on travel behavior in Istanbul: A panel data analysis. *Sustain. Cities Soc.* **2021**, *65*, 102619. [CrossRef]
52. Scorrano, M.; Danielis, R. Active mobility in an Italian city: Mode choice determinants and attitudes before and during the COVID-19 emergency. *Res. Transp. Econ.* **2021**, *86*, 101031. [CrossRef]
53. Consumer Data Research Centre, UK. Bikeshare Activity during COVID-19, Online Story. Available online: <https://data.cdrc.ac.uk/stories/uk-bikeshare-activity-during-covid-19> (accessed on 17 May 2021).
54. Heydari, S.; Konstantinou, G.; Beshoodi, A.W. Effect of the COVID-19 pandemic on bike sharing demand and hire time: Evidence from Santander Cycles in London. *PLoS ONE* **2021**, *16*, e0260969. [CrossRef] [PubMed]
55. Kubal'ak, S.; Kalasova, A.; Hajnik, A. The Bike-Sharing System in Slovakia and the Impact of COVID-19 on This Shared Mobility Service in Selected City. *Sustainability* **2021**, *13*, 6544. [CrossRef]
56. Jobe, J.; Griffin, G.P. Bike share response to COVID-19. *Transp. Res. Interdiscip. Perspect.* **2021**, *10*, 100353. [CrossRef]
57. Bergantino, A.S.; Intini, M.; Tangari, L. Influencing factors for potential bike-sharing users: An empirical analysis during the COVID-19 pandemic. *Res. Transp. Econ.* **2021**, *86*, 101028. [CrossRef]
58. Hu, S.; Xiong, C.; Liu, Z.; Zhang, L. Examining spatiotemporal changing patterns of bike-sharing usage during COVID-19 pandemic. *J. Transp. Geogr.* **2021**, *91*, 102997. [CrossRef]
59. Nikitas, A.; Tsigdinos, S.; Karolemeas, C.; Kourmpa, E.; Bakogiannis, E. Cycling in the era of covid-19: Lessons learnt and best practice policy recommendations for a more bike-centric future. *Sustainability* **2021**, *13*, 4620. [CrossRef]
60. Institute for Transportation & Development Policy. *The Bikeshare Planning Guide*; Institute for Transportation & Development Policy: New York, NY, USA, 2018; p. 23. Available online: <https://www.transformative-mobility.org/assets/publications/The-Bikeshare-Planning-Guide-ITDP-Datei.pdf> (accessed on 11 January 2022).
61. Wikipedia. Available online: <https://en.wikipedia.org/wiki/Igoumenitsa> (accessed on 6 January 2022).
62. Wikipedia. Available online: <http://population.city/greece/chania/> (accessed on 6 January 2022).
63. Population. City. Available online: <http://population.city/greece/rodos/> (accessed on 6 January 2022).
64. Hania. News, National Edition. Available online: <https://hania.news/2021/12/02/343986/> (accessed on 10 December 2021).
65. Money-Tourism.gr, National Edition. Available online: <https://money-tourism.gr/meiomenes-83-4-oi-afixeis-allodapon-toyriston-sti-rodos-8mino-2020/> (accessed on 10 December 2021).
66. Tsakalidis, A.; Gkoumas, K.; Pekar, F. Digital transportation supporting transport decarbonization: Technological development in EU-funded research and innovation. *Sustainability* **2020**, *12*, 3762. [CrossRef]
67. Gkoumas, K.; dos Santos, F.L.M.; Stepniak, M.; Pekar, F. Research and Innovation Supporting the European Sustainable and Smart Mobility Strategy: A Technology Perspective from Recent European Union Projects. *Appl. Sci.* **2021**, *11*, 1981. [CrossRef]
68. Alonso Raposo, M.; Ciuffo, B.; Alves Dias, P.; Ardente, F.; Aurambout, J.; Baldini, G.; Baranzelli, C.; Blagoeva, D.; Bobba, S.; Braun, R.; et al. *The Future of Road Transport*; Publications Office of the European Union: Luxembourg, 2019. [CrossRef]
69. Tsakalidis, A.; Gkoumas, K.; Grosso, M.; Pekar, F. TRIMIS: Modular Development of an Integrated Policy-Support Tool for Forward-Oriented Transport Research and Innovation Analysis. *Sustainability* **2020**, *12*, 194. [CrossRef]

Article

Mapping of Hydrothermal Alteration Zones in the Kelâat M'Gouna Region Using Airborne Gamma-Ray Spectrometry and Remote Sensing Data: Mining Implications (Eastern Anti-Atlas, Morocco)

Younes Mamouch ^{1,*}, Ahmed Attou ¹, Abdelhalim Miftah ¹, Mohammed Ouchchen ², Bouchra Dadi ^{3,4}, Lahsen Achkouch ¹, Yassine Et-tayea ¹, Abdelhamid Allaoui ⁵, Mustapha Boualoul ⁵, Giovanni Randazzo ^{6,*}, Stefania Lanza ^{7,*} and Anselme Muzirafuti ^{6,*}

Citation: Mamouch, Y.; Attou, A.; Miftah, A.; Ouchchen, M.; Dadi, B.; Achkouch, L.; Et-tayea, Y.; Allaoui, A.; Boualoul, M.; Randazzo, G.; et al. Mapping of Hydrothermal Alteration Zones in the Kelâat M'Gouna Region Using Airborne Gamma-Ray Spectrometry and Remote Sensing Data: Mining Implications (Eastern Anti-Atlas, Morocco). *Appl. Sci.* **2022**, *12*, 957. <https://doi.org/10.3390/app12030957>

Academic Editor: Saro Lee

Received: 3 December 2021

Accepted: 13 January 2022

Published: 18 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- ¹ Laboratory Physico-Chemistry of Processes and Materials, Research Team Geology of the Mining and Energetics Resources, Faculty of Sciences and Technology, Hassan First University of Settat, Settat 26002, Morocco; attouahmed@hotmail.com (A.A.); a.miftah@uhp.ac.ma (A.M.); l.achkouch@uhp.ac.ma (L.A.); y.et-tayea@uhp.ac.ma (Y.E.-t.)
 - ² Laboratory of Applied Geology and Geo-Environment, Faculty of Science, Ibn Zohr University, Agadir 80000, Morocco; mohammed.ouchchen@edu.uiz.ac.ma
 - ³ Structural Geology and Thematic Mapping Laboratory, Earth Sciences Department, Faculty of Science, Ibn Zohr University, Agadir 80000, Morocco; bouchra.dadi@edu.uiz.ac.ma
 - ⁴ Laboratory for Sustainable Innovation and Applied Research, Universiapolis, Technical University of Agadir, Agadir 80000, Morocco
 - ⁵ Laboratory of Geoenvironment and Environment, Department of Geology, Faculty of Sciences, University of Moulay Ismail, BP 11201 Zitoune, Meknes 50000, Morocco; abdelhamid.gaig11@gmail.com (A.A.); boualoul@gmail.com (M.B.)
 - ⁶ Dipartimento di Scienze Matematiche e Informatiche, Scienze Fisiche e Scienze della Terra, Università degli Studi di Messina, Via F. Stagno d'Alcontres, 31-98166 Messina, Italy
 - ⁷ GeoloGIS s.r.l. Spin Off of Università degli Studi di Messina, 31-98166 Messina, Italy
- * Correspondence: y.mamouch@uhp.ac.ma (Y.M.); giovanni.randazzo@unime.it (G.R.); stefania.lanza@unime.it (S.L.); anselme.muzirafuti@unime.it (A.M.); Tel.: +212-695-504-000 (Y.M.)

Abstract: The mapping of hydrothermal alteration zones associated with mineralization is of paramount importance in searching for metal deposits. For this purpose, targeting alteration zones by analyzing airborne geophysical and satellite imagery provides accurate and reliable results. In the Kelâat M'Gouna inlier, located in the Saghro Massif of the Moroccan Anti Atlas, natural gamma-ray spectrometry and ASTER satellite data were used to map hydrothermal alteration zones. Natural gamma-ray spectrometry data were processed to produce maps of Potassium (K in %), Uranium (eU in ppm), Thorium (eTh in ppm) and ratios of K/eTh and K/eU. In addition, four-band ratios were computed, on ASTER data, to map the distribution of clay minerals, phyllic minerals, propylitic minerals, and iron oxides. The combined results obtained from geophysical and satellite data were further exploited by fuzzy logic modelling using the Geographic Information System (GIS) to generate a mineral prospectivity map. Seven hydrothermal alteration zones likely to be favorable for mineralization have been identified. They show a spatial correlation with (i) known surface prospects and mineral occurrences, (ii) the granite-encasing contact zone, and (iii) the fault zones (Sidi Flah and Tagmout faults). This research therefore provides important information on the prospecting of mineral potential in the study area.

Keywords: mineral exploration; natural gamma-ray spectrometry; ASTER; fuzzy logic modelling; Kelâat M'Gouna inlier; Eastern Anti-Atlas; Morocco

1. Introduction

The mapping of hydrothermal alteration zones associated with mineralized systems is of great importance in mineral exploration, especially in the early stages of metal de-

posit exploration [1–11]. However, the hydrothermal alteration minerals formed can vary significantly, depending on the chemical composition of the primary rock and hydrothermal fluids, the nature of the host, temperature and pressure conditions, and the tectonic setting [12].

Many methods have been proposed to map the spatial distribution of hydrothermal alteration zones. The method of processing multispectral satellite images, especially those of the ASTER sensor, is qualified among the most used approaches [4,6–8,10,11,13–18]. Apart from remote sensing data, processing of geophysical data contributes assuredly to minerals exploration through their different geophysical components (gravimetric, magnetometric, electrical, electromagnetic and natural gamma spectrometric methods) [19–24]. The natural gamma-ray spectrometric method has been widely and successfully used in mineral exploration and identification of alteration zones [20,21,25–30]. However, few studies have been tested the application of this method in conjunction with the open accessible free satellite images.

In this regard, we used satellite data from the ASTER sensor, airborne natural gamma-ray spectrometry and surface geological data to identify potential mineral exploration sites in the area of Kelâat M’Gouna/Morocco. This area is located in the Jbel Saghro massif, one of the most important massifs of Morocco’s Anti-Atlas chain, which is recognized for its composition in mineral raw material (Figure 1). This area has recently been the subject of mining exploration activity that has led to discovering three gold deposits: Ismlal, Talat-n-Tabarought and Tawrirt-n-Çwalkh. These deposits are controlled by complex hydrothermal processes such as silicification, chloritization, hematization and sericitization [31,32]. These processes occurred under specific geodynamic conditions that characterized the geological history of the Anti-Atlas.

The aim of this study is to conduct a geological mapping of mineral composition in Kelâat M’Gouna area by combining geophysical and remote sensing data. This study was conducted with main objective of (i) accurately mapping hydrothermal alteration zones; (ii) developing a mineral prospectivity maps by combining thematic layers using fuzzy logic modelling; (iii) relating mineral formation processes to hydrothermal events; and (iv) clarifying the characteristics of the mineral formations in the different tectonic units of the study area.

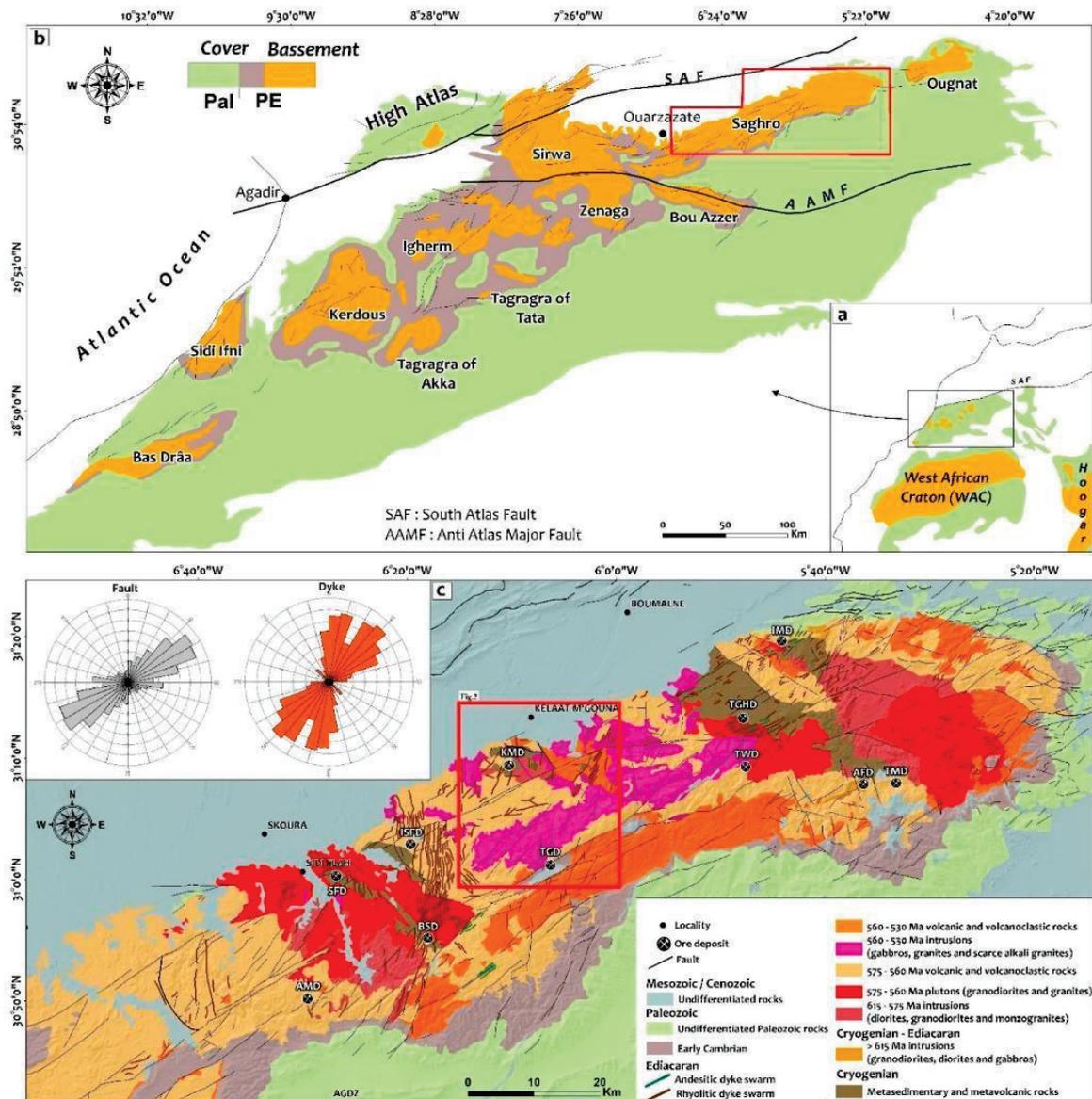


Figure 1. (a) Location of the Moroccan Anti-Atlas Range relative to the West African Craton [33]. (b) General geological map of the Anti-Atlas showing its main Precambrian inliers [34,35], modified. (c) Geological map of the Saghro Massif with its main metalliferous deposits (IMD: Imeter Deposit; TGHD: Taghassa Deposit; KMD: Kelâat M’Gouna Deposits; ISFD: Issarfane; BSD: Bouskour Deposit; AMD: Amzwaro Deposit; TMD: Tizi Moudou Deposit; AFD: Asfalou Deposit; TWD: Tiwit Deposit; TGMD: Tagmout Deposit; SFD: Sidi Flah Deposit). In addition, their rose diagrams showing the trends of faults on the left and dykes on the right. The study area is marked by the red polygon [36,37], modified.

2. Material and Methods

2.1. Geological Context of the Study Area

The Saghro massif is located NE of the major accident of the Anti-Atlas (Figure 1b), which is interpreted as a Pan-African suture following the identification of a complex ophiolite that stakes it [38]. It is subdivided into three domains, (i) the Western Saghro corresponding to the Sidi Flah and Bouskour inliers, (ii) the Central Saghro corresponding to the Kelâat M’Gouna inlier and (iii) the Eastern Saghro formed by the Boumalne and Imiter inliers (Figure 1).

The study area corresponds to the geological sheet of Kelâat M’Gouna, at 1:50,000 scales. It is geographically attached to the Kelâat M’Gouna inlier, between meridians 6°00 W and 6°15 W and parallels 31°00 N and 31°15 N. This area contains a wide range of geological formations ranging in age from Cryogenian to the present day (Figure 2). The Lower Cryogenian formations, which are volcano-sedimentary and metamorphic, are the oldest in the Jbel Saghro massif and correspond to turbidites interbedded with basic volcanic flows and intruded by gabbros, diorites and granites [32,39–42]. They are overlain in anomalous contact, by Upper Cryogenian conglomerates, sandstones, limestones, cinerites, rhyolites and andesites [36,43], and they are overlain by the formations of the Ouarzazate Group, which in turn, comprises two discordant sets on top of each other [31]. The first set corresponds to the Lower Ediacaran or the Lower Ouarzazate Group, composed of potassium-rich volcanic and granitic formations that intrude all the previous geological formations. The second set corresponds to the Upper Ediacaran or Upper Ouarzazate Group, consisting of detrital and volcanic formations intruded by rhyolite and microgranite dikes. The formations of the Upper Ouarzazate Group are tectonically continuous with the Adoudounian sedimentary formations. The Paleozoic formations are poorly represented in the study area. They are limited to the Tagmout graben (Figure 2), where they are exposed in conglomerates, pink sandstones with basalt intercalation, Paradoxides shales and Tabanites sandstones [31]. The Cenozoic-Quaternary sedimentary deposits correspond to the filling of the Ouarzazate basin.

2.2. Structural and Tectonic Context of the Study Area

The Saghro Massif is affected by the major phase of the Pan-African Orogeny (B1) dated at 685 ± 15 Ma [44]. This phase is responsible for the emplacement of diorite, quartz diorite and granodiorite massifs along N130° troughs at Bouskour [45,46] and at Boumalne-Dadès [47]. This phase is followed by the late phase (B2) of weak intensity and without metamorphic transformation. It is responsible for the development of granitic massifs in the Bouskour and Ougnat inliers.

The formations, which outcrop on the Kelâat M’Gouna sheet, have undergone the various tectonic events that affected the Anti-Atlas. These events are reflected by the dominance of brittle structures classified in general according to their orientation in three families:

- a. The first family-oriented NE-SW is the most dominant and with plurikilometer lengths. The most important faults of this family are the Sidi Flah fault and the faults bordering the Tagmout graben (Figure 2). The Sidi Flah fault shows a sinister set related to its reactivation under the effects of the Hercynian orogeny [31]. The Tagmout graben corresponds to the western extension of the great graben located in the heart of the Saghro massif [48,49].
- b. The second family is oriented NNW-SSE and shows a dexterous set in coherence with the antithetic movement [31].
- c. The third family is oriented NNE-SW and corresponds to dolerite and rhyolite dikes that were emplaced in the echelon distension fractures of the synthetic Riedel faults playing in shear in the same stress field around 564 Ma [50].

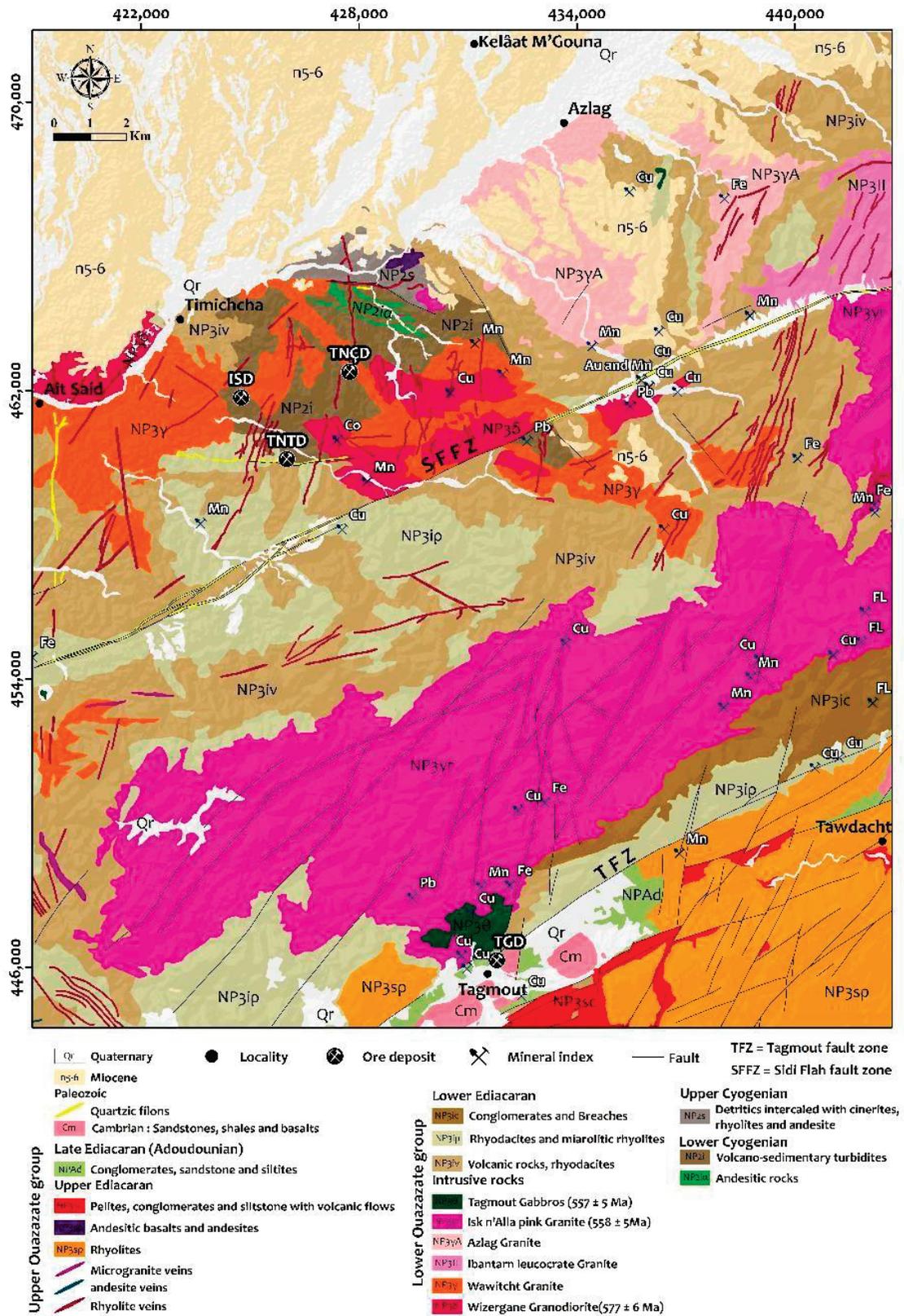


Figure 2. Geological map of the study area extracted from the geological map 1:50,000 of Kelâat M'Gouna, showing its main ore deposits (ISD: Ismlal gold Deposit; TNÇD: Taourirt-n-Çwalh gold Deposit; TNTD: Talat-n-Tbarought gold Deposit; TGD: Tagmout copper Deposit).

2.3. Ore Deposits of the Study Area

Many distinct categories of mineralization have been identified on the Kelâat M'Gouna sheet and constitute about 50 showings and deposits such as gold, copper, iron, lead, fluorine, manganese, cobalt and silver (Figure 2). The most important mineralization is represented by gold deposits represented by three gold occurrences that are so far in the development phase: Ismlal, Talat-n-Tabarought and Tawrirt-n-Çwalh [31,32,41]. According to mineralogical, textural, structural and chemical aspects of these occurrences, two main types of mineralization have been distinguished: an old porphyry type system, followed by a younger epithermal type system [41,51]. At Ismlal, gold mineralization is hosted in Lower Cryogenian volcano-sedimentary turbidites (NP2i), intruded by granodiorites of Lower Ediacaran age (NP3i). It occurs as quartz veins ranging from $N0^\circ$ to $N120^\circ$, in breccias of general orientation $N120^\circ$ and dissemination in the volcano-sedimentary turbidites of NP2i. Gold grades are estimated at 0.5 g/t. The length of this mineralized zone is about 800 m, and its width is about 200 m [31,32]. At Talat-n-Tabarought, the gold-bearing structure is in the form of the NW-SE, and NE-SW trending T. Gold grades (between 0.1 and 0.3 g/t) is low compared to those at Ismlal [31]. The mineralization host is NP2i sandstone, and it presents quartz micro veins at the edge of pyrite granodiorite and tourmaline granodiorite intrusions. At Tawrirt-n-Çwalh, the gold mineralization is discontinuous in an NNE-SW direction and is 800 m in extent. The host of this mineralization corresponds to a metamorphic sandstone-pelitic alternation of NP2i age injected by quartz veins and potassic feldspars with gold visible to the naked eye [31,41]. Gold grades are interesting and range from 1 to 9 g/t [31,32,41]. In addition, copper mineralization is well recognized in the region, the most important of which is Tagmout which is located south of the study area (Figure 2). It is a hydrothermal vein deposit hosted by gabbros, monzogabbros and granites of the Upper Neoproterozoic. The metallogenic study of the deposit identified a copper paragenesis dominated by chalcopyrite, chalcocite, bornite, covellite, cuprite, grey copper and malachite [31].

2.4. Geophysical and Remote Sensing Data

2.4.1. Radiometric Data

Airborne geophysics based on gamma-ray spectrometry estimates the concentration of Potassium (K), Uranium (eU) and Thorium (eTh) at the earth's surface down to 30–45 cm [52–55]. These radioelements occur in widely varying concentrations in the rocks that form the earth's crust. The Table 1 shows the concentrations of the three radioelements in the main rocks categories.

The airborne gamma-ray spectrometry data used in this study were acquired in the Moroccan Anti Atlas in 1998 by the company Géoterrx-Dighem for the Ministry of Energy and Mines. The measurements were made with an average ground clearance of 60 m by the Exploranium GR-830/3 spectrometer. The flight lines are oriented $N315^\circ$ and spaced at 500 m. The flight line crossings are $N45^\circ$ and 4000 m apart. Gamma-ray emissions were recorded from the ground and air over an energy range of 0 to 3 MeV with 8 downward and 2 upward crystals, respectively. Count rates were determined within three windows corresponding to natural radiogenic concentrations of Potassium (K, 1.46 MeV), Uranium (U, 1.76 MeV) and Thorium (Th, 2.62 MeV). The radiometric data were recorded at a lower frequency (1 Hz) and with an average spacing of 63 m. The recorded data then underwent a series of corrections which are: (i) activity time correction; (ii) calculation of effective ground clearance at standard temperature and pressure conditions; (iii) subtraction of cosmic and helicopter noise; (iv) subtraction of radon background (assessed by upward facing detector measurements); (v) Compton effect correction and (vi) attenuation corrections.

These radiometric data were provided to us in the form of digital maps. Using ArcGIS 10.3 software, these maps were first georeferenced and digitized to create a digital database (by digitizing the intersection of the iso-value curves and the flight lines). They were then interpolated using the inverse distance weighting (IDW) interpolation method to obtain maps representing the horizontal variation of radioactive concentrations of three

elements: (i) Potassium (K in %), (ii) equivalent Thorium (eTh in ppm), and (iii) equivalent Uranium (eU in ppm). Furthermore, ratios of K/eTh and K/eU were calculated to delineate Potassium enrichment zones as indicators of potential mineral resource-related alteration zones [25,28,56–58] (Figure 3).

Table 1. Radioelement concentrations in main categories of rocks [53,56].

Rock Type	Potassium (%)		Uranium (ppm)		Thorium (ppm)	
	Range	Mean	Range	Mean	Range	Mean
Acid Extrusives	1.0–6.2	3.1	0.8–16.4	4.1	1.1–41.0	11.9
Acid Intrusives	1.0–7.6	3.4	0.1–30.0	4.5	0.1–253.1	25.7
Intermediate Extrusives	0.01–2.5	1.1	0.2–2.6	1.1	0.4–6.4	2.4
Intermediate Intrusives	0.1–6.2	2.1	0.1–23.4	3.2	0.4–106.0	12.2
Basic Extrusives	0.06–2.4	0.7	0.03–3.3	0.8	0.05–8.8	2.2
Basic Intrusives	0.01–2.6	0.8	0.01–5.7	0.8	0.03–15.0	2.3
Ultrabasic	0–0.8	0.3	0–1.6	0.3	0–7.5	1.4
Chemical Sedimentary Rocks	0.02–8.4	0.6	0.03–26.7	3.6	0.03–132.0	14.9
Carbonates	0.01–3.5	0.3	0.03–18.0	2	0.03–10.8	1.3
Detrital Sedimentary Rocks	0.01–9.7	1.5	0.1–80.0	4.8	0.2–362.0	12.4
Metamorphosed Igneous Rocks	0.1–6.1	2.5	0.1–148.5	4	0.1–104.2	14.8
Metamorphosed Sedimentary Rocks	0.01–5.3	2.1	0.1–53.4	3	0.1–91.4	12

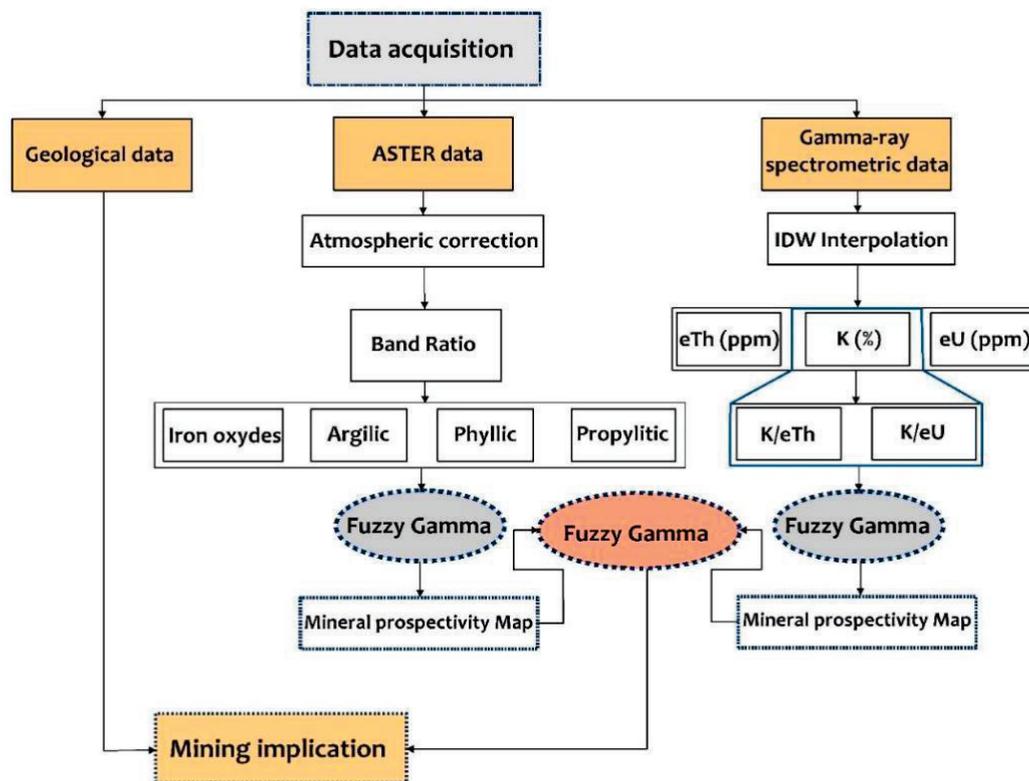


Figure 3. The methodological flowchart used in this study.

The image was pre-processed using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) method [59–61]. This correction process requires the luminance image and generates a reflectance corrected image. Subsequently, we calculated the band

ratios “Bands Ratios” related to the depth of the absorption band “RBD” which is a technique that has been used for many years in remote sensing for mapping hydrothermal alteration minerals [7,15,62–66]. In the present study, four alteration mineral assemblage domains were calculated: (i) clay alteration (kaolinite and montmorillonite), (ii) phyllic alteration (sericite and illite), (iii) propylitic alteration (epidote, chlorite and carbonates), and (iv) alteration ferric iron uptake (hematite, goethite and jarosite) (Figure 3).

2.4.2. ASTER Satellite Imagery

The space remote sensing data used in this study correspond to an ASTER sensor image acquired on 9 September 2003, with a cloud cover of the order of 0%. Its spectral resolution makes it possible to map alteration minerals related to mineralization processes since it contains six spectral bands in the short wave infrared region ranging from 1.60 to 2.430 μm [13]. Other characteristics and features of the Aster sensor are presented in Table 2.

Table 2. Characteristics and performance of the ASTER sensor [67].

Subsystem	Band No.	Spectral Range (μm)	Radiometric Resolution	Absolute Accuracy	Spatial Resolution (m)	Signal Quantization (Bits)
VNIR	1	0.52–0.60	$\text{NE}\Delta\rho \leq 0.5\%$	$\leq \pm 4\%$	15	8
	2	0.63–0.69				
	3	0.78–0.86				
	3N	0.78–0.86				
SWIR	4	1.60–1.70	$\text{NE}\Delta\rho \leq 0.5\%$	$\leq \pm 4\%$	30	8
	5	2.145–2.185	$\text{NE}\Delta\rho \leq 1.3\%$			
	6	2.185–2.225	$\text{NE}\Delta\rho \leq 1.3\%$			
	7	2.235–2.285	$\text{NE}\Delta\rho \leq 1.3\%$			
	8	2.295–2.365	$\text{NE}\Delta\rho \leq 1.0\%$			
	9	2.360–2.430	$\text{NE}\Delta\rho \leq 1.3\%$			
TIR	10	8.125–8.475	$\text{NE}\Delta T \leq 0.3 \text{ K}$	$\leq 3 \text{ K (200–240 K)}$	90	12
	11	8.475–8.825		$\leq 2 \text{ K (240–270 K)}$		
	12	8.925–9.275		$\leq 1 \text{ K (270–340 K)}$		
	13	10.25–10.95		$\leq 2 \text{ K (340–370 K)}$		
	14	10.95–11.65				

2.4.3. Fuzzy Logic Modelling of Radiometric and ASTER Satellite Imagery

Fuzzy logic modelling is a widely and successfully used technique in mining mapping and mainly for the development of mineral prospectivity maps [8,66,68–70]. Mathematically, it is a form of multi-valued logic based on the fuzzy set theory where the real values of the variables are included in the interval [0–1]; zero corresponds to non-membership, and 1 corresponds to full membership [71]. This method was first proposed in 1965 by Zadeh and is defined as follows:

$$A_{ij} = \{ (x_{ij}, \mu_A) / x_{ij} \in X_i \}, 0 \leq \mu_A \leq 1$$

where μ_A is called the degree of membership function of x in A and X corresponds to a set of layers X_i ($i = 1, 2, 3, \dots, n$), and each layer to r classes defined as ($j = 1, 2, 3, \dots, r$).

The degree of membership μ_A plays in the interval [0–1], so: If $0 \leq \mu_A < 0.5$: x_{ij} is not conducive to mineralization; if $\mu_A = 0.5$: we cannot determine is x_{ij} conducive to mineralization or not; if $0.5 < \mu_A \leq 1$ it means that x_{ij} is conducive to mineralization.

In this study, we applied this analysis to produce mineral prospectivity maps following the methodological approach presented in the Figure 3. The first map combines the four layers corresponding to the band ratios extracted from the ASTER image (Table 3). Meanwhile, the second map combines the geophysical data corresponding to the K, K/eTh ratio and K/eU ratio layers. These layers were fuzzified individually using the linear membership function. Then the fuzzy gamma operator was applied to combine our thematic layers. The choice of this operator is based on the fact that it is a compromise between the fuzzy algebraic sum and its product [72]. In addition, it is possible to develop improvements on the input of the resulting maps [73]. After several trials, we took 0.72 as the value of parameter γ . The fuzzification parameters used for the input data are represented in the Table 4.

Table 3. Band ratios used in this analysis for mapping weathering zones in the study area.

Bands Ratios	Equations	Target Minerals
CLMI	(band 4 + band 6)/band 5	Alunite/kaolinite/montmorillonite
PHMI	(band 5 + band 7)/band 6	Sericite/muscovite/illite/smectite
PRMI	(band 7 + band 9)/band 8	Epidote/chlorite/carbonates
IOI	(band 5/band 3) + (band 1/band 2)	Hematite/goethite/jarosite

Table 4. Fuzzy membership parameters used for input layers.

Data Origin	Input Layer	Membership Type	Fuzzy Operator
Aster Dataset	Argilic	Lineare	$\gamma = 0.72$
	Phyllic	Lineare	
	Propylitic	Lineare	
	Iron oxydes	Lineare	
Spectrometry gamma Dataset	K	Lineare	$\gamma = 0.72$
	K/eTh	Lineare	
	K/eU	Lineare	

3. Results

3.1. Mapping of Hydrothermal Alteration Zones

3.1.1. Contribution of the Radiometry

The resulting radiometric maps provide a synthetic view of the heterogeneities of the geological formations encountered based on the radiometric signature of the different rock units and the structural trends that affect them. The highest concentrations of radioactive elements (more than 5.70% K, 15.80 ppm eTh and 8.30 ppm eU) are concentrated in the SSE part of the map (Figure 4), which could be related to the presence of a large granitic massif called “pink granites of Isk n’Alla” (Figure 2). This contrast sequence is also marked by low to moderate amounts observed, especially if we go towards the North of the study area (0.20 to 0.45% K, 5.25 to 7.24 ppm Th and 0.91 to 1.29 ppm U) (Figure 4). The geological map in Figure 2 and radioelement contents in the main rock categories (Table 1) were used to geologically interpret the spatial distributions of potassium, thorium and uranium.

View that Potassium is more mobile than thorium in terms of element mobility during chemical alteration processes [75,76]. The K/eTh ratio is often considered the best indicator of Potassium enrichment zones related to hydrothermal alteration. The authors of [77] show that the Potassium/Thorium ratio is nearly constant in most rocks and generally ranges from 0.17 to 0.2 (K/eTh). Values of the k/eTh ratio that exceed this range could be due to hydrothermal alteration processes associated with the emplacement of magmatic-hydrothermal mineralization. In the case of the study area, the resulting K/eTh values allow us to distinguish seven anomalous potassium domains (Figure 4d). Domain 1 presents extremely high K/eTh values reaching 0.65, associated with alteration zones probably linked to mineralized zones. Indeed, this domain is characterized by the outcrop of several showings, namely copper, manganese, silver, etc . . . (Figure 4d). Domains 2, 3 and 5 are respectively associated with young granites “pink granite of Isk n’Alla,” rhyolites and rhyodacites of lower Ediacaran and granites of Wawicht whose ratio values exceed 0.4 (Figure 4d). Furthermore, domain 4 is related to the Sidi Flah fault zone, which hosts several mineralized showings such as gold, copper, lead and manganese (Figure 4d). Domain 6 is related to gold occurrences closely related to the Wawicht granite and is characterized by high values of K/eTh ratios (>0.35). To the north of the study area is domain 7, where K/eTh values vary considerably and sometimes reach 3.35. This area is associated with the Azlag granite, which hosts a copper occurrence (Figures 2 and 4d).

3.1.3. Contribution of Aster Data

The results of the extracted band ratios for mapping clay alteration minerals (kaolinite, alunite and montmorillonite), phyllite alteration (sericite, muscovite and illite), propylitic alteration (epidote, chlorite and carbonates) and iron oxide alteration (hematite, goethite and jarosite) are shown in Figure 5.

The band ratio $(b_4 + b_6)/b_5$ shows the spatial distribution of clay alteration minerals (blue pixels), which are mostly mapped in the Lower Cryogenian turbidites, in the Upper Cryogenian units, in the Lower Ouarzazate Group units and very locally in the Quaternary unit along the Oueds (Figure 4a).

The band ratio $(b_5 + b_7)/b_6$ illustrates the surface distribution of phyllic alteration minerals in yellow pixels (Figure 4b), which are mapped in the Upper Cryogenian conglomerate and pelite units with intercalations of cinerites, rhyolites and andesites and the Lower Ouarzazate Group granite unit. These bands are mapped because of their mineralogical composition’s high content of sericite, muscovite, and illite. The band ratio $(b_7 + b_9)/b_8$, corresponds to the propylitic alteration mineral index shown in green pixels; it shows a distribution more or less similar to the phyllic alteration mineral index and is generally related to the Upper Cryogenian volcano-sedimentary unit and the Lower Ouarzazate Group granite unit (Figure 4c).

The band ratio $(b_5/b_3) + (b_1/b_2)$ shows the surface distribution of iron oxides in red pixels (Figure 4d). Compared with the geological map, the high abundance of all these mapped minerals is typically associated with Cryogenian and Lower Ouarzazate Group units. It shows a strong correlation with the mineral occurrences and prospects in the study area.

metal deposits, especially the gold deposits. Benziane et al., 2008 [31], showed that the three gold prospects in the study area are associated with hydrothermal alteration phenomena. According to these authors, silicification is expressed either by developing quartz stockworks or by dispersion in the host volcano-sedimentary formations, following a general NE-SW direction. Generally, this silicification is accompanied locally by opening zones filled with potassium feldspar and tourmaline. In this respect, ratios of K/eTh and K/eU were calculated to highlight enrichment zones in this element. Chloritization is manifested by the development of large zones in the soft greenish levels of the volcano-sedimentary series. For this purpose, the Aster image band ratio $(b7 + b9)/b8$ was calculated to map the spatial distribution of high chlorite content zones. Hematization is materialized by the development of iron oxides and hydroxides, following the fracturing planes, which are generally oriented $N20^\circ$. This type of alteration mineral was mapped using the band ratio $(b5/b3) + (b1/b2)$. Finally, the band ratio $(b5 + b7)/b6$ was used to map the sericitization in the study area in a subordinate manner [31]. In the form of thematic layers, this database was fuzzified by the fuzzy logic modelling technique by combining the two mineral prospectivity maps, Aster and radiometric. Subsequently, a mineral prospectivity map for the study area was generated. A total of seven prospective value areas were identified (Figure 7), and they are mainly associated with alteration zones:

- Area 1 is mainly associated with alteration zones in the Ouarzazate Group rhyolitic formations. This elongated area straddles a system of NE-SW and NNE-SSW faults. These faults probably served as flow paths for hydrothermal mineralization. The documented mineral occurrences also show a close spatial relationship with the fault systems in the study area, particularly in the Tagmout copper deposit.
- Area 2 and 3 are located at the contact zone between the pink granite of Isk n'Alla and the volcanic and rhyodacite formations of the Lower Ouarzazate Group. These formations are crossed by a swarm of rhyolitic dikes of variable direction of the Upper Ouarzazate Group. At outcrop, both zones host some Cu, Mn and Fe showings.
- Area 4 stakes the ENE-WSW Sidi Flah fault zone, which extends over several kilometers. From a mining point of view, several showings, including Au, Cu, Pb and Mn, have been reported along and near this fault. This type of fault may favor the circulation of mineralizing hydrothermal fluids from a deep source that may enrich the host rocks and tectonic structures [22,23].
- Area 5 coincides with hydrothermal alteration associated with the Wawicht granite and volcanic rocks of the Lower Ouarzazate Group.
- Area 6 is associated with the Cryogenian basement formations or its Ediacaran cover. Mineraally, it is a prospective area as it hosts three gold prospects. According to Benzian et al., 2008 [31]; Tuduri et al., 2018 [32], most gold mineralization in the study area is restricted to Precambrian formations such as Cryogenian basement turbid volcano-sedimentary and Ediacaran granites. This indicates that gold prospecting in the areas adjacent to the three indicated prospects is likely to be most successful in the altered zones recognized in basement rocks and its Ediacaran cover.
- Area 7 concerns an alteration zone associated with the Azlag granite and the volcano-sedimentary cover of the lower Ouarzazate group. Due to Miocene deposits that may hide other mineral occurrences, only one copper showing has been reported. Consequently, a geophysical study will be necessary to follow the rooting of these surface anomalies, which would be related to tectonic structures at depth.

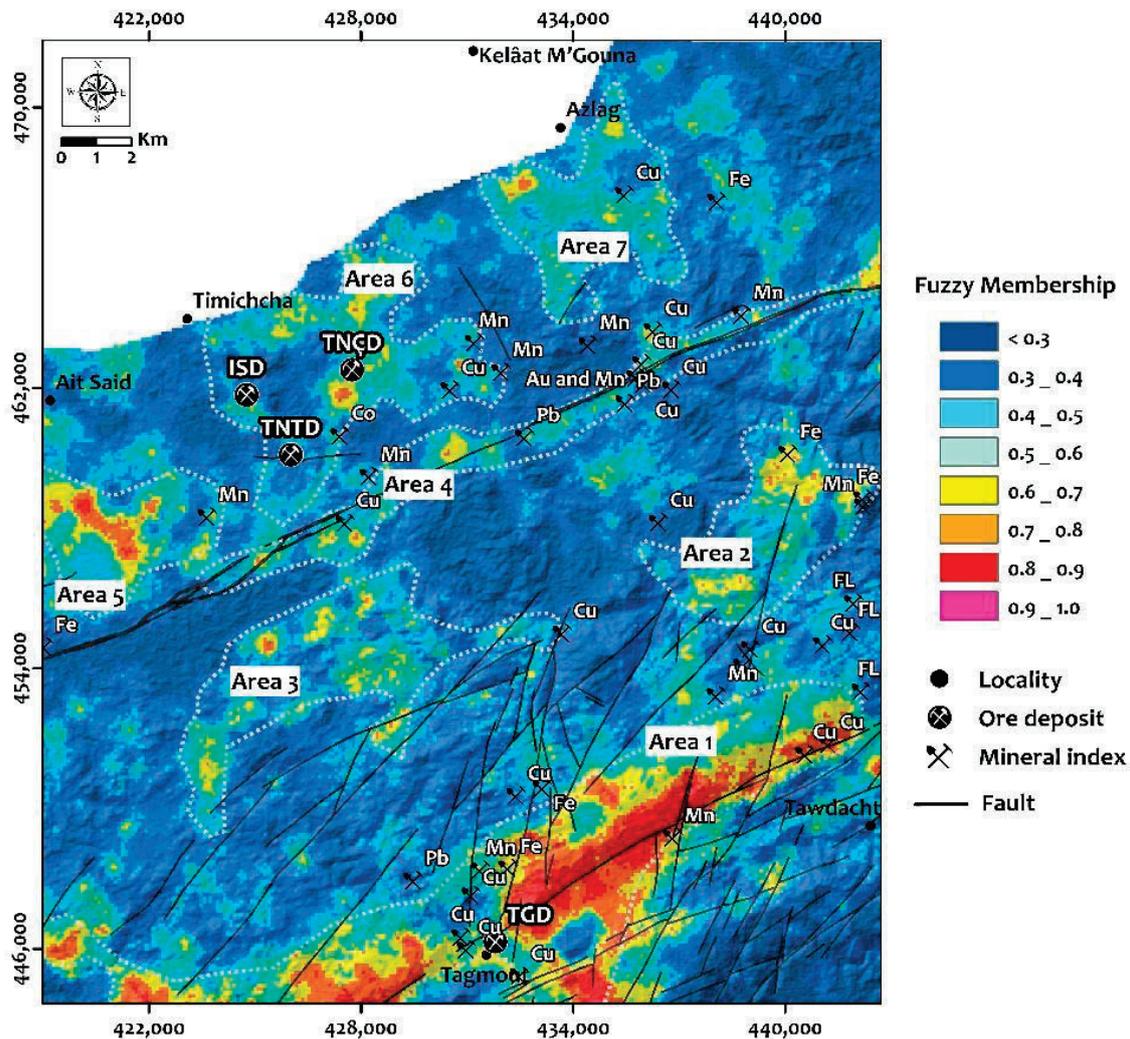


Figure 7. Combination of mineral prospectivity maps.

5. Conclusions

The study demonstrates the importance of combining natural gamma-ray spectrometry and ASTER data in the early stages of mineral exploration. This combination was applied to target areas with high mining potential in the Kelâat M'Gouna inlier. The results obtained allowed detailed mapping of hydrothermal alteration zones related to mineralization. Maps of Potassium (K in %), Uranium (eU in ppm), Thorium (eTh in ppm) and ratios of K/eTh and K/eU were generated to delineate the high concentrations of radioactive elements related to the altered zones, particularly in Potassium.

Band ratios extracted from the ASTER image were calculated to visualize the spatial distribution of specific minerals in the alteration zones. Clay, phyllite, propylitic minerals and iron oxides were mapped in some lithologies that host several mineral occurrences.

The mineral prospectivity maps generated by the fuzzy logic modelling allowed us to locate the alteration zones. Seven anomalous zones were then distinguished. The geological data showed that these zones are located in the contact zones between the granitic massifs, especially the Wawicht, Isk n'Alla and Azlag granites and their host rocks formed by the Ediacaran volcano-sedimentary rocks of the Ouarzazate Group. In addition, most of these zones have been mapped in rocks that host the prospects and, mining showings already indicated, notably those of gold. Other anomalous zones have been mapped in fault zones, mainly in NE-SW and N-S trends, such as the Sidi Flah fault zone and the Tagmout graben zone. To this end, it is recommended to carry out a detailed structural study in conjunction

with geophysics to locate, delineate, and follow the deep rooting of the metallic bodies and tectonic structures that may plug the mineralization in the study area.

Author Contributions: Conceptualization, Y.M., A.A. (Ahmed Attou), A.M. (Abdelhalim Miftah) and M.O.; Data curation, Y.M. and M.O.; Formal analysis, Y.M., A.A. (Ahmed Attou) and M.O.; Funding acquisition, Y.M. and A.M. (Anselme Muzirafuti); Investigation, Y.M.; Methodology, Y.M., M.O., B.D. and Y.E.-t.; Project administration, Y.M. and A.A. (Ahmed Attou); Resources, Y.M., G.R., S.L. and A.M. (Anselme Muzirafuti); Software, Y.M., M.O. and L.A.; Supervision, A.A. (Ahmed Attou) and A.M. (Abdelhalim Miftah); Validation, Y.M., A.A. (Ahmed Attou), A.M. (Abdelhalim Miftah), A.A. (Abdelhamid Allaoui), M.B., G.R., S.L., A.M. (Anselme Muzirafuti), M.O. and B.D.; Visualization, Y.M., A.A. (Ahmed Attou), A.M. (Abdelhalim Miftah), A.M. (Anselme Muzirafuti) and M.O.; Writing—original draft, Y.M. and M.O.; Writing—review and editing, Y.M., A.A. (Ahmed Attou), A.M. (Abdelhalim Miftah), M.O., B.D., L.A., Y.E.-t., A.A. (Abdelhamid Allaoui), M.B., G.R., S.L. and A.M. (Anselme Muzirafuti). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abdelnasser, A.; Kumral, M.; Zoheir, B.; Karaman, M.; Weihed, P. REE geochemical characteristics and satellite-based mapping of hydrothermal alteration in Atud gold deposit, Egypt. *J. Afr. Earth Sci.* **2018**, *145*, 317–330. [\[CrossRef\]](#)
2. Gabr, S.S.; Hassan, S.M.; Sadek, M.F. Prospecting for new gold-bearing alteration zones at El-Hoteib area, South Eastern Desert, Egypt, using remote sensing data analysis. *Ore Geol. Rev.* **2015**, *71*, 1–13. [\[CrossRef\]](#)
3. Mars, J.C.; Rowan, L.C. Spectral assessment of new ASTER SWIR surface reflectance data products for spectroscopic mapping of rocks and minerals. *Remote Sens. Environ.* **2010**, *114*, 2011–2025. [\[CrossRef\]](#)
4. Mars, J.C.; Rowan, L.C. ASTER spectral analysis and lithologic mapping of the Khanneshin carbonatite volcano, Afghanistan. *Geosphere* **2011**, *7*, 276–289. [\[CrossRef\]](#)
5. Pour, A.B.; Hashim, M.; Park, Y.; Hong, J.K. Mapping alteration mineral zones and lithological units in Antarctic regions using spectral bands of ASTER remote sensing data. *Geocarto Int.* **2018**, *33*, 1281–1306. [\[CrossRef\]](#)
6. Rowan, L.C.; Hook, S.J.; Abrams, M.J.; Mars, J.C. Mapping hydrothermally altered rocks at Cuprite, Nevada, using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), a new satellite-imaging system. *Econ. Geol.* **2003**, *98*, 1019–1027. [\[CrossRef\]](#)
7. Rowan, L.C.; Mars, J.C. Lithologic mapping in the Mountain Pass, California area using advanced spaceborne thermal emission and reflection radiometer (ASTER) data. *Remote Sens. Environ.* **2003**, *84*, 350–366. [\[CrossRef\]](#)
8. Sekandari, M.; Masoumi, I.; Pour, A.B.; Muslim, A.M.; Hossain, M.S.; Misra, A. ASTER and WorldView-3 satellite data for mapping lithology and alteration minerals associated with Pb-Zn mineralization. *Geocarto Int.* **2020**, 1–31. [\[CrossRef\]](#)
9. Sheikhrasimi, A.; Pour, A.B.; Pradhan, B.; Zoheir, B. Mapping hydrothermal alteration zones and lineaments associated with orogenic gold mineralization using ASTER data: A case study from the Sanandaj-Sirjan Zone, Iran. *Adv. Sp. Res.* **2019**, *63*, 3315–3332. [\[CrossRef\]](#)
10. Zhang, X.; Pazner, M.; Duke, N. Lithologic and mineral information extraction for gold exploration using ASTER data in the south Chocolate Mountains (California). *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 271–282. [\[CrossRef\]](#)
11. Pour, A.B.; Park, Y.; Park, T.-Y.S.; Hong, J.K.; Hashim, M.; Woo, J.; Ayoobi, I. Regional geology mapping using satellite-based remote sensing approach in Northern Victoria Land, Antarctica. *Polar Sci.* **2018**, *16*, 23–46. [\[CrossRef\]](#)
12. Jébrak, M.; Marcoux, E.; Laithier, M. *Géologie des Ressources Minérales*; Ministère des Ressources Naturelles et de la Faune: Quebec City, QC, Canada, 2008; ISBN 2551237378.
13. Abrams, M. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): Data products for the high spatial resolution imager on NASA's Terra platform. *Int. J. Remote Sens.* **2000**, *21*, 847–859. [\[CrossRef\]](#)
14. Abrams, M.; Yamaguchi, Y. Twenty years of ASTER contributions to lithologic mapping and mineral exploration. *Remote Sens.* **2019**, *11*, 1394. [\[CrossRef\]](#)
15. Ninomiya, Y. Mapping quartz, carbonate minerals, and mafic-ultramafic rocks using remotely sensed multispectral thermal infrared ASTER data. In Proceedings of the Thermosense XXIV, Orlando, FL, USA, 1–5 April 2002; International Society for Optics and Photonics: Bellingham, WA, USA, 2002; Volume 4710, pp. 191–202.
16. Safari, M.; Maghsoudi, A.; Pour, A.B. Application of Landsat-8 and ASTER satellite remote sensing data for porphyry copper exploration: A case study from Shahr-e-Babak, Kerman, south of Iran. *Geocarto Int.* **2018**, *33*, 1186–1201. [\[CrossRef\]](#)

17. Muzirafuti, A.; Boualoul, M.; Barreca, G.; Allaoui, A.; Bouikbane, H.; Lanza, S.; Crupi, A.; Randazzo, G. Fusion of Remote Sensing and Applied Geophysics for Sinkholes Identification in Tabular Middle Atlas of Morocco (the Cause of El Hajeb): Impact on the Protection of Water Resource. *Resources* **2020**, *9*, 51. [\[CrossRef\]](#)
18. Et-Tayea, Y.; Rachid, A.; Attou, A.; Nasri, H.; Mamouch, Y.; El Khazanti, F.; El Azhari, H.; El Haddar, A.; Aziz, A. A new bentonite deposit prospected in the Cap des Trois Fourches area (north-eastern Rif, Morocco) using spectrometry by satellite imagery coupled with mineralogical, chemical, and microstructural investigations. *Arab. J. Geosci.* **2021**, *14*, 1–14. [\[CrossRef\]](#)
19. El Azzab, D.; Ghfir, Y.; Miftah, A. Geological interpretation of the rifian foreland gravity anomalies and 3D modeling of their Hercynian granites (Northeastern Morocco). *J. Afr. Earth Sci.* **2019**, *150*, 584–594. [\[CrossRef\]](#)
20. Elkhateeb, S.O.; Abdellatif, M.A.G. Delineation potential gold mineralization zones in a part of Central Eastern Desert, Egypt using Airborne Magnetic and Radiometric data. *NRIAG J. Astron. Geophys.* **2018**, *7*, 361–376. [\[CrossRef\]](#)
21. Miftah, A.; El Azzab, D.; Attou, A.; Manar, A.; Rachid, A.; Ramhy, H. Geochemical mapping of radioactive elements using helicopter-borne gamma-ray spectrometry (Tiouit, Eastern Anti-Atlas, Morocco): Or occurrence and environmental impact. *J. Afr. Earth Sci.* **2018**, *139*, 392–402. [\[CrossRef\]](#)
22. Miftah, A.; El Azzab, D.; Attou, A.; Rachid, A.; Ouchchen, M.; Soulaïmani, A.; Soulaïmani, S.; Manar, A. Combined analysis of helicopter-borne magnetic and stream sediment geochemical data around an ancient Tiouit gold mine (Eastern Anti-Atlas, Morocco): Geological and mining interpretations. *J. Afr. Earth Sci.* **2021**, *175*, 104093. [\[CrossRef\]](#)
23. Ouchchen, M.; Boutaleb, S.; El Azzab, D.; Abioui, M.; Mickus, K.L.; Miftah, A.; Echogdali, F.Z.; Dadi, B. Structural interpretation of the Igherm region (Western Anti Atlas, Morocco) from an aeromagnetic analysis: Implications for copper exploration. *J. Afr. Earth Sci.* **2021**, *176*, 104140. [\[CrossRef\]](#)
24. Ranjbar, H.; Masoumi, F.; Carranza, E.J.M. Evaluation of geophysics and spaceborne multispectral data for alteration mapping in the Sar Cheshmeh mining area, Iran. *Int. J. Remote Sens.* **2011**, *32*, 3309–3327. [\[CrossRef\]](#)
25. Abd El Nabi, S.H. Role of γ -ray spectrometry in detecting potassic alteration associated with Um Ba'anib granitic gneiss and metasediments, G. Meatiq area, Central Eastern Desert, Egypt. *Arab. J. Geosci.* **2013**, *6*, 1249–1261. [\[CrossRef\]](#)
26. Elawadi, E.; Ammar, A.; Elsirafy, A. Mapping surface geology using airborne gamma-ray spectrometric survey data—A case study. In Proceedings of the 7th SEGJ International Symposium, Sendai, Japan, 24–26 November 2004.
27. Grasty, R.L.; Shives, R.B.K. Applications of gamma ray spectrometry to mineral exploration and geological mapping. In Proceedings of the Workshop Presented at Exploration, Houston, TX, USA, 4–5 October 1997; Volume 97.
28. Maden, N.; Akaryalı, E. Gamma ray spectrometry for recognition of hydrothermal alteration zones related to a low sulfidation epithermal gold mineralization (eastern Pontides, NE Türkiye). *J. Appl. Geophys.* **2015**, *122*, 74–85. [\[CrossRef\]](#)
29. Shives, R.B.K.; Charbonneau, B.W.; Ford, K.L. The detection of potassic alteration by gamma-ray spectrometry—recognition of alteration related to mineralization. *Geophysics* **2000**, *65*, 2001–2011. [\[CrossRef\]](#)
30. Youssef, M.A.S.; Elkhodary, S.T. Utilization of airborne gamma ray spectrometric data for geological mapping, radioactive mineral exploration and environmental monitoring of southeastern Aswan city, South Eastern Desert, Egypt. *Geophys. J. Int.* **2013**, *195*, 1689–1700. [\[CrossRef\]](#)
31. Benziane, F.; Yazidi, A.; Saadane, A.; Yazidi, M.; EL Fahssi, A.; Stone, B.D.; Walsh, G.; Burton, W.C.; Aleinikoff, J.N.; Ejjouani, H.; et al. Carte géologique du Maroc au 1/50 000, feuille KELÂAT M'GOUNA—Notice explicative. *Notes Mémoires Serv. Géologique Maroc* **2008**, *468*, 136.
32. Tuduri, J.; Chauvet, A.; Barbanson, L.; Bourdier, J.-L.; Labriki, M.; Ennaciri, A.; Badra, L.; Dubois, M.; Ennaciri-Leloix, C.; Sizaret, S. The jbel saghro au (–ag, cu) and ag–hg metallogenetic province: Product of a long-lived ediacaran tectono-magmatic evolution in the moroccan anti-atlas. *Minerals* **2018**, *8*, 592. [\[CrossRef\]](#)
33. Thiéblemont, D.; Chêne, F.; Liégeois, J.P.; Ouabadi, A.; Le Gall, B.; Maury, R.C.; Jalludin, M.; Ouattara Gbélé, C.; Tchaméni, R.; Fernandez-Alonso, M. Geological Map of Africa at 1: 10 Million Scale. In Proceedings of the 35th International Geology Congress, Cape Town, South Africa, 28 August–2 September 2016.
34. Gasquet, D.; Ennih, N.; Liégeois, J.-P.; Soulaïmani, A.; Michard, A. The pan-african belt. In *Continental Evolution: The Geology of Morocco*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 33–64.
35. Thomas, R.J.; Chevallier, L.P.; Gresse, P.G.; Harmer, R.E.; Eglington, B.M.; Armstrong, R.A.; De Beer, C.H.; Martini, J.E.J.; De Kock, G.S.; Macey, P.H. Precambrian evolution of the Sirwa window, Anti-Atlas orogen, Morocco. *Precambrian Res.* **2002**, *118*, 1–57. [\[CrossRef\]](#)
36. Hindermeier, J. Le précambrien-III du sarho. *Comptes Rendus Hebd. Séances L'Acad. Sci.* **1953**, *237*, 1024–1026.
37. Ouguir, H.; Macaudière, J.; Dagallier, G.; Qadrouci, A.; Leistel, J. Cadre structural du gîte Ag-Hg d'Imiter (Anti Atlas, Maroc); implication métallogénique. *Bull. Soc. Geol. Fr.* **1994**, *165*, 233–248.
38. Leblanc, M. Ophiolites Précambriennes dans le PII de l'Anti-Atlas Central (Maroc). Ph.D. Thesis, University of Paris VI, Paris, France, 1975.
39. El Baghdadi, M.; El Boukhari, A.; Jouider, A.; Benyoucef, A.; Nadem, S. Calc-alkaline arc I-type granitoid associated with S-type granite in the Pan-African belt of eastern Anti-Atlas (Saghro and Ougnat, South Morocco). *Gondwana Res.* **2003**, *6*, 557–572. [\[CrossRef\]](#)
40. Baidada, B.; Cousens, B.; Alansari, A.; Soulaïmani, A.; Barbey, P.; Ilmen, S.; Ikenne, M. Geochemistry and Sm–Nd isotopic composition of the Imiter Pan-African granitoids (Saghro massif, eastern Anti-Atlas, Morocco): Geotectonic implications. *J. Afr. Earth Sci.* **2017**, *127*, 99–112. [\[CrossRef\]](#)

41. Tuduri, J. Processus de Formation et Relations Spatio-Temporelles des Minéralisations à or et Argent en Contexte Volcanique Précambrien (Jbel Saghro, Anti-Atlas, Maroc). Implications sur les Relations Déformation-Magmatisme-Volcanisme-Hydrothermalisme. Ph.D. Thesis, University of Orléans, Orléans, France, 2005.
42. Walsh, G.J.; Benziane, F.; Aleinikoff, J.N.; Harrison, R.W.; Yazidi, A.; Burton, W.C.; Quick, J.E.; Saadane, A. Neoproterozoic tectonic evolution of the jebel Saghro and Bou Azzer—El Graara inliers, eastern and central Anti-Atlas, Morocco. *Precambrian Res.* **2012**, *216*, 23–62. [[CrossRef](#)]
43. Bajja, A.; Greiling, R.O.; Rocci, G. Pan-African andesites and dacites in the eastern Anti-Atlas: Syn-subduction or post-collisional? *Z. Dtsch. Geol. Ges.* **1998**, 1–12. [[CrossRef](#)]
44. Clauer, N. *Géochimie Isotopique du Strontium des Milieux Sédimentaires. Application à la Géochronologie de la Couverture du Craton Ouest-Africain*; Persée-Portail des Revues Scientifiques en SHS: Aubervilliers, France, 1976; Volume 45.
45. Ezzouhairi, H. Etude Pétrographique, Géochimique et Structurale des Formations Plutoniques du Précambrien II de la Boutonnière de Bouskour (Sarhro Occidental Anti-Atlas, Maroc). Ph.D. Thesis, University of Marrakech, Marrakech, Morocco, 1989.
46. Nerci, K. La Boutonnière Protérozoïque de Sidi Flah (Saghro, Anti-Atlas, Maroc): Contribution à L'étude de ses Minéralisations à Pb, Zn, Cu, Ag et de Cadre Géologique. Ph.D. Thesis, University of Rabat, Rabat, Morocco, 1988; 222p.
47. Rjimati, E.C.; Derre, C.; Lecolle, M.; Lillie, F.; Nerci, K. *Caractéristiques de la tectonique pan-africaine dans le Jbel Saghro (Anti-Atlas)*, Notes et mémoires du Service géologique. 1992; 387–394.
48. Elsass, P. Analyse Tectonique du Graben de Tagmout-Tine Ouayour (Jbel Sarhro, Maroc). *Notes Mem.* **1977**, *268*, 23.
49. Hawkins, M.P.; Beddoe-Stephens, B.; Gillespie, M.R.; Loughlin, S.; Barron, H.F.; Barnes, R.P.; Powell, J.H.; Waters, C.N.; Williams, M. *Carte géologique du Maroc au 1/50 000, feuille Tiwit*, Notes et Mémoires du Service Géologique du Maroc. 2001; Volume 404.
50. Henares Romero, J.; López Casado, C.; Sanz de Galdeano, C.; Delgado, J.; Peláez Montilla, J.A. Stress fields in the Iberian-Maghrebi region. *J. Seismol.* **2003**, *7*, 65–78. [[CrossRef](#)]
51. Derré, C.; Lécolle, M. *Altérations hydrothermales dans le Protérozoïque supérieur du Saghro (Anti-Atlas oriental). Relations avec les minéralisations*, Chronique de la Recherche Minière. 1999; 39–63.
52. Dickson, B.L.; Scott, K.M. Interpretation of aerial gamma-ray surveys-adding the geochemical factors. *AGSO J. Aust. Geol. Geophys.* **1997**, *17*, 187–200.
53. Erdi-Krausz, G.; Matolin, M.; Minty, B.; Nicolet, J.P.; Reford, W.S.; Schetselaar, E.M. *Guidelines for Radioelement Mapping Using Gamma Ray Spectrometry Data: Also as Open Access e-Book*; International Atomic Energy Agency (IAEA): Vienna, Austria, 2003; ISBN 9201083033.
54. Minty, B.R.S. Fundamentals of airborne gamma-ray spectrometry. *AGSO J. Aust. Geol. Geophys.* **1997**, *17*, 39–50.
55. Wilford, J. Airborne gamma-ray spectrometry. *Geophys. Remote Sens. Methods Regioith Explor.* **2002**, 46–52.
56. El-Sadek, M.A. Radiospectrometric and magnetic signatures of a gold mine in Egypt. *J. Appl. Geophys.* **2009**, *67*, 34–43. [[CrossRef](#)]
57. Ghoneim, S.M.; Abd El Nabi, S.H.; Yehia, M.A.; Salem, S.M. Using air-borne gamma ray spectrometry and remote sensing data for detecting alteration zones around Wadi Saqia area, Central Eastern Desert, Egypt. *J. Afr. Earth Sci.* **2021**, *178*, 104181. [[CrossRef](#)]
58. Ribeiro, V.B.; Mantovani, M.S.M. Gamma spectrometric and magnetic interpretation of Cabaçal copper deposit in Mato Grosso (Brazil): Implications for hydrothermal fluids remobilization. *J. Appl. Geophys.* **2016**, *135*, 223–231. [[CrossRef](#)]
59. Adler-Golden, S.; Berk, A.; Bernstein, L.S.; Richtsmeier, S.; Acharya, P.K.; Matthew, M.W.; Anderson, G.P.; Allred, C.L.; Jeong, L.S.; Chetwynd, J.H. FLAASH, a MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations. In Proceedings of the 7th Annual JPL Airborne Earth Science Workshop, Pasadena, CA, USA, 12–16 January 1998; JPL Publication: La Cañada Flintridge, CA, USA, 1998; Volume 97, pp. 9–14.
60. Adler-Golden, S.M.; Matthew, M.W.; Bernstein, L.S.; Levine, R.Y.; Berk, A.; Richtsmeier, S.C.; Acharya, P.K.; Anderson, G.P.; Felde, J.W.; Gardner, J.A. Atmospheric correction for shortwave spectral imagery based on MODTRAN4. In Proceedings of the Imaging Spectrometry V, Denver, CO, USA, 18–23 July 1999; International Society for Optics and Photonics: Bellingham, WA, USA, 1999; Volume 3753, pp. 61–69.
61. Matthew, M.W.; Adler-Golden, S.M.; Berk, A.; Richtsmeier, S.C.; Levine, R.Y.; Bernstein, L.S.; Acharya, P.K.; Anderson, G.P.; Felde, G.W.; Hoke, M.L. Status of atmospheric correction using a MODTRAN4-based algorithm. In Proceedings of the Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, Orlando, FL, USA, 24–28 April 2000; International Society for Optics and Photonics: Bellingham, WA, USA, 2000; Volume 4049, pp. 199–207.
62. Crosta, A.P.; De Souza Filho, C.R.; Azevedo, F.; Brodie, C. Targeting key alteration minerals in epithermal deposits in Patagonia, Argentina, using ASTER imagery and principal component analysis. *Int. J. Remote Sens.* **2003**, *24*, 4233–4240. [[CrossRef](#)]
63. Mia, B.; Fujimitsu, Y. Mapping hydrothermal altered mineral deposits using Landsat 7 ETM+ image in and around Kuju volcano, Kyushu, Japan. *J. Earth Syst. Sci.* **2012**, *121*, 1049–1057. [[CrossRef](#)]
64. Ninomiya, Y. Rock type mapping with indices defined for multispectral thermal infrared ASTER data: Case studies. In Proceedings of the Remote Sensing for Environmental Monitoring, GIS Applications, and Geology II, Crete, Greece, 23–27 September 2002; International Society for Optics and Photonics: Bellingham, WA, USA, 2003; Volume 4886, pp. 123–132.
65. Ninomiya, Y. A stabilized vegetation index and several mineralogic indices defined for ASTER VNIR and SWIR data. In Proceedings of the IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477), Toulouse, France, 21–25 July 2003; Volume 3, pp. 1552–1554.
66. Zhang, N.; Zhou, K. Mineral prospectivity mapping with weights of evidence and fuzzy logic methods. *J. Intell. Fuzzy Syst.* **2015**, *29*, 2639–2651. [[CrossRef](#)]

67. Yamaguchi, Y.; Kahle, A.B.; Tsu, H.; Kawakami, T.; Pniel, M. Overview of advanced spaceborne thermal emission and reflection radiometer (ASTER). *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1062–1071. [[CrossRef](#)]
68. Carranza, E.J.M.; Hale, M. Geologically constrained fuzzy mapping of gold mineralization potential, Baguio district, Philippines. *Nat. Resour. Res.* **2001**, *10*, 125–136. [[CrossRef](#)]
69. Ghanbari, Y.; Hezarkhani, A.; Ataei, M.; Pazand, K. Mineral potential mapping with fuzzy models in the Kerman–Kashmar Tectonic Zone, Central Iran. *Appl. Geomat.* **2012**, *4*, 173–186. [[CrossRef](#)]
70. Zhang, N.; Zhou, K.; Du, X. Application of fuzzy logic and fuzzy AHP to mineral prospectivity mapping of porphyry and hydrothermal vein copper deposits in the Dananhu-Tousuquan island arc, Xinjiang, NW China. *J. Afr. Earth Sci.* **2017**, *128*, 84–96. [[CrossRef](#)]
71. Zadeh, L.A. Fuzzy logic, neural networks, and soft computing. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh*; World Scientific: Singapore, 1996; pp. 775–782.
72. Zimmermann, H.-J.; Zysno, P. Latent connectives in human decision making. *Fuzzy Sets Syst.* **1980**, *4*, 37–51. [[CrossRef](#)]
73. Kim, Y.-H.; Choe, K.-U.; Ri, R.-K. Application of fuzzy logic and geometric average: A Cu sulfide deposits potential mapping case study from Kapsan Basin, DPR Korea. *Ore Geol. Rev.* **2019**, *107*, 239–247. [[CrossRef](#)]
74. Galbraith, J.H.; Saunders, D.F. Rock classification by characteristics of aerial gamma-ray measurements. *J. Geochem. Explor.* **1983**, *18*, 49–73. [[CrossRef](#)]
75. Fall, M.; Baratoux, D.; Ndiaye, P.M.; Jessell, M.; Baratoux, L. Multi-scale distribution of Potassium, Thorium and Uranium in Paleoproterozoic granites from eastern Senegal. *J. Afr. Earth Sci.* **2018**, *148*, 30–51. [[CrossRef](#)]
76. Fall, M.; Baratoux, D.; Jessell, M.; Ndiaye, P.M.; Vanderhaeghe, O.; Moyon, J.F.; Baratoux, L.; Bonzi, W.M.-E. The redistribution of thorium, uranium, potassium by magmatic and hydrothermal processes versus surface processes in the Saraya Batholith (Eastern Senegal): Insights from airborne radiometrics data and topographic roughness. *J. Geochem. Explor.* **2020**, *219*, 106633. [[CrossRef](#)]
77. Hoover, D.B.; Heran, W.D.; Hill, P.L. *The Geophysical Expression of Selected Mineral Deposit Models*; CiteseerX: Princeton, NJ, USA, 1992.

Article

Location Planning of Charging Stations for Electric Buses in Public Transport Considering Vehicle Scheduling: A Variable Neighborhood Search Based Approach

Nils Olsen [†] and Natalia Kliewer ^{*,†}

Department of Information Systems, Freie Universität Berlin, Garystr. 21, 14195 Berlin, Germany; nils.olsen@fu-berlin.de

* Correspondence: natalia.kliewer@fu-berlin.de

† These authors contributed equally to this work.

Abstract: Many public transport companies have recently launched projects testing the operation of electric buses. Progressively, traditional combustion engine buses are being replaced by electric buses. In such cases, some stops on bus lines are equipped with charging technology. Combustion engine buses can operate for an entire day without having to refuel. By contrast, electric buses have considerably shorter ranges and need to recharge their batteries throughout a day. For cost-efficient use of electric buses, charging stations must be located within the road network so that required deadhead trips are as short as possible, but attention must also be paid to construction costs. In contrast to vehicle scheduling, which is a more short-term planning task of public transport companies, location planning of charging stations is a long-term planning problem and requires a simultaneous solving of both optimization problems. Specifically, location planning and vehicle scheduling have to be considered simultaneously in order to open up optimization potentials by comparison to sequential planning, since locations of charging stations directly influence the resulting vehicle rotations. To this purpose, we present a novel solution method for the simultaneous optimization of location planning of charging stations and vehicle scheduling for electric buses in public transport, using variable neighborhood search. By a computational study using real-world public transport data, we show that a simultaneous consideration of both problems is necessary because sequential planning generally leads to either infeasible vehicle rotations or to significant increases in costs. This is especially relevant for public transport companies that start operating electric bus fleets.

Keywords: location planning; vehicle scheduling; electric buses; charging stations; partial charging

Citation: Olsen, N.; Kliewer, N. Location Planning of Charging Stations for Electric Buses in Public Transport Considering Vehicle Scheduling: A Variable Neighborhood Search Based Approach. *Appl. Sci.* **2022**, *12*, 3855. <https://doi.org/10.3390/app12083855>

Academic Editors: Giovanni Randazzo, Anselme Muzirafuti, Dimitrios S. Paraforos and Stefania Lanza

Received: 2 March 2022

Accepted: 8 April 2022

Published: 11 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last years, awareness of climate change and sustainable operations has increased significantly throughout the entire economy and public life. Electromobility is currently considered a highly relevant technology in order to make public transport systems more sustainable and environmentally friendly. Therefore, traditional buses with combustion engines are being progressively replaced by electric buses. Electrically powered buses facilitate a locally emission-free movement which leads to minimal emission levels of greenhouse gases, dust particles, and nitrogen oxides. Seeking to improve the quality of life, especially in congested urban areas, electric buses enable much more quietly operations [1].

At present, the electric energy required for powering electric buses is either provided by batteries or is generated by fuel cells from hydrogen, methanol, or similar fuel [2]. Due to the lower energy density of modern electric batteries compared to common tank capacities for hydrogen or methanol, battery-powered buses involve the greatest challenges for bus operations. For this reason, we focus on battery electric buses (BEBs) within this work. However, the methodology and results of this work can be transferred to any other type of electric engine. We will consider electric bus and battery electric bus as synonyms.

Traditional combustion engine buses can often operate for an entire day without having to refuel. By contrast, modern BEBs have only a fraction of the ranges of combustion engine buses and need to recharge their batteries several times a day [3]. Nowadays, BEBs are charged overnight at vehicle depots after the completion of their daily operations. In addition, the vehicles are charged at charging stations during shorter waiting periods while operating (opportunity charging). Energy transmission occurs either conductively by a wire or inductively. In some cases, the vehicle batteries are also replaced with a fully charged battery (battery swapping).

With a view, for example, to the current real-world bus project at the Schiphol Airport in Amsterdam, the Netherlands, the bus company Connexxion operates with up to 100 BEBs at the present time [4]. Electric VDL Citea buses are operated within this project, with batteries capable of storing 215 kWh which results in a range between 80 and 120 km. The batteries are charged inductively with fast charging systems. Most modern electric buses like the *Irizar ie Bus* are able to store about 350 kWh and may operate up to 17 h in urban bus systems without charging [5].

In recent years, many other public transport companies have launched similar pilot projects testing the operation of BEBs. An overview on current projects is provided by [6]. Most projects initiated aim towards substituting diesel buses with BEBs during the daily services while retaining cost-minimal vehicle rotations. In such cases, charging systems are established at some stops on the bus lines to facilitate the recharging of the vehicle batteries during operation. For a cost-efficient deployment of BEBs, the charging stations must be built within the road network so that deadhead trips are as short as possible or are not necessary at all. Longer deadhead trips increase the operational costs and may lead to higher demands for buses.

Therefore, construction costs for charging stations as well as the buses' purchase and operational costs have to be considered at the planning stage. The planning process of public transport companies consists principally of strategic, tactical, and operational planning tasks, which differ with regard to the time periods considered. Figure 1 provides an overview of the planning process. Strategic planning comprises the network design and line planning. The network design determines stop points and necessary infrastructure, particularly including the distribution of charging stations within the road network. In this scope, specific technical aspects such as energy grids' transmission capacities or restrictions imposed by local conditions may be considered [7,8]. Within the tactical planning, timetables are constructed according to the previously planned lines. Operational planning determines the deployment of vehicles and personnel.

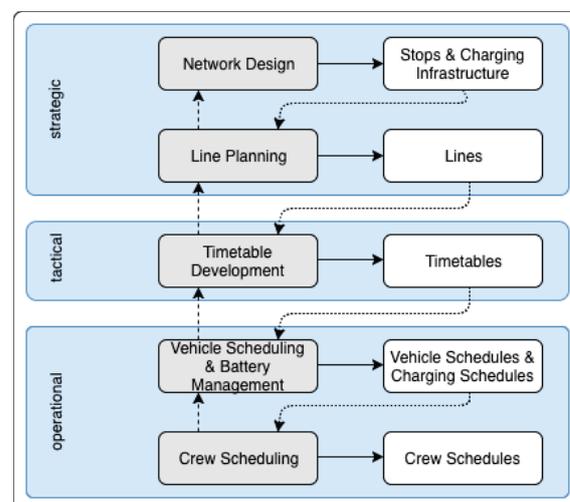


Figure 1. Overview of the planning process arising for companies in public transport when deploying BEBs.

The first operational planning task is vehicle scheduling, which specifies the vehicle deployment for operating service trips offered daily. Service trips denote trips to transport passengers from a departure stop via intermediate stops to an arrival stop at fixed times determined by a timetable. The objective is to assign the set of service trips to vehicles at minimum costs. As part of this task, each service trip must be covered exactly once, each vehicle must execute a feasible sequence of trips (vehicle rotation) without time overlaps, and each vehicle must start and end its rotation at the same depot. This optimization problem commonly refers to the term *Vehicle Scheduling Problem* (VSP). Between successive service trips a vehicle can perform deadhead trips without transporting passengers if necessary. If BEBs are considered within vehicle scheduling, restricted operating ranges due to limited battery capacities and battery charging must be taken into account. This extended optimization problem is commonly denoted as the *Electric Vehicle Scheduling Problem* (E-VSP). While charging, a vehicle stops at a charging station for a specific time period depending on the battery's remaining energy (State of Charge, SoC). Batteries can be either fully or partially charged. The task of determining when, where, and to what extent a battery is charged is denoted as battery management which is closely related to vehicle scheduling.

Unlike vehicle scheduling, which is a more short-term planning task in operational planning, location planning of charging stations is a long-term planning task belonging to strategic network planning and requires a simultaneous optimization of location planning of charging stations and vehicle scheduling for BEBs. Both optimization problems have to be considered simultaneously in order to open up optimization potentials by comparison to sequential planning. At the present time, there are solution approaches to the E-VSP considering fixed locations of charging stations determined in advance, on the one hand. On the other hand, location planning problems for charging stations are being solved to provide for the operation of cost-minimal vehicle rotations computed for buses without range limitations by BEBs. Both approaches belong to a sequential planning.

Simultaneous problem solving is always applicable when a public transport company fully or partially substitutes its fleet of diesel buses with BEBs for the first time. This is particularly the case because charging stations are not usually available within public transport systems yet and need to be built. Furthermore, it is expected that in the future private energy companies will operate networks of charging stations, especially within urban areas, that can be used by vehicles and buses. Some of these networks already exist, such as *E.on Drive* in Germany, but it is expected that such offers will be expanded in the future [9]. In this scenario, each transport company has to pay a usage fee in order to charge a vehicle at specific stations. While location planning of charging stations is a long-term planning problem, vehicle scheduling is carried out every time the timetable changes. However, the simultaneous approach is still applicable because then it is based on the modified timetable and the set of charging stations provided by the energy companies. The construction costs for building a charging station then correspond to the usage fees.

In this paper, we present a novel solution method for the simultaneous optimization of location planning of charging stations and vehicle scheduling for BEBs in public transport to open up potentials for cost savings in comparison with a sequential planning. To do so, we develop a solution approach based on *Variable Neighborhood Search* (VNS), which has been successfully applied to real-world combinatorial optimization problems in a variety of application areas [10]. We propose a heuristic solution approach because the E-VSP and the location planning problem are both difficult to solve, especially with regard to larger instances. Following Lenstra and Rinnooy [11] and Yang and Sun [12], both problems are NP-hard. Simultaneous problem solving is expected to be no less difficult [13]. Within our solution approach we incorporate complete as well as partial charging procedures of the vehicle batteries. By a computational study, we prove the need for simultaneous optimization as opposed to sequential planning. We show that simultaneous problem solving is necessary because sequential planning generally leads to either infeasible vehicle

rotations or to significant increases in costs. Further on, we discern that the incorporation of partial charging procedures leads in principle to major cost savings.

This paper is outlined as follows: In Section 2 we provide an overview of existing work about scheduling of electric vehicles and location planning of charging stations for BEBs. In Section 3 we define the problem to be solved formally. Following this, we introduce the metaheuristic solution method in Section 4. In Section 5 we perform comprehensive computational experiments and analyze the results in order to make key statements. We provide conclusions and present potentials for further research in Section 6.

2. Literature Overview

In this section, we give an overview of related work. As mentioned above, existing work can generally be divided into scheduling of BEBs assuming fixed locations of charging stations and location planning of charging stations for given vehicle rotations. Consequently, we begin by discussing existing solution approaches for scheduling BEBs in public transport. We then present literature on location planning of charging stations.

2.1. Scheduling Electric Buses

As one of the first contributions dealing with alternative engine types within vehicle scheduling, Stasko and Gao [14] present a solution method for the VSP taking into account different engine options. The solution approach is based on integer programming. Engines powered by compressed natural gas (CNG) are considered besides combustion engines. The approach aims at reducing emission levels within vehicle scheduling.

Reuer et al. [15] consider a mixed fleet of vehicles consisting of electrically powered buses and buses without range limitations within the basic VSP. The authors apply a time-space network based exact solution method for the VSP introduced by Kliewer et al. [16] to solve the enhanced optimization problem. Solutions obtained to this problem contain optimal flow values through the network. Therefore, strategies for flow decomposition are necessary to obtain vehicle rotations. The authors analyze six strategies for flow decomposition that aim at maximizing the proportion of feasible vehicle rotations for BEBs. Battery charging is assumed to be performed within constant time periods. The authors show that a simple substitution of traditional buses with BEBs leads to widely infeasible vehicle schedules.

Haghani and Banihashemi [17] consider a fleet consisting entirely of range restricted vehicles. They consider vehicle scheduling with route and time constraints in order to limit the lengths and durations of vehicle rotations. However, battery charging is not considered. The authors propose one exact and two heuristic solution models together with techniques for reducing the problem sizes in order to solve even larger-scale problem instances. Chao and Xiaohong [18] consider battery swapping in addition to limited operating ranges of BEBs within the VSP. To solve the problem, a solution method based on a Non-dominated Sorting Genetic Algorithm (NSGA-II) is introduced. A case study based on real-world data taken from a project in Shanghai is performed to analyze the solution approach. Li [19] addresses vehicle scheduling of BEBs with either battery swapping or charging and presents a model for restricting the maximum route distance. Both fast charging and battery swapping are presumed to be performed within constant time windows, but the time for fast charging depends on the location. Adler and Mirchandani [20] deal with scheduling of BEBs incorporating charging procedures at given charging stations located within the road network. To solve the problem, they present a column-generation approach. A heuristic method is presented to obtain necessary initial solution. The algorithm is based on a greedy algorithm and computes vehicle rotations under consideration of range limitations and charging. In this work, again full chargings of vehicle batteries are assumed.

As one of the first authors, Wen et al. [21] address the E-VSP with partial chargings. They present an exact solution method based on mixed integer programming and an adaptive large neighborhood search heuristic approach. The results demonstrate that the exact solution methods is only applicable to small problem instances. However, the

heuristic solution approach also solves larger instances in a reasonable amount of time. van Kooten Niekerk et al. [22] also consider partial charging procedures of BEBs. The authors introduce a solution approach based on column generation. Charging times depend linearly on a battery's SoC. Furthermore, battery aging and time-dependent energy prices are considered. The authors show that in some cases, the consideration of partial charging procedures leads to cost savings.

Recently, Wang et al. [23] proposed an exact solution method for the E-VSP based on dynamic programming. Within this contribution, battery aging is particularly considered. The objective of the solution method is minimize the total costs especially incorporating costs for battery replacements during the life spans of the vehicles deployed. By a computational study, the authors analyze the influence of different working loads, battery management, and working temperatures of batteries on resulting vehicle schedules.

2.2. Location Planning of Charging Stations for Electric Buses

At the present time, only few publications deal with location planning of charging stations for BEBs in public transport. Kunith et al. [2] present a mixed integer linear optimization model for determining locations for charging stations for a bus route. The model is based on a set covering problem. The objective is to minimize the number of charging stations needed. The authors consider constraints imposed by the buses' operation and the battery charging process. In addition, different energy consumption scenarios are considered to reflect external influencing factors on the buses' energy consumption, such as traffic volume and weather conditions. Standard optimization libraries are used for solving the problem.

Berthold et al. [24] propose a mixed integer linear program in order to determine optimal locations of charging stations for the electrification of a single bus line in Mannheim. The problem is solved by using standard optimization libraries. Furthermore, partial charging procedures and battery aging effects over several time periods are considered. Since the problem is very complex, the solution approach is not suitable for larger instances. Xylia et al. [25] develop a dynamic optimization model to establish a charging infrastructure for BEBs in Stockholm, Sweden, considering restricted waiting times at intermediate stops on service trips given by the schedule and different currents of the charging systems imposed by local conditions. They provide statements about the application possibilities of BEBs in urban areas and effects on vehicle rotations. Within both works, no line changes of the buses used are considered.

Liu et al. [26] consider energy consumption uncertainties within location planning of charging stations for BEBs in public transport. Therefore, the authors propose a robust optimization model represented by a mixed integer linear program. Using real-world data, the authors show that the proposed solution model can provide optimal locations for charging stations that are robust against uncertain energy consumption of BEBs. Lin et al. [27] introduce a spatial-temporal model for a large-scale planning of charging-stations for BEBs in public transport. The authors consider characteristics of BEBs operation and plug-in fast charging technologies. The model is represented by a mixed-integer second-order cone programming formulation with high computational efficiency. A case study using data from Shenzhen, China is used to analyse the robustness of the solution model to timetable changes.

Based on the solution method presented in this paper, Stumpe et al. [13] present an exact mathematical model for integrated optimization of vehicle scheduling with BEBs and location planning for charging stations. The authors particularly perform a robustness analysis and study the impact of technological aspects such as battery capacity, charging power, and energy consumption as well as economic issues containing investment costs for charging stations and electric buses. A computational study points out that the exact solution model introduced is not capable of solving realistic problem instances to optimality.

Regarding related optimization problems in the scope of transportation, there are some contributions dealing with the charging infrastructure for electric vehicles. Regarding

Vehicle Routing Problems (VRP) with electric vehicles, Worley et al. [28] propose a solution approach for the simultaneous determination of optimal locations for charging stations and vehicle routes. They show that this approach leads to lower total costs of the vehicle deployment by comparison to locations of charging stations known a priori. Schiffer and Walther [29] also deal with the simultaneous determination of locations for charging stations and routes for electric vehicles. The authors extend this optimization problem by considering uncertain characteristics of the customers to be served. Uncertain spatial customer distributions, demand, and service time windows are particularly addressed. The authors introduce a robust optimization approach based on adaptive large neighborhood search. Vehicle routing comprises different challenges and conditions than vehicle scheduling and therefore needs other solution approaches. Consequently, it is not possible to draw concrete statements with regard to the E-VSP.

2.3. Summary and Need for Further Research

Table 1 presents the main characteristics of the presented literature. As described there, there is no existing work that deals with scheduling of BEBs and location planning of charging stations simultaneously. However, as underlined by Worley et al. [28] with regard to vehicle routing, a simultaneous optimization opens up potentials for cost savings. It is to be expected that a simultaneous problem solving will also be beneficial for scheduling of BEBs in public transport. In addition, partial charging procedures have not yet been considered sufficiently within the scope of scheduling BEBs. As shown by van Kooten Niekerk et al. [22] for fixed locations of charging stations, the incorporation of partial charging procedures facilitates further optimization potentials. Simultaneous problem solving under consideration of partial charging procedures forms the basic idea of our contribution.

Table 1. Overview of the main characteristics of related literature.

Reference	E-VSP	E-VRP	Mixed veh.fleet	Electric veh.fleet	w/o Line Changes	Fixed chrg.stat.	Fixed veh.rot.	Partial Charging
Stasko and Gao [14]	•		•			•	•	
Haghani and Banihashemi [17]	•			•		•		
Worley et al. [28]		•		•				
Chao and Xiaohong [18]	•			•		•		
Li [19]	•			•		•		
Reuer et al. [15]	•		•			•	•	
Adler and Mirchandani [20]	•			•		•		
Wen et al. [21]	•			•		•		•
Berthold et al. [24]	•			•	•		•	•
van Kooten Niekerk et al. [22]	•			•		•		•
Xyliaa et al. [25]	•			•	•		•	•
Liu et al. [26]	•			•			•	•
Schiffer and Walther [29]		•		•				•
Lin et al. [27]	•			•			•	•
Wang et al. [23]	•			•		•		•
Stumpe et al. [13]	•			•				•

3. Problem Description and Cost Model

In this section, we present the *Electric Vehicle Scheduling Problem with Location Planning of Charging Stations* (E-VSP-LP) as the key problem being solved in this paper. In the following, we first introduce the parameters of the problem. Afterwards, we introduce decision variables and the objective function.

We assume a public transportation network given by a set $S = \{s_1, \dots, s_n\}$ of $n \in \mathbb{N}$ stop points also containing the set of vehicle depots $D \subseteq S$. Service trips are defined by a given timetable as a set $T = \{t_1, \dots, t_m\}$ with $m \in \mathbb{N}$. A service trip $t \in T$ is characterized by its departure and arrival time as well as its departure and arrival stop. For any pair $(s_i, s_j) \in S \times S$ of stop points there is a specific distance and travel time that can be different depending on whether the trip is a service or deadhead trip. In our study, we do not consider opportunity charging of BEBs during the execution of service trips. Consequently, the set S contains the departure and arrival stop of each service trip $t \in T$ as well as the set of depots. The aim is to assign the service trips contained in T to a set of BEBs that are substantially determined by their battery capacities. There may be other specifications such as vehicle dimensions or passenger capacities. Each combination of these features is denoted as a *vehicle type*. To recharge the vehicle batteries, charging stations can be built at each stop point of S . The installed charging system at a charging station considerably influences the time needed for charging. A vehicle can be either fully or partially charged, which also affects the charging time.

For the deployment of a BEB fixed costs $c_{fixed}^{bus} > 0$ incur independently of the executed trips. Each charging or trip operated during a vehicle rotation results in operational costs. Therefore, we consider time costs per hour $c_{time}^{bus} > 0$ and for the distances covered of $c_{distance}^{bus} > 0$. The equipment of stop points with charging technology causes fixed costs $c_{fixed}^{charging} > 0$. These costs may be different, depending on the type of the charging system to be installed or the location. For instance, it is more expensive to build a charging station at a busy crossing than in a quiet side street.

We define decision variables $y_s \in \{0, 1\}$, $\forall s \in S$ and $x_v \in \{0, 1\}$, $\forall v \in V$ denoting the decision whether a charging station is built at stop point s or respectively, whether a vehicle v is used or not. The objective of the simultaneous optimization problem is to minimize the total costs for a given timetable and potential locations of charging stations. Accordingly, fixed costs for BEBs as well as charging stations and operational costs for the buses' operation must be minimized. The objective function can be formulated as

$$\min \underbrace{\sum_{s \in S} c_{fixed}^{charging} \cdot y_s}_{\text{location planning}} + \underbrace{\sum_{v \in V} c_{fixed}^{bus} \cdot x_v}_{\text{vehicle costs}} + \underbrace{\sum_{v \in V} \sum_{t \in v} (c_{time}^{bus} \cdot dur(t) + c_{distance}^{bus} \cdot len(t))}_{\text{operative costs}}. \quad (1)$$

vehiclescheduling

A trip's duration is specified by $dur(t) \geq 0$ and a trip's length by $len(t) \geq 0$. The objective function's value may be interpreted as the total costs caused by a first investment into an electrification of a public transport company's fleet and infrastructure for a specific timetable period. Variable costs for the maintenance of the charging infrastructure or battery replacements are not considered within this work.

In this paper, we solve the E-VSP-LP heuristically as large real-world instances cannot be solved to optimality in an acceptable time [13]. For that reason, we do not present a formal model at this point. However, we refer to Stumpe et al. [13] for a comprehensive mathematical problem formulation and further insights.

4. A Variable Neighborhood Search Based Solution Method for the E-VSP-LP

In this section, we discuss our solution approach for the E-VSP-LP. The objective is to find vehicle rotations for BEBs and locations for charging stations simultaneously and at a minimum cost. We begin by presenting the basic procedure of our heuristic solution method. The solution method consists primarily of generating initial solutions first and then finding new solutions with lower total costs. To do so, we introduce a savings algorithm for generating initial solutions in Section 4.2. Afterwards, we present an algorithm for improvement based on VNS in Section 4.3.

4.1. General Approach

Algorithm 1 provides the main procedure of our solution method. The set of scheduled service trips to be assigned and an initial set of charging stations, together with their locations, serve as the input data. Already existing charging infrastructure, for example due to the implementation of previous pilot projects, may be included in the set of charging stations. Usually, at the beginning of the algorithm the set of charging stations is empty. The algorithm basically consists of two consecutive steps: First, we use a savings algorithm to generate initial sets of vehicle rotations for BEBs and charging stations (l. 1). Subsequently, we use this initial solution as the input for an improvement method based on VNS, which we denote as BVNS (l. 2). The algorithm terminates by returning the best solution found. The two key Algorithms 2 and 3 are explained in the following sections.

Algorithm 1 Main Variable Neighborhood Search.

Input: scheduled service trips T , charging stations S

Output: vehicle rotations V , charging stations \bar{S}

- 1: $(V', S') \leftarrow SA(T, S)$;
 - 2: $(V'', S'') \leftarrow BVNS(V', S')$;
 - 3: **return** V'', S'' ;
-

4.2. Savings Algorithm for Generating Initial Solutions

The savings algorithm was first introduced by Clarke and Wright [30] to solve VRPs heuristically. The objective of vehicle routing is to determine an optimal set of routes seeking to service a number of customers with a fleet of vehicles. Following Cordeau et al. [31], the savings algorithm is one of the most commonly used methods for vehicle routing in practice. Starting from routes each containing one customer the basic procedure is to compute cost savings iteratively for merging two routes into the same one. Within each iteration the merging that results in the highest saving is performed. A saving consists of fixed and operative costs saved. This procedure terminates when no further mergings can be performed. While this algorithm has been applied generally to VRPs, we adapt this algorithm hereinafter in order to apply the same procedure to the E-VSP-LP.

Algorithm 2 shows the procedure for generating initial solutions to the E-VSP-LP formally using the idea of cost savings. The set of scheduled service trips to be assigned and an initial set of charging stations, together with their locations, serve as the input data. The algorithm begins by adding a vehicle rotation for each scheduled service trip, now containing only the associated trip together with a deadhead trip from and to the depot (l. 4). If these vehicle rotations are not feasible for BEBs the entire optimization problem is infeasible. Within each iteration of the algorithm those two vehicle rotations (l. 7 and 8) are merged that lead to a feasible rotation and entail the highest saving. Therefore, the set of service trips of both rotations to be merged are processed consecutively, in order of departure times (l. 9). Since the SoC mostly influences the feasibility of a vehicle rotation besides temporal restrictions the algorithm aims at adding charging procedures as often as possible. For this purpose, starting with a new and empty vehicle rotation (l. 10), four different cases are considered for each service trip of the rotations to be merged. First, we check whether a charging procedure can be performed at an existing charging station of S before executing the current service trip, taking into account necessary deadhead trips (l. 12). If this can be done, necessary deadhead trips, the charging procedure, and the service trip are added (l. 13). If this is not possible, we examine whether the current service trip can be executed without detours to charging stations (l. 14). If the SoC is insufficient, we check whether the current service trip can be executed by building a new charging station at the trip's departure stop and performing a charging procedure (l. 16). Lastly, the same is checked but for the latest position of the vehicle, which is less strict because the deadhead trip is executed after the charging procedure (l. 18). If none of these options can be carried

out, the current merging is aborted (l. 20). When a merging is feasible, the saving for merging two vehicle rotations $v, w \in V$ into a new rotation \bar{v}, \bar{w} is given by

$$s(v, w) = c_{fix}^{bus} - \delta \cdot c_{fix}^{charging} - (o(\bar{v}, \bar{w}) - o(v) - o(w)) \quad (2)$$

where $o(v) \geq 0, \forall v \in V$ denotes the operational costs for each vehicle rotation and $\delta \in \mathbb{N}$ the number of additionally respectively fewer needed charging stations. After each iteration the merging is performed that involves the highest positive saving (l. 25). Then, the set S of charging stations is modified, the new vehicle rotation is added, and the rotations merged are removed (l. 26 and 27). If no positive savings exist, the algorithm terminates and returns the sets of vehicle rotations and charging stations (l. 29). Hence, solutions generated by this procedure are always feasible.

The procedure of Algorithm 2 is based on the heuristic solution method proposed for the E-VSP by Adler and Mirchandani [20]. Within this algorithm, the charging stations are assumed to be known a priori and cannot be changed. However, within Algorithm 2, we extend the procedure from Adler and Mirchandani [20] significantly by incorporating location planning for charging stations.

Algorithm 2 Savings Algorithm (SA).

Input: scheduled service trips T , charging stations S

Output: vehicle rotations V , charging stations \bar{S}

```

1:  $V \leftarrow \emptyset$ 
2:  $\bar{S} \leftarrow S$ 
3: for all  $t \in T$  do
4:   Add a vehicle rotation to  $V$  containing only  $t$ ;
5: end for
6: while TRUE do
7:   for all  $v \in V$  do
8:     for all  $w \in V \setminus \{v\}$  do
9:       Determine the set  $\bar{T}$  of service trips of  $v \cup w$ ;
10:      Create a new vehicle rotation  $\bar{v}$  without trips;
11:      for all  $t \in \bar{T}$  do
12:        if  $\bar{v}$  can be recharged at an existing charging station and execute  $t$  then
13:          Add necessary deadhead trips, charging,  $t$  to  $\bar{v}$ ;
14:        else if  $\bar{v}$  can execute  $t$  then
15:          Add necessary deadhead trips,  $t$  to  $\bar{v}$ ;
16:        else if  $\bar{v}$  can be recharged at the departure stop of  $t$  and execute  $t$  then
17:          Add charging station to  $\bar{S}$ , necessary deadhead trips, charging,  $t$  to  $\bar{v}$ ;
18:        else if  $\bar{v}$  can be recharged at its current position and execute  $t$  then
19:          Add charging station to  $\bar{S}$ , charging, necessary deadhead trips,  $t$  to  $\bar{v}$ ;
20:        else break;
21:      end if
22:    end for
23:  end for
24:  end for
25:  Make move with the highest saving  $s(v, w)$ ;
26:  Remove rotations  $v$  and  $w$  from  $V$ ; Add  $\bar{v}$  to  $V$ ;
27:  Add new charging stations to  $\bar{S}$ ;
28:  if No positive savings exist then
29:    return  $V, \bar{S}$ ;
30:  end if
31: end while

```

4.3. Variable Neighborhood Search for Improvement

To finding new solutions with lower total costs, we use a VNS based solution method. VNS was first introduced by Hansen et al. [10]. Solution approaches based on VNS have already been extensively considered in the literature and have been proven to be suitable for numerous practical problems with realistic data sizes [32]. The underlying concept of VNS is a systematic change of neighborhoods, both in an improvement phase to find a local optimum and in a perturbation phase to escape from local optima. In the perturbation phase, a so-called shaking method is applied, which exerts a stochastic influence on an incumbent solution by performing stochastic changes. Even this procedure can cause a deterioration in the objective function value it has used to escape from local optima. In the improvement phase, a local search method is used to find new solutions with lower total costs.

Adapting the basic VNS concept to solve the E-VSP-LP thus requires the definition of a problem specific neighborhood structure and methods for shaking, a local search, and changing the neighborhood. Algorithm 3 provides the procedure for our solution method. The algorithm follows the *basic* VNS adapted from Hansen et al. [33]. Note that the following procedure is applicable not only for solutions generated by Algorithm 2 but also for every possible feasible solution.

Algorithm 3 Basic Variable Neighborhood Search (BVNS).

Input: vehicle rotations V , charging stations S , t_{max} , k_{max}

Output: vehicle rotations V , charging stations S

```

1:  $t \leftarrow 0$ 
2: while  $t < t_{max}$  do
3:    $k \leftarrow 1$ ;
4:   while  $k \leq k_{max}$  do
5:      $(V', S') \leftarrow \text{SHAKE}(V, S, k)$ ;
6:      $(V'', S'') \leftarrow \text{BESTIMPROVEMENT}(V', S', k)$ ;
7:      $(V, S, k) \leftarrow \text{NEIGHBORHOODCHANGE}((V, S), (V'', S''), k)$ ;
8:   end while
9:    $t \leftarrow \text{CPU TIME}()$ ;
10: end while
11: return  $(V, S)$ ;

```

We first define a neighborhood N_k of size $k \in \mathbb{N}$ by selecting k vehicle rotations. The choice of the vehicle rotations will be made randomly from the entire set in order to incorporate stochastic influences. It follows the maximum neighborhood size $k_{max} \in \mathbb{N}$ as the number of vehicles used within the incumbent solution. After each iteration of shaking and local search, a neighborhood change is performed. In this step, the objective function values of the incumbent and improved solution are compared. If the improved solution is better than the incumbent, it is accepted and the size of the neighborhood is reset to the smallest possible value. Otherwise, the size of the neighborhood is increased and the procedure is repeated. The procedure terminates when the maximum computational time is exceeded. Algorithm 4 shows the procedure formally.

Algorithm 4 NEIGHBORHOODCHANGE.

Input: solutions (V, S) , (V', S') , neighborhood size k , objective function f
Output: solution (V, S) , neighborhood size k

```

1: if  $f(V', S') < f(V, S)$  then
2:    $(V, S) \leftarrow (V', S')$ ;
3:    $k \leftarrow 1$ ;
4: else  $k \leftarrow k + 1$ ;
5: end if
6: return  $(V, S), k$ ;

```

Second, we use Algorithm 5 as the local search method within Algorithm 3 for improving a solution. As the total costs of a solution consist of operational costs for deadheading as well as fixed costs for vehicles and charging stations, Algorithm 5 combines the three following Algorithms 6–8, each aiming towards reducing one cost component. In Algorithm 5, the move is performed that involves the highest cost saving.

Algorithm 5 BESTIMPROVEMENT.

Input: neighborhood N_k , objective function f
Output: neighborhood N_k

```

1: return  $\min_f \{EXST(N_k), SST(N_k), SCP(N_k)\}$ ;

```

Algorithm 6 is used to reduce operational costs for deadheading by exchanging service trips between different vehicle rotations of a corresponding neighborhood. Therefore, a saving is computed for each pair of service trips for the neighborhood's set of vehicles that can be exchanged, and the move with the highest saving is returned.

Algorithm 6 Exchange of Service Trips (EXST).

Input: neighborhood N_k
Output: neighborhood N_k

```

1: for all  $v \in V$  do
2:   for all  $w \in V \setminus \{v\}$  do
3:     for all  $t_v \in v$  do
4:       for all  $t_w \in w$  do
5:         if  $t_v$  and  $t_w$  can be exchanged then
6:           Compute saving;
7:         end if
8:       end for
9:     end for
10:   end for
11: end for
12: Perform exchange with the highest saving;
13: return  $N_k$ ;

```

Algorithm 7 aims at inserting service trips of vehicle rotations with a lower number of service trips into vehicle rotations with a higher number of service trips, again based on a neighborhood. If an insertion is possible, a saving is computed containing proportionate fixed costs for the remaining service trips, fixed costs for additional charging stations, and operational costs for possible detours. Again, the best move found is returned. The algorithm attempts to omit vehicle rotations whereby no service trips are being executed any more.

Algorithm 7 Shift Service Trips (SST).

Input: neighborhood N_k , fixed costs for an BEB c_{fix}^{bus}
Output: neighborhood N_k

- 1: **for all** $v \in V$ **do**
- 2: **for all** $w \in V : |ST_w| < |ST_v|$ **do**
- 3: **for all** $t_w \in w$ **do**
- 4: **if** t_w can be inserted in v **then**
- 5: Compute saving $(c_{fix}^{bus} / |ST_w|)$ less the costs for newly built charging stations
and additional operational costs;
- 6: **end if**
- 7: **end for**
- 8: **end for**
- 9: **end for**
- 10: **end for**
- 11: Perform move with the highest saving, omit a vehicle if no trips are being performed;
- 12: **return** N_k ;

Algorithm 8 aims at decreasing the number of charging stations used by moving charging procedures from less frequented charging stations to higher frequented charging stations, considering the vehicle rotations of a neighborhood. The move is returned that is feasible and entails the highest saving including proportionate fixed costs for remaining charging procedures at a specific charging station and operational costs for additional detours. Similar to Algorithm 7, this procedure aims at omitting charging stations where chargings are no longer being performed at a specific stop point.

Algorithm 8 Shift Charging Procedures (SCP).

Input: neighborhood N_k , charging stations S
Output: neighborhood N_k , charging stations S

- 1: Sort S by the number of charging procedures performed within the entire set of vehicle rotations in ascending order;
- 2: **for** $s = 1$ **to** $|S| - 1$ **do**
- 3: **for all** $t = |S|$ **to** $s + 1$ **do**
- 4: **if** A charging of a vehicle in N_k is performed at s and can be shifted to t **then**
- 5: Compute saving $(c_{fix}^{charging} / |CP_s|)$ less additional operational costs;
- 6: **end if**
- 7: **end for**
- 8: **end for**
- 9: Perform move with the highest saving, omit a charging stations if no chargings are being performed;
- 10: **return** N_k, S ;

While stochastic influences on incumbent solutions are already incorporated by the random selection of a neighborhood's set of vehicles, the Algorithm 9 is applied additionally within Algorithm 3. This approach is intended to enable more stochastic changes to the procedure aiming to escape from local optima. Shaking is based on the procedures given by Algorithms 6–8. Within each method call of Algorithm 9, one of the three algorithms is randomly applied if the corresponding move is feasible. This is done even though the objective function value is being worsened.

Algorithm 9 SHAKE.

Input: neighborhood N_k
Output: neighborhood N_k

- 1: Choose EXST, SST or SCP as f at random;
 - 2: **if** $f(N_k)$ is feasible **then**
 - 3: **return** $f(N_k)$;
 - 4: **else** Go to 1
 - 5: **end if**
-

4.4. Inserting Partial Chargings

In our computational study, which follows this section, we incorporate complete and partial charging procedures. So far, the algorithms presented operate with any kind of charging procedures. However, we need more algorithmic effort in order to incorporate partial chargings within Algorithms 2 and 3. To that purpose, we consider the following Algorithm 10 by Olsen and Kliewer [34]. It is applied to each vehicle rotation that is generated respectively modified within the solution procedure. As a result, Algorithm 10 either returns the set of partial charging procedures that have to be inserted into the corresponding vehicle rotation or its infeasibility. Only if a resulting vehicle rotation is feasible is it taken into further consideration.

Algorithm 10 checks iteratively, after each trip of a rotation, whether the SoC has been violated (l. 2). If this is the case, the previous trips are considered (l. 3). Each trip that begins or ends at a charging station represents a charging opportunity (l. 5). If no such possibilities are found the vehicle rotations is infeasible (l. 9). Over all charging possibilities determined, the one performed at the most highly frequented charging station is processed (l. 11). This aims at reducing the number of charging stations by shifting charging procedures from less to more highly frequented charging stations. In the next step, the vehicle rotation is divided at the specific charging station into two sub-rotations containing the previous and subsequent trips. Then, both sub-rotations are processed by the algorithm. In the case that all sub-rotations are feasible, the algorithm terminates (l. 13). If a charging station is no longer needed it is omitted. If at least one sub-rotation is infeasible, the next charging opportunity is processed (l. 15 and l.16).

Algorithm 10 Inserting Partial Chargings (PCP).**Input:** vehicle rotation v , set S of charging stations**Output:** vehicle rotation v , feasibility or infeasibility of v

```

1: for all  $t_1 \in v$  do
2:   if SoC after executing  $t_1$  is not sufficient then
3:     for all  $t_2 \in v$  previous to  $t_1$  do
4:       if Departure stop is a charging station then
5:         Save charging opportunity;
6:       end if
7:     end for
8:     if Set of charging opportunity is empty then
9:       return  $v$ , infeasible;
10:    end if
11:    Add charging opportunity at the highest frequented charging station;
12:    if Vehicle rotation can be performed then
13:      return  $v$ , feasible;
14:    else
15:      Exclude charging opportunity from the set of all opportunities;
16:      Go to 8;
17:    end if
18:  end if
19: end for

```

5. Computational Analysis

In the following, we perform our computational experiments. We first present the instances to be solved and the problem parameters. Then, we look at the results of a sequential planning approach. In this case, location planning of charging stations and vehicle scheduling of BEBs are solved one by one. Therefore, our analysis is twofold: First, we discuss the results of solving a location planning problem for charging stations to enable the operation of given cost-optimal vehicle rotations computed for traditional buses without the range limitations of BEBs. Second, we present the results of solving an E-VSP given the locations of charging stations computed in the previous step. Last, we analyze the results of simultaneous problem solving using our heuristic solution method provided by Algorithm 3 for the E-VSP-LP and compare the results to the sequential planning approaches. We specifically investigate the impact of considering complete and partial charging procedures on solutions.

5.1. Experimental Design

Our computational experiments are performed on 10 real-world instances that are inspired by real-world public transport data. The instances are characterized by different numbers of stop points and service trips as well as different distributions of service trips over a day. To simplify the analysis, the instances' labels reflect the numbers of service trips and stop points. The instances' distributions of cumulative service trips over the day are presented by Figure 2. The figure shows that the instances differ substantially with regard to the distributions. It is worth mentioning that the last five instances consist of subsets of the service trips taken from instance t3067_s209 for runtime reasons. In the case of instances t1580_s209 and t1487_s209 the original set of service trips was halved randomly, and in the case of instances t1060_s209, t1074_s209, and t933_s209 the set was divided into three parts also in a random way.

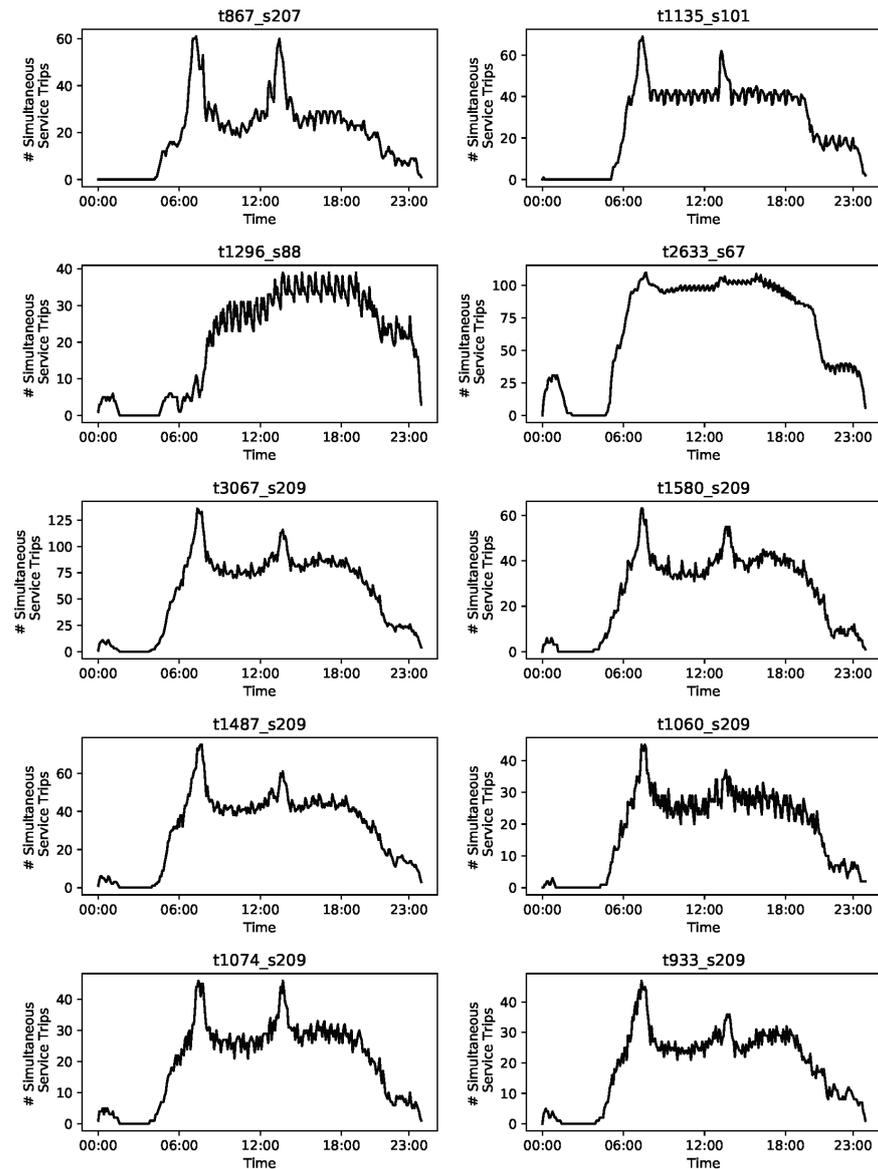


Figure 2. Profiles of cumulative service trips.

Within our experiments, we presume a single vehicle depot, a single vehicle type, and a single charging system. Accordingly, each timetabled service trip can be executed by every available BEB. Additionally, each BEB can charge its battery at every charging station. With regard to the practical implementations of BEBs, we assume that three buses at most can be charged at a charging station at the same time. This is because building sites for charging systems are usually restricted, especially in urban areas. In our study, we distinguish between complete and partial charging procedures. In order to incorporate battery aging, we presume that a battery's SoC ranges between 20% and 80% of a battery's capacity as indicated by Fernandez et al. [35]. In our experiments, we first presume that a vehicle is always charged up to a SoC of 80%. After that, we consider partial chargings. In that regard, the threshold until a battery is charged may vary depending on the idle times at charging stations. Irrespective of the threshold until a battery is charged during its rotation, we assume that a vehicle always begins its rotation with a fully charged battery.

Following Stamati and Bauer [36], charging modern batteries is a nonlinear and therefore complex procedure. The current during a charging process is of particular importance. As demonstrated by Olsen and Kliever [34], the current decreases quickly

when a battery is charged to over 80% of its capacity. Below this threshold, the current is almost constant. For that reason, we assume a constant current and thus linear charging times for vehicle batteries within this paper. We assume that 5 kWh can be transferred into a vehicle battery per minute. In our study we consider chargings before the start or after the end of service trips. To reflect the lower consumption of BEBs on deadhead trips we therefore assume a consumption of 1.5 kWh per kilometer and of 1.8 kWh per kilometer driving on service trips. These parameters are inspired by the data of the previously introduced project at the Schiphol Airport in Amsterdam. At present, there is a wide range of battery capacities offered on the market that range between approximately 60 and 300 kWh. Based on this, we consider different battery capacities of 60, 120, 300 and 500 kWh within our experiments. A battery capacity of 500 kWh may be considered as a future development of battery research. Since we consider only one vehicle type at the same time, we conduct our study for each capacity. Based on Stamati and Bauer [36], a BEB in use and equipped with a 60-kWh battery causes fixed costs of about 350,000 monetary units. Measured by the battery sizes this results to fixed costs for the other vehicles of 365,000, 405,000, and 450,000 monetary units. With regard to the operational costs, we presume 0.5 units per driven kilometer and 50 units per hour of operation. Again based on the bus project in Amsterdam, the equipment of a stop point with charging technology is incorporated with fixed costs of 200,000 monetary units. We use the term "monetary units" here since we assume that these units are roughly comparable—at least in terms of scale—and, based on this, that monetary units form a system of imputed cost components.

5.2. Location Planning of Charging Stations for the Electrification of Cost-Minimal Vehicle Rotations, Computed without Range Limitations

We begin our computational analysis by discussing the results of solving a location planning problem for charging stations for the electrification of given cost-minimal vehicle rotations computed without range limitations. The vehicle rotations were generated using the exact optimization method for the traditional VSP by Kliewer et al. [16], which is based on a time-space network. In order to enable the operation of these rotations by BEBs, charging stations are added to the network and charging procedures are inserted into the vehicle rotations. Partial charging procedures are performed, since the idle times at potential charging stations are given by the vehicle rotations. The objective is to maximize the proportion of vehicle rotations that are feasible for BEBs. Ideally, this procedure should ensure the holistic operation of the timetabled service trips by BEBs. For this purpose, we adapt the location planning problem for charging stations introduced by Berthold et al. [24] and solve it using standard optimization libraries.

Table 2 provides the results of solving a location planning problem for charging stations, containing the proportions and absolute numbers of feasible vehicle rotations for BEBs together with the numbers and proportions of charging stations needed for each instance and each battery capacity. Additionally, the optimal number of vehicles used is indicated when no range limitations are considered. If the totality of all vehicle rotations is feasible for BEBs, the operational and total costs are specified for subsequent analyses. First, we observe that in the vast majority of cases the holistic electrification of vehicle rotations by means of inserting charging procedures is not possible. It is apparent that this observation holds regardless of the instance to be solved. However, the proportion of feasible vehicle rotations grows with increasing battery capacities. We can observe that every instance can be entirely served by BEBs in the case of a battery capacity of 500 kWh. In some cases, this situation already occurs with a battery capacity of 300 kWh and in a single case with 120 kWh. However, none of the instances can be entirely served by BEBs with a battery capacity of 60 kWh. Regarding a battery capacity of 60 kWh, the proportions of feasible vehicle rotations fluctuates widely and ranges between 7.25% and 79.63%. With regard to charging stations, it can be concluded that the numbers of stop points equipped with charging technology decreases significantly when the battery capacities grow. Instance t1296_s88 shows the biggest reduction in the number of charging stations needed from

48.86% to 6.81%. The operational costs of feasible vehicle rotations decrease slightly when the battery capacities grow, which can be attributed to fewer charging procedures required.

In summary, a sequential planning solving at first a standard VSP without incorporating the special features of BEBs and subsequently a location planning problem for charging stations is generally insufficient, leading to widely infeasible solutions. This approach is only suitable if the ranges of BEBs rise sharply in the future. The results obtained serve as lower bounds for the numbers of BEBs used and as an upper bound for the numbers of charging stations needed in the evaluation of the simultaneous solution approach.

Table 2. Results of solving a location planning problem for charging stations for the electrification of cost-minimal vehicle rotations computed without range restrictions incorporating partial charging procedure.

Instance	Battery Capacity (kWh)	# Vehicles	# Stations	Operat. Costs (Mio)	Feasible Rotations	Total Costs (Mio)
t876_s207	60	69	47 (22.71%)	-	5 (7.25%)	-
	120	69	44 (21.25%)	-	31 (44.93%)	-
	300	69	33 (15.94%)	-	62 (89.86%)	-
	500	69	7 (3.38%)	1.127	69 (100%)	33.93
t1135_s101	60	75	33 (32.67%)	-	43 (57.33%)	-
	120	75	27 (26.73%)	-	69 (92%)	-
	300	75	15 (14.85%)	1.351	75 (100%)	35.48
	500	75	2 (1.98%)	1.349	75 (100%)	35.60
t1296_s88	60	47	43 (48.86%)	-	28 (59.68%)	-
	120	47	32 (36.36%)	-	42 (80.37%)	-
	300	47	25 (28.40%)	-	42 (80.37%)	-
	500	47	6 (6.81%)	0.114	47 (100%)	22.76
t2633_s67	60	125	29 (43.28%)	-	74 (58.4%)	-
	120	125	21 (32.34%)	-	80 (64%)	-
	300	125	17 (25.37%)	-	117 (93.6%)	-
	500	125	8 (11.94%)	2.652	125 (100%)	60.91
t3067_s209	60	165	90 (43.06%)	-	88 (53.33%)	-
	120	165	69 (33.01%)	-	154 (93.33%)	-
	300	165	39 (18.66%)	-	162 (96.97%)	-
	500	165	14 (6.69%)	3.045	165 (100%)	80.79
t1580_s209	60	75	55 (26.31%)	-	39 (52%)	-
	120	75	45 (21.53%)	-	61 (81.33%)	-
	300	75	20 (9.56%)	1.342	75 (100%)	36.71
	500	75	7 (3.34%)	1.319	75 (100%)	36.82
t1487_s209	60	89	53 (25.35%)	-	46 (51.79%)	-
	120	89	37 (17.71%)	-	87 (97.76%)	-
	300	89	24 (11.48%)	1.696	89 (100%)	43.74
	500	89	7 (3.34%)	1.672	89 (100%)	43.47
t1060_s209	60	54	43 (20.57%)	-	43 (79.63%)	-
	120	54	30 (14.35%)	0.988	54 (100%)	28.20
	300	54	13 (6.22%)	0.987	54 (100%)	26.11
	500	54	3 (1.43%)	0.985	54 (100%)	26.04
t1074_s209	60	56	39 (18.66%)	-	31 (55.36%)	-
	120	56	33 (15.78%)	-	52 (92.86%)	-
	300	56	16 (7.65%)	0.986	56 (100%)	27.67
	500	56	6 (2.87%)	0.985	56 (100%)	27.69
t933_s209	60	54	35 (16.74%)	-	23 (42.60%)	-
	120	54	25 (11.96%)	-	32 (77.78%)	-
	300	54	15 (7.17%)	0.971	54 (100%)	26.59
	500	54	4 (1.91%)	0.963	54 (100%)	26.26

5.3. Scheduling of Electric Buses Given Fixed Locations of Charging Stations

We now discuss the results of solving an E-VSP with given locations of charging stations. The set of charging stations determined by the previous experiment within Section 5.2 serves as the input, since this set is already optimal if corresponding solutions are feasible for BEBs. Following Section 1, the objective of the E-VSP is to minimize the

number of buses in use and the operational costs for deadheading while covering each service trip. In order to ensure comparability, partial chargings are performed. Because the E-VSP is NP-hard and exact solution methods are not suitable for solving large real-world instances in general, as in our experiments, we solve the E-VSP heuristically here.

To do so, we use our main solution method from Algorithm 1 in a reduced version. Within both Algorithms 2 and 3, which represent the main components of Algorithm 1, we disable modifications of the charging stations. Within Algorithm 2, we only allow the assignment of service trips to vehicles without charging or with detours to existing charging stations. The other two cases are omitted. Within Algorithm 3, we modify the Algorithms 5 and 9 by disabling Algorithm 8 within each procedure. This approach means that the set of charging stations cannot change in this experiment. While the following results are not necessarily optimal due to the heuristic solving, we show that they provide reasonable bounds for our analysis within the next section.

An overview of the results of this experiment is given by Table 3, providing the numbers of vehicles used as well as operational and total costs. The number of charging stations is taken from the previous experiment. In contrast to that, now each solution is feasible, which was to be expected because of the constraints imposed by the E-VSP. Consequently, the total costs are specified for each instance and each battery capacity, containing fixed costs for buses used and charging stations as well as operational costs. First of all, the results show that in most cases where feasible vehicle rotations were computed in the first experiment described in Section 5.2, the solving of an E-VSP provides similar results regarding the numbers of vehicles used and total costs. In some cases, the number of vehicles required is slightly higher than in the first experiment, which can be traced back to the heuristic solution approach. Furthermore, the operational costs are marginally increased. However, the solutions of this experiment converge towards the optimal solutions and thus provide a reasonable benchmark for subsequent analyses. Regarding the numbers of vehicles used, one can observe that the fewer the proportions of feasible vehicle rotations determined within the first experiment, the more vehicles are required when solving the E-VSP. This is understandable because the closely-timed service trips of the vehicle rotations when no range limitations are considered do not provide enough time for rechargings. This leads to an increasing demand for vehicles. For example, considering instance t1580_s209, the optimal numbers of vehicles used is obtained for battery capacities of 500 kWh and 300 kWh. As the proportion of feasible vehicle rotations reduces rapidly for 120 kWh and 60 kWh within the first experiment (81.33% respectively 52%), the need for additional vehicles rises significantly (6 respectively 12 additional vehicles). Regarding the operational costs, we note that higher demands for vehicles generally leads to decreasing operational costs. This is because less deadhead trips and chargings have to be performed due to the shorter rotations.

In conclusion, solving an E-VSP with given locations of charging stations always leads to feasible vehicle rotations, which is in contrast to the first experiment. However, this solution approach generally entails increases in costs due to additional deadhead trips, likely leading to increasing demands for vehicles. The results obtained serve as upper bounds for the analysis of the simultaneous problem solving to be conducted in the following section.

Table 3. Results of solving an E-VSP given locations of charging stations by reduced Algorithm 1 incorporating partial charging procedures.

Instance	Battery Capacity (kWh)	# Vehicles	# Stations	Operat. Costs (Mio)	Total Costs (Mio)
t876_s207	60	80 (+10)	47	1.019	40.77
	120	75 (+6)	44	1.060	39.72
	300	72 (+3)	33	1.098	38.51
	500	69 (+0)	7	1.157	33.96
t1135_s101	60	87 (+12)	33	1.187	39.89
	120	82 (+7)	27	1.219	37.90
	300	76 (+1)	15	1.265	35.79
	500	75 (+0)	2	1.379	35.62
t1296_s88	60	61 (+14)	43	0.082	32.18
	120	49 (+3)	32	0.112	28.16
	300	49 (+3)	25	0.108	25.79
	500	47 (+0)	6	0.116	22.77
t2633_s67	60	144 (+19)	29	2.193	59.84
	120	137 (+12)	21	2.438	57.69
	300	131 (+6)	17	2.514	59.82
	500	126 (+1)	8	2.621	61.32
t3067_s209	60	179 (+14)	90	2.681	87.83
	120	171 (+6)	69	2.820	82.49
	300	169 (+4)	39	2.871	81.07
	500	166 (+1)	14	2.994	81.19
t1580_s209	60	87 (+12)	55	1.133	45.33
	120	81 (+6)	45	1.289	42.10
	300	75 (+0)	20	1.367	36.74
	500	75 (+0)	7	1.323	36.83
t1487_s209	60	101 (+12)	53	1.421	50.02
	120	92 (+3)	37	1.573	44.40
	300	90 (+1)	24	1.682	44.13
	500	89 (+0)	7	1.753	43.53
t1060_s209	60	59 (+5)	43	0.952	32.35
	120	55 (+1)	30	0.991	28.56
	300	54 (+0)	13	0.989	26.11
	500	54 (+0)	3	0.986	26.04
t1074_s20	60	64 (+8)	39	0.897	33.05
	120	59 (+3)	33	0.971	30.76
	300	57 (+1)	16	0.988	28.07
	500	56 (+0)	6	0.994	27.69
t933_s209	60	64 (+10)	35	0.956	32.06
	120	58 (+4)	25	0.961	28.38
	300	55 (+1)	15	0.970	26.99
	500	54 (+0)	4	0.969	26.27

5.4. Simultaneous Optimization of Vehicle Scheduling and Charging Infrastructure

We now discuss the results of simultaneous optimization of scheduling of BEBs and location planning for charging stations, i.e., solving the E-VSP-LP, using our solution method given by Algorithm 1. We begin by presenting the results obtained by Algorithm 2 for generating initial solutions. Then, we discuss the results of Algorithm 3 for finding new solutions with lower total costs. In this experiment we consider complete as well as partial charging procedures in order to enable a comparison with the previous experiments. We conclude this chapter by a runtime analysis.

5.4.1. Summary of Results for Generating Initial Solutions

Table 4 provides the results of using Algorithm 2 for generating initial solutions containing feasible sets of vehicle rotations and charging stations. The results contain the total and operational costs as well as the numbers of buses and charging stations used for each instance and each battery capacity. Additionally, the differences in the total costs

are specified when enabling partial charging procedures, and the best solutions found are in bold.

Table 4. Results of Algorithm 2 for generating initial vehicle rotations for electric buses and locations for charging stations considering complete and partial charging procedures.

Instance	Battery Cap.(kWh)	Complete Chargings				Partial Chargings			
		# Vehicles	# Stations	Operat. Costs (Mio)	Tot. Costs Costs (Mio)	# Vehicles	# Stations	Operat.	Tot. Costs
t876_s207	60	90	2	1.620	33.62	86	2	1.621	32.22 (−1.40)
	120	76	1	1.392	29.38	75	1	1.397	29.02 (−0.36)
	300	76	1	1.322	32.35	75	1	1.322	31.94 (−0.40)
	500	76	1	1.307	35.75	75	1	1.307	35.31 (−0.45)
t1135_s101	60	107	1	1.990	39.69	104	2	1.991	38.89 (−0.80)
	120	94	2	1.644	36.45	92	3	1.644	35.97 (−0.48)
	300	91	1	1.528	38.63	89	2	1.529	38.07 (−0.56)
	500	80	1	1.493	37.74	79	1	1.501	37.30 (−0.44)
t1296_s88	60	86	6	0.729	32.33	82	7	0.730	31.18 (−1.15)
	120	74	3	0.487	28.25	71	4	0.489	27.40 (−0.84)
	300	64	1	0.408	26.58	62	1	0.412	25.77 (−0.81)
	500	58	1	0.384	26.73	56	1	0.391	25.84 (−0.89)
t2633_s67	60	148	18	3.818	60.11	151	16	3.709	60.56 (+0.44)
	120	144	16	3.307	59.87	146	14	3.292	60.08 (+0.21)
	300	139	12	2.978	62.27	141	11	2.787	62.64 (+0.37)
	500	134	6	2.892	64.69	135	5	2.815	64.82 (+0.12)
t3067_s209	60	182	50	3.618	79.81	180	46	3.621	78.12 (−1.70)
	120	178	48	3.346	80.31	175	43	3.378	78.00 (−2.31)
	300	174	36	3.114	82.58	171	33	3.164	80.67 (−1.91)
	500	171	12	3.087	83.03	169	12	3.096	82.14 (−0.89)
t1580_s209	60	108	5	1.966	41.01	109	2	1.721	40.37 (−0.65)
	120	98	1	1.601	37.62	98	1	1.583	37.60 (−0.02)
	300	91	1	1.474	38.58	91	1	1.462	38.57 (−0.01)
	500	87	1	1.287	40.68	87	1	1.276	40.67 (−0.01)
t1487_s209	60	124	4	2.464	46.86	118	4	2.464	44.76 (−2.10)
	120	102	1	1.940	39.42	99	2	1.940	38.58 (−0.85)
	300	102	1	1.797	43.36	98	2	1.792	41.98 (−1.38)
	500	98	1	1.752	46.10	95	2	1.751	45.01 (−1.10)
t1060_s209	60	82	2	1.490	30.69	78	3	1.493	29.54 (−1.15)
	120	66	2	1.218	25.81	64	3	1.219	25.33 (−0.48)
	300	63	1	1.132	26.89	61	2	1.134	26.34 (−0.56)
	500	60	1	1.112	28.36	57	2	1.121	27.27 (−1.09)
t1074_s209	60	86	2	1.496	32.09	85	5	1.499	32.49 (+0.40)
	120	72	1	1.216	27.74	71	4	1.218	28.13 (+0.39)
	300	72	1	1.132	30.54	71	3	1.194	30.69 (+0.16)
	500	67	1	1.105	31.50	66	3	1.184	31.63 (+0.13)
t933_s209	60	81	6	1.527	31.37	82	7	1.498	31.95 (+0.57)
	120	66	1	1.171	25.51	67	2	1.169	26.12 (+0.61)
	300	65	1	1.081	27.66	66	2	1.089	28.32 (+0.66)
	500	60	1	1.044	28.29	61	2	1.075	29.03 (+0.73)

We first compare the results to the first experiment conducted and described in Section 5.2. We observe that in two of the 17 cases, when the first experiment lead to feasible vehicle rotations, the total costs obtained by the application of the savings algorithm were already lower by comparison to solving a location planning problem for charging stations. In the other cases, higher total costs are obtained. In general, the higher total costs arise from higher demands for vehicles needed within the savings algorithm. Regarding each instance, the numbers of vehicles used has increased, which is reasonable due to the heuristic solution procedure of the savings algorithm. The solving of instance *t1296_s88* leads to the highest increase of 23.4%. By contrast, the number of charging stations used decreases in every case. In some cases, such as instance *t1060_s209*, the number of charging stations needed is enormously reduced (30 to two). However, since the costs for additional vehicles prevail over the cost savings arising from the lower number of charging stations used, the total costs increase. This holds true both for complete and partial charging procedures. Regarding these two charging procedures, the total costs obtained are lower in seven of the ten instances for all battery capacities when partial

charging procedures are enabled. On average, total cost savings of about 1.2% are achieved. Only in three cases are the total costs higher when considering partial chargings.

We now compare the initial solutions with the results obtained and described in Section 5.3. With regard to the total costs, our observations are twofold: In those cases in which the solving of a location planning problem led to infeasible vehicle rotations, the application of Algorithm 2 leads to lower total costs by comparison to the results obtained by solving an E-VSP. In the other cases where feasible solutions were obtained, the total costs are higher, arising from a higher demand for vehicles needed as indicated previously. Basically, the results computed by Algorithm 2 merely serve as the input for improvement methods and thus do not serve as the final results. For this reason, the clarified statements are not particularly significant. In the next section, we present the results of improvement using our solution approach based on VNS.

5.4.2. Summary of Results for Variable Neighborhood Search for Improvement

In order to carry out a final comparison between sequential planning and simultaneous problem solving, we now present the results of our solution method given by Algorithm 3 for finding new solutions with lower total costs. We use the initial solutions presented in the previous section as the input data. Table 5 shows the results, containing numbers of vehicles and charging stations used, as well as operational and total costs for each instance and each battery capacity. Additionally, the differences in the total costs are specified when enabling partial charging procedures, and the best solutions found are in bold.

Again, we first compare the results to solving a location planning problem for charging stations at given vehicle rotations. In those cases, where feasible solutions were computed and shown in Section 5.2, the total costs obtained by applying Algorithm 3 are almost of the same quality. In some cases, the total costs are slightly higher, which is most likely due to the heuristic solving. However, in certain scenarios, even better solutions are achieved which can be explained by the utilization of the degrees of freedom. Simultaneous problem solving enables shorter and fewer deadhead trips to charging stations, leading to lower operational and fixed costs for vehicles. This effect would be intensified if exact solution methods were used. As the sequential planning approach leads mostly to infeasible solutions, the simultaneous problem solving is generally preferable.

We now discuss the results with regard to solving an E-VSP with given locations of charging stations as carried out and described in Section 5.3. The most significant observation is that the total costs obtained by the simultaneous problem solving are always below the results of solving an E-VSP with fixed charging stations. This holds true for each combination of instance and battery capacity. The primary reasons for this are that the VNS based approach leads either to the same or slightly higher numbers of vehicles. Similarly, considerably lower numbers of charging stations needed are achieved due to the simultaneous solution procedure, leading to significant cost savings. Additionally, the operational costs are reduced for the most part, which can be explained by the shorter deadhead trips to charging stations. As the cost savings exceed the increased costs for additional vehicles, the solutions generated entail significantly lower total costs. It is interesting to observe that the greatest costs savings are achieved for instances that contain peak times of cumulative service trips over the day. This can be explained by the fact that peak times of service trips over the day allow the vehicles to recharge their batteries during times with reduced offers. In conclusion, simultaneous problem solving enables significant cost savings and is always preferable to solving an E-VSP with given locations of charging stations.

Table 5. Results of Algorithm 3 containing vehicle schedules for electric buses and charging infrastructure after 100.000 iterations considering complete and partial charging procedures.

Instance	Battery Cap.(kWh)	Complete Chargings				Partial Chargings			
		# Vehicles	# Stations	Operat. Costs (Mio)	Tot. Costs Costs (Mio)	# Vehicles	# Stations	Operat.	Tot. Costs
t876_s207	60	79	6	1.317	30.46	77	4	1.348	29.29 (−1.17)
	120	76	3	1.334	30.57	75	2	1.392	29.26 (−1.31)
	300	74	3	1.317	32.03	73	2	1.322	31.38 (−0.65)
	500	73	2	1.254	34.60	72	1	1.277	33.92 (−0.68)
t1135_s101	60	86	31	1.592	39.44	85	30	1.617	38.86 (−0.58)
	120	81	22	1.512	36.57	79	20	1.555	35.38 (−1.19)
	300	77	13	1.617	36.05	76	13	1.656	35.68 (−0.37)
	500	76	2	1.267	35.96	75	2	1.293	35.54 (−0.42)
t1296_s88	60	58	37	0.089	26.63	56	32	0.089	27.68 (−1.95)
	120	49	24	0.112	23.99	49	21	0.112	23.24 (−0.75)
	300	49	21	0.108	25.20	48	20	0.110	24.55 (−0.65)
	500	48	9	0.112	23.96	48	7	0.113	23.46 (−0.50)
t2633_s67	60	139	21	2.217	56.11	138	19	2.219	55.26 (−0.85)
	120	136	18	2.445	56.58	135	16	2.450	55.72 (−0.86)
	300	130	16	2.528	59.17	129	14	2.534	58.27 (−0.90)
	500	128	7	1.609	61.95	127	6	1.617	61.26 (−0.69)
t3067_s209	60	182	48	2.627	78.32	177	36	2.694	73.64 (−4.68)
	120	172	37	2.796	74.82	170	27	2.809	71.60 (−3.22)
	300	171	26	2.854	78.60	169	18	2.819	75.76 (−2.84)
	500	169	12	2.894	81.94	167	11	2.937	80.83 (−1.11)
t1580_s209	60	80	41	1.698	39.94	79	39	1.706	39.10 (−0.84)
	120	79	41	1.751	40.83	78	36	1.754	39.22 (−1.61)
	300	77	14	1.318	36.00	76	12	1.324	35.10 (−0.90)
	500	75	8	1.317	37.06	75	7	1.318	36.81 (−0.25)
t1487_s209	60	99	38	1.448	45.59	96	31	1.451	42.81 (−2.80)
	120	92	31	1.567	42.89	91	24	1.569	40.78 (−2.11)
	300	92	23	1.534	44.54	90	19	1.561	42.76 (−1.78)
	500	90	6	1.494	43.49	89	6	1.533	43.08 (−0.41)
t1060_s209	60	59	37	0.951	30.85	57	31	0.958	28.65 (−2.19)
	120	56	30	0.982	28.92	56	27	0.983	28.17 (−0.75)
	300	55	15	0.983	27.00	54	13	0.988	26.10 (−0.90)
	500	55	2	0.984	26.23	54	3	0.985	26.03 (−0.20)
t1074_s209	60	64	26	0.913	29.81	62	23	0.912	28.36 (−1.45)
	120	59	19	0.963	27.24	57	19	0.968	26.52 (−0.73)
	300	57	14	0.981	27.56	56	16	0.981	27.66 (+0.00)
	500	56	7	0.982	27.93	56	4	0.983	27.18 (−0.75)
t933_s209	60	61	27	0.939	29.03	60	24	0.941	27.94 (−1.10)
	120	58	23	0.948	27.86	56	19	0.951	26.14 (−1.73)
	300	55	15	0.964	26.98	54	15	0.970	26.59 (−0.40)
	500	55	4	0.959	26.70	54	4	0.962	26.26 (−0.45)

Lastly, we investigate the impact of enabling partial charging procedures within vehicle rotations. The results clearly specify that the incorporation of partial chargings is more realistic and opens up optimization potentials. The number of vehicles as well as charging stations used is lower in almost all cases. This leads to significant cost savings up to 4.68% compared to the best solution found for one of the two sequential approaches. On average, savings of 1.17% over all instances and battery capacities can be observed. The same total costs are achieved in only one case. Furthermore, the more vehicles are used, the higher the cost savings are. For this reason, the cost savings generally decrease when the battery capacities increase.

It is worth noting that the clarified statements would also hold true for exact solution methods for the E-VSP-LP. Exact solving would even strengthen the results because of the expected lower total costs. Figure 3 illustrates the key statements made within this chapter. The figure provides an overview of the total costs obtained by the different solution approaches presented for the instances t1060_s209, t1135_s101, and t3067_s209 and for all battery capacities. The instances are chosen among all instances presented since they cover characteristic problem sizes and distributions of cumulative service trips over the day. Comparable behavior is to be expected for instances with similar characteristics not shown here. It is important to note that the total costs are only specified for feasible solutions.

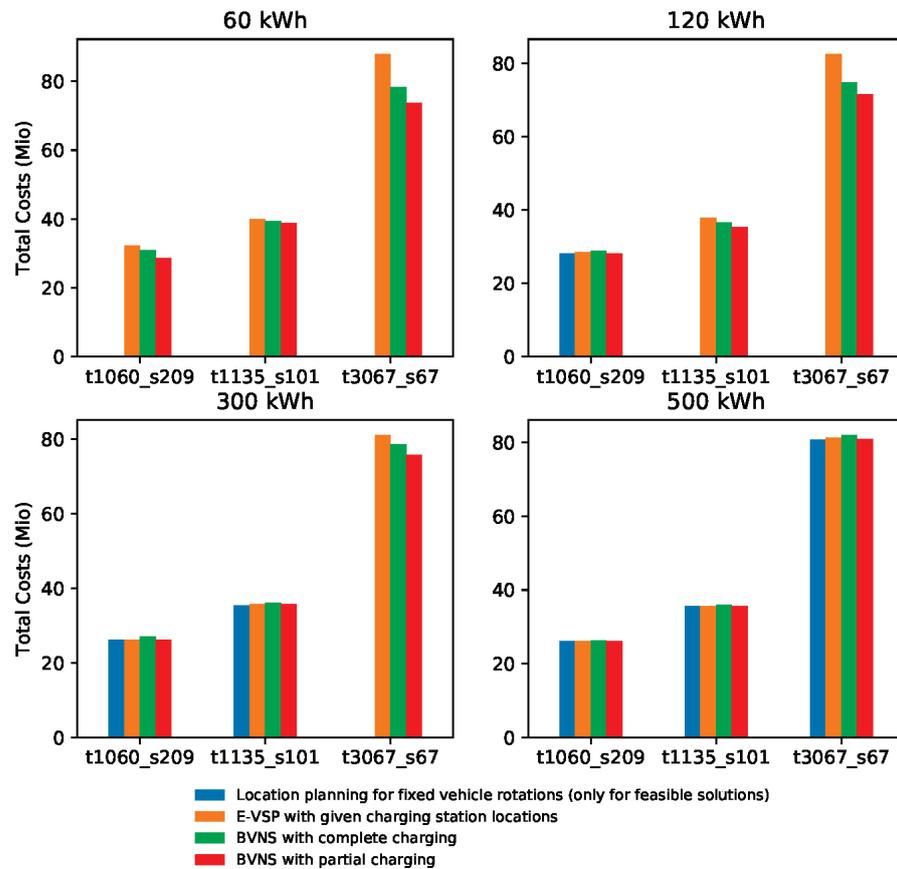


Figure 3. Overview of total costs obtained by the different solution approaches presented for all battery capacities.

5.4.3. Convergence Analysis

The experiments were performed on a common desktop computer (Intel(R) Core(TM) i7-6700 HQ @CPU 2.60GHz 2.59GHz, 16GB RAM). The solution method is implemented in Java. The computational analysis was carried out using Python 3.10.

Figure 4 provides an overview of the convergence behaviour for all problem instances solved by Algorithm 3. In order to facilitate comparison between the different instances, the total costs obtained are normalized. Each figure contains data for the first 20.000 runs. For none of the instances solved a total run time of 10 h was exceeded. The results basically prove reasonable convergence behaviours towards the minimum total costs for all instances. However, particular differences between the instances can be observed. The lower the number of service trips, the faster the total costs obtained by Algorithm 3 decrease. It is noteworthy that the number of stop points has no visible influence on the speed of convergence.

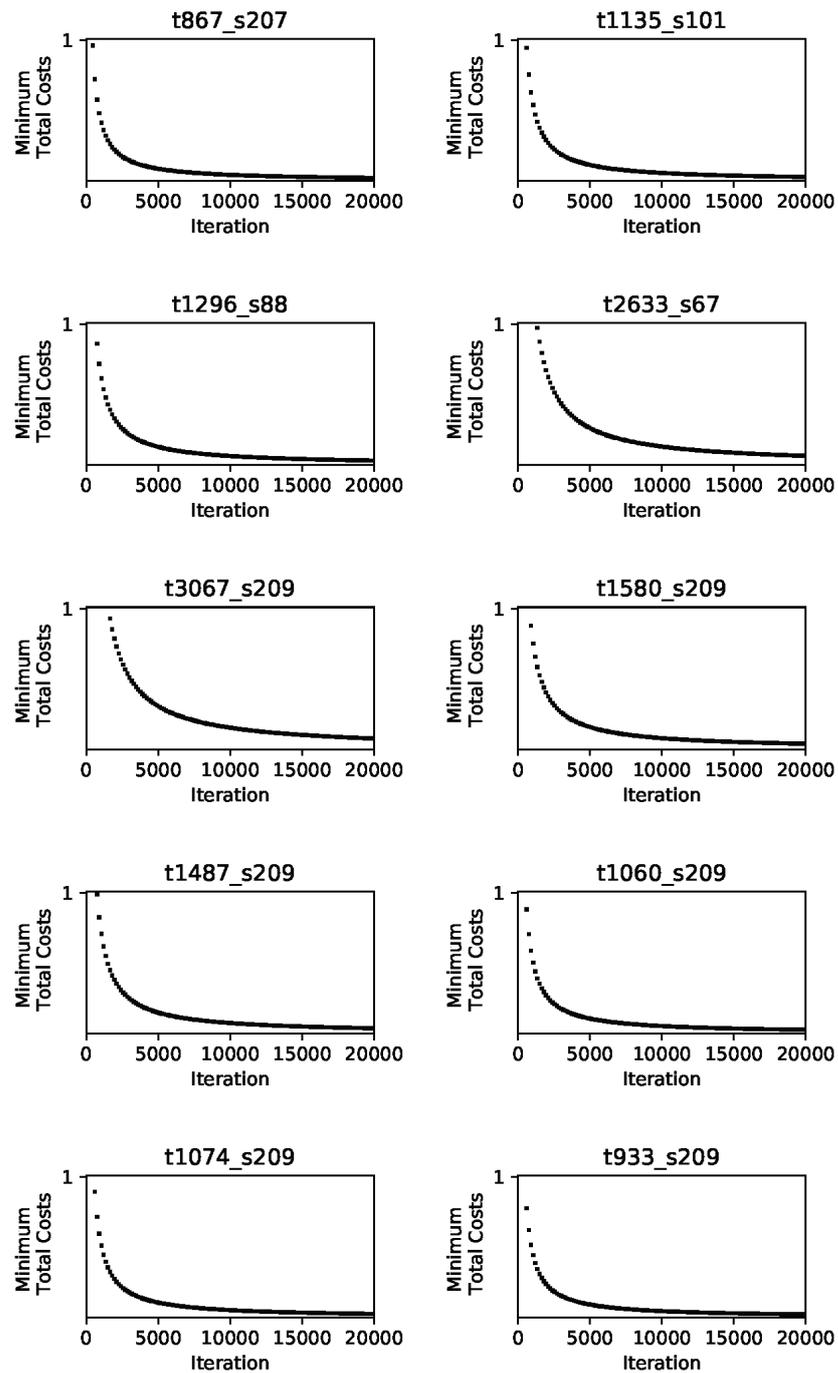


Figure 4. Overview of convergence behaviour for all problem instances solved by Algorithm 3.

6. Conclusions

We have introduced a novel solution method for simultaneous optimization of location planning of charging stations and vehicle scheduling for BEBs in public transport. To do so, we introduced the E-VSP-LP, which extends the standard E-VSP to incorporate location planning of charging stations. To solve the problem we developed a metaheuristic solution method based on VNS, as both problems are difficult to solve. To generate the necessary initial solutions we adapted the traditional savings algorithm. To evaluate the solution approach we performed a computational study based on real-world public transport data, with up to 3000 service trips and different battery capacities of the buses deployed. We also focused on a consideration of complete and partial battery charging procedures of the

batteries within vehicle rotations. In our study we compared the simultaneous solution approach to sequential planning to tackle the underlying problems.

Our experiments showed that simultaneous solving of location planning of charging stations and vehicle scheduling of BEBs is necessary as opposed to sequential planning. First, we demonstrated that sequential planning, first solving a standard VSP and afterwards a location planning problem for charging stations, generally leads to infeasible vehicle rotations for BEBs with regard to current battery technologies. Second, solving an E-VSP with given locations of charging stations entails significant increases in costs. Solving the E-VSP-LP, on the one hand, ensures the feasibility of the vehicle rotations. On the other hand, significantly lower total costs are achieved by comparison to solving an E-VSP, due to the higher degrees of freedom. This is particularly relevant for public transport companies that start operating electric bus fleets. With regard to complete and partial battery chargings, we found large cost savings in most cases when enabling partial chargings within the vehicle rotations.

Our paper can be extended by the following aspects. First, the proposed models do not deal with multiple depots. Incorporating this extension would most likely open up further potentials for cost savings, as already shown for the traditional VSP. Second, our solution method solves the E-VSP-LP heuristically. Exact solution approaches would be interesting for a better verification of the quality of heuristic solution methods. In addition, an interesting path for future research would be to develop additional algorithms for the generation of initial solutions as well as for improvement. Finally, more accurate models regarding the technical aspects of BEBs may be considered. It is conceivable to presume uncertain energy consumptions that may depend on weather conditions or the volume of traffic.

Author Contributions: Conceptualization, N.O. and N.K.; methodology, N.O. and N.K.; software, N.O. and N.K.; validation, N.O. and N.K.; formal analysis, N.O. and N.K.; investigation, N.O. and N.K.; resources, N.O. and N.K.; data curation, N.O. and N.K.; writing—original draft preparation, N.O. and N.K.; writing—review and editing, N.O. and N.K.; visualization, N.O. and N.K.; supervision, N.K.; project administration, N.K.; funding acquisition, N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been financially supported by the German Research Foundation [grant KL 2152/5-1]. The publication of this article was funded by Freie Universität Berlin.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Requia, W.J.; Mohamed, M.; Higgins, C.D.; Arain, A.; Ferguson, M. How clean are electric vehicles? Evidence-based review of the effects of electric mobility on air pollutants, greenhouse gas emissions and human health. *Atmos. Environ.* **2018**, *185*, 64–77. [\[CrossRef\]](#)
2. Kunitz, A.; Goehlich, D.; Mendelevitch, R. Planning and Optimization of a Fast Charging Infrastructure for Electric Urban Bus Systems. In Proceedings of the 3rd International Conference on Traffic and Transport Engineering, Lisbon, Portugal, 17–18 April 2014; pp. 43–51.
3. Deilami, S.; Muyeen, S.M. An Insight into Practical Solutions for Electric Vehicle Charging in Smart Grid. *Energies* **2020**, *13*, 1545. [\[CrossRef\]](#)
4. Connexion-Electric Bus Fleet at Schiphol Airport Benefits from Smart Control. The Mobility House. 2022. Available online: https://www.mobilityhouse.com/int_en/our-references/connexion-electric-bus-fleet (accessed on 6 April 2022).
5. Irizar ie Bus. Irizar e-Mobility. 2022. Available online: <https://irizar-emobility.com/en/vehicles/irizar-ie-bus> (accessed on 6 April 2022).
6. E-Bus-Projekte in Deutschland. VDV Die Verkehrsunternehmen. 2022. Available online: <https://www.vdv.de/e-bus-projekt.aspx> (accessed on 6 April 2022).

7. Alonso, M.; Amaris, H.; Germain, J.; Galan, J. Optimal Charging Scheduling of Electric Vehicles in Smart Grids by Heuristic Algorithms. *Energies* **2014**, *7*, 2449–2475. [CrossRef]
8. Märkle-Huß, J.; Feuerriegel, S.; Neumann, D. Cost minimization of large-scale infrastructure for electricity generation and transmission. *Omega* **2020**, *96*, 102071. [CrossRef]
9. Electric Car Sharing Points for Business. eon. 2021. Available online: <https://www.eonenergy.com/for-your-business/large-energy-users/energy-solutions/electric-vehicle-charging> (accessed on 7 June 2021).
10. Hansen, P.; Mladenovic, N.; Perez, J.A.M. Variable neighbourhood search: Methods and applications. *Ann. Oper. Res.* **2010**, *175*, 367–407. [CrossRef]
11. Lenstra, J.K.; Rinnooy, A.H.G.K. Complexity of vehicle routing and scheduling problems. *Networks* **1981**, *11*, 221–227. [CrossRef]
12. Yang, J.; Sun, H. Battery swap station location-routing problem with capacitated electric vehicles. *Comput. Oper. Res.* **2015**, *55*, 217–232. [CrossRef]
13. Stumpe, M.; Rössler, D.; Schryen, G.; Kliewer, N. Study on Sensitivity of Electric Bus Systems under Simultaneous Optimization of Charging Infrastructure and Vehicle Schedules. *Euro J. Transp. Logist.* **2021**, *10*, 100049. [CrossRef]
14. Stasko, T.; Gao, H. Reducing transit fleet emissions through vehicle retrofits, replacements, and usage changes over multiple time periods. *Transp. Res. Part D Transp. Environ.* **2010**, *15*, 254–262. [CrossRef]
15. Reuer, J.; Kliewer, N.; Wolbeck, L. The electric vehicle scheduling problem—a study on time-space network based and heuristic solution approaches. In Proceedings of the Conference on Advanced Systems in Public Transport (CASPT), Rotterdam, The Netherlands, 19–23 July 2015.
16. Kliewer, N.; Mellouli, T.; Suhl, L. A time–space network based exact optimization model for multi-depot bus scheduling. *Eur. J. Oper. Res.* **2006**, *175*, 1616–1627. [CrossRef]
17. Haghani, A.; Banihashemi, M. Heuristic approaches for solving large-scale bus transit vehicle scheduling problem with route time constraints. *Transp. Res. Part A* **2002**, *36*, 309–333. [CrossRef]
18. Chao, Z.; Xiaohong, C. Optimizing Battery Electric Bus Transit Vehicle Scheduling with Battery Exchanging: Model and Case Study. *Intell. Integr. Sustain. Multimodal Transp. Syst.-Procedia-Soc. Behav. Sci.* **2013**, *13*, 2725–2736. [CrossRef]
19. Li, J.Q. Transit Bus Scheduling with Limited Energy. *Transp. Sci.* **2014**, *48*, 521–539. [CrossRef]
20. Adler, J.D.; Mirchandani, P.B. The Vehicle Scheduling Problem for Fleets with Alternative-Fuel Vehicles. *Transp. Sci.* **2017**, *51*, 441–456. [CrossRef]
21. Wen, M.; Linde, E.; Ropke, S.; Mirchandani, P.; Larsen, A. An adaptive large neighborhood search heuristic for the Electric Vehicle Scheduling Problem. *Comput. Oper. Res.* **2016**, *76*, 73–83. [CrossRef]
22. van Kooten Niekerk, M.E.; van den Akker, J.M.; Hoogeveen, J.A. Scheduling Electric Vehicles. *Public Transp.* **2017**, *9*, 155–176. [CrossRef]
23. Wang, J.; Kang, L.; Liu, Y. Optimal scheduling for electric bus fleets based on dynamic programming approach by considering battery capacity fade. *Renew. Sustain. Energy Rev.* **2020**, *130*, 109978. [CrossRef]
24. Berthold, K.; Förster, P.; Rohrbeck, B. Location Planning of Charging Stations for Electric City Buses. In *Operations Research Proceedings*; Springer: Cham, Switzerland, 2015; pp. 237–242. [CrossRef]
25. Xyliaa, M.; Leducc, S.; Patrizioci, P.; Silveiraa, S.; Kraxnerc, F. Developing a dynamic optimization model for electric bus charging infrastructure. *Transp. Res. Procedia* **2017**, *27*, 776–783. [CrossRef]
26. Liu, Z.; Song, Z.; He, Y. Planning of Fast-Charging Stations for a Battery Electric Bus System under Energy Consumption Uncertainty. *Transp. Res. Rec.* **2018**, *2672*, 96–107. [CrossRef]
27. Lin, Y.; Zhang, K.; Shen, Z.J.M.; Ye, B.; Miao, L. Multistage large-scale charging station planning for electric buses considering transportation network and power grid. *Transp. Res. Part C Emerg. Technol.* **2019**, *107*, 423–443. [CrossRef]
28. Worley, O.; Klabjan, D.; Sweda, T.M. Simultaneous vehicle routing and charging station siting for commercial electric vehicles. In Proceedings of the IEEE International Electric Vehicle Conference (IEVC), Greenville, SC, USA, 4–8 March 2012. [CrossRef]
29. Schiffer, M.; Walther, G. Strategic planning of electric logistics fleet networks: A robust location-routing approach. *Omega* **2018**, *80*, 31–42. [CrossRef]
30. Clarke, G.; Wright, J.W. Scheduling of vehicles from a central delivery depot to a number of delivery points. *Oper. Res.* **1964**, *12*, 546–581. [CrossRef]
31. Cordeau, J.F.; Laporte, G.; Savelsbergh, M.; Vigo, D. Vehicle Routing. *Transp. Handb. Oper. Res. Manag. Sci.* **2007**, *14*, 195–224.
32. Brimberg, J.; Hansen, P.; Mladenovic, N.; Taillard, E. Improvements and comparison for heuristics for solving the multisource Weber problem. *Oper. Res.* **2000**, *48*, 444–460. [CrossRef]
33. Hansen, P.; Mladenovic, N.; Brimberg, J.; Perez, J.A.M. Variable Neighborhood Search. In *Handbook of Metaheuristics*; International Series in Operations Research & Management Science; Gendreau, M., Potvin, J.Y., Eds.; Springer: Boston, MA, USA, 2010; Volume 146. [CrossRef]
34. Olsen, N.; Kliewer, N. Scheduling Electric Buses in Public Transport: Modeling of the Charging Process and Analysis of Assumptions. *Logist. Res.* **2020**, *13*, 4.
35. Fernandez, I.J.; Calvillo, C.F.; Sanchez-Mirallas, A.; Boal, J. Capacity fade and aging models for electric batteries and optimal charging strategy for electric vehicles. *Energy* **2013**, *60*, 35–43. [CrossRef]
36. Stamati, T.; Bauer, P. On-road charging of electric vehicles. In Proceedings of the IEEE Transportation Electrification Conference and Expo (ITEC), Detroit, MI, USA, 16–19 June 2013; pp. 1–8. [CrossRef]

Article

Digitalizing Maritime Containers Shipping Companies: Impacts on Their Processes

Pedro-Luis Sanchez-Gonzalez *, David Díaz-Gutiérrez and Luis R. Núñez-Rivas

Escuela Técnica Superior de Ingenieros Navales, Universidad Politécnica de Madrid (UPM), Av. Arco de la Victoria, 4, 28040 Madrid, Spain; david.diaz@upm.es (D.D.-G.); luisramon.nunez@upm.es (L.R.N.-R.)

* Correspondence: pl.sanchez@alumnos.upm.es

Abstract: Key analysts are emphasizing the importance of the digitalization especially of the supply chain. This work aims to improve maritime shipping companies by introducing digitalization in their operations. This objective is achieved analyzing the impact of maritime container shipping companies' digitalization. This analysis requires as input the Business Process Model (BPMo) and an inventory of digital applications to verify how the BPMo changes when deploying the applications, define the prerequisites necessary for this deployment, and identify the key performance indicators (KPIs) to track it. The impact of the deployment of the applications has been quantified by using four performance dimensions: Costs, Time, Quality, and Flexibility. The results show that the impacts are different per application, with changes in the processes, the addition of new ones, and the decommissioning of others. The impact of digitalization is high when trying to deploy all the applications at the same time. Companies can leverage this work, which requires reviewing the documented impacts in their processes and the applications' prerequisites as well as updating their existing balanced scorecard, incorporating the application's KPIs. A list of 10 applications has been identified as "quick wins"; then, applications can be the starting point for digitalizing a company.

Keywords: digitalization; BPM; business process model; artificial intelligence; big data; virtual reality; internet of things; cloud computing; digital security; additive engineering

Citation: Sanchez-Gonzalez, P.-L.; Díaz-Gutiérrez, D.; Núñez-Rivas, L.R. Digitalizing Maritime Containers Shipping Companies: Impacts on Their Processes. *Appl. Sci.* **2022**, *12*, 2532. <https://doi.org/10.3390/app12052532>

Academic Editor: Vicente Julian

Received: 4 February 2022

Accepted: 26 February 2022

Published: 28 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Maritime transportation defied the COVID-19 disruption, laying the foundations for a transformation in global supply chains [1]. Maritime shipping companies form the backbone of maritime transportation; therefore, they have been forced to changes to follow the global chain changes.

In this context, the key analysts are emphasizing the importance of the digitalization especially of the supply chain, with high investments in artificial intelligence, real-time transportation visibility, etc. [2]. Given its relevance and their intermodal global operations, maritime transportation industry digitalization is key for the supply chain's digitalization.

Digitalization in the maritime transportation industry is being studied these days following different streams: Munim et al. focused on big data and artificial intelligence [3]; Plaza-Hernández studied the integration of IoT technologies in the industry [4]; Kapidani et al. looked at the industry digitalization from a sustainability point of view [5]; Kapnissis et al. investigated blockchain adoption in the industry [6]; Tijan et al. reviewed the drivers, success factors, and barriers to digital transformation in the maritime transport sector [7]. These articles are just a few examples that illustrate the relevance of digitalization research in the maritime transportation industry.

The research from the team of the present paper published in 2019 in *Sensors* (ISSN 1424-8220) showed that when looking at the three different industrial sectors that compose the maritime transportation industry (ship design and shipbuilding; shipping; and ports), its digitalization is moving at different speeds in the different domains and industrial sectors defined in the aforementioned *Sensors* paper:

- Autonomous vehicles and robotics (hereafter, robotics).
- Artificial intelligence (AI).
- Big data.
- Virtual, augmented, and mixed reality (VR).
- The internet of things (IoT).
- Cloud and edge computing (hereafter, the cloud).
- Digital security.
- Three-dimensional (3D) printing and additive engineering (3DP).

The size of the maritime transportation industry makes it necessary to focus on one of their industrial sectors; therefore, this work is limited to shipping.

Any change to the operations of maritime shipping companies requires understanding of how they operate. Business Process Management (BPM) is the science that monitors how work is performed in an organization in order to ensure consistent outcomes and to take advantage of opportunities for improvement [8]; this makes BPM an optimal technique for understanding maritime shipping companies' operations.

Few published works make use of BPM for analyzing the maritime transportation sector. Lyridis et al. [9] made use of BPM to optimize operations of a shipping company for one specific route. Elbert et al. [10] resorted to BPM for ports optimization, thereby analyzing the chains taking place at ports when ships arrive or depart and the interactions with ground organizations. Cimino et al. [11] also relied on BPM for analyzing the impact of Information and Communication Technology (ICT) for ports optimization. Finally, Nikitakos et al. [12] partially used BPM in part to evaluate ICTs in the Greek-owned shipping sector.

The research being presented in this article aims to improve maritime shipping companies by introducing digitalization in their operations while being aware that the implementation of a successful business process model does not automatically bring about the same benefits for all companies [13] but rather is a starting point for understanding the problems. Given the importance of maritime container shipping companies for the maritime transportation industry, this research focuses on these companies. Since there are different types of maritime container shipping companies, those used in this study are companies that have their own fleet of vessels used both nautically and commercially by the company.

The contributions of this work are as follows:

1. To contribute to the digitalization of the industry via the analysis of the impacts of digitalizing the aforementioned process model;
2. To generate the key performance indicators (KPIs) that will allow a phased approach for the deployment of the processes' digitalization;
3. To identify a list of "quick wins": applications that given their optimal results on the analysis could be considered as the starting point for digitalizing a company.

This work is divided into the following sections: Section 2 describes the methodology used in the study; Section 3 includes the results of the impact of maritime container shipping companies processes' digitalization as well as its analysis and discussion; and finally, Section 4, summarizes the conclusions.

2. Approach and Methodology

Since the hypothesis that needs to be proved is that the impacts of maritime containers shipping companies' digitalization is different per application and that these applications can be grouped or clustered according to their impact in the company's operations, the first step was performing an impact analysis. The impact analysis performed in this work required two inputs: the maritime containers shipping companies' Business Process Model (BPMo) and the digital applications used for digitalizing the BPMo.

The lack of published process models for the companies that are the object of this research has required the development of a BPMo. The developed BPMo departed from the Cross-Industry version of the Process Classification Framework© [14,15] from the

American Productivity & Quality Centre (APQC) since there is no version for maritime shipping companies. Figure 1 has the “look and feel” from the APQC, which was used as a starting process model.

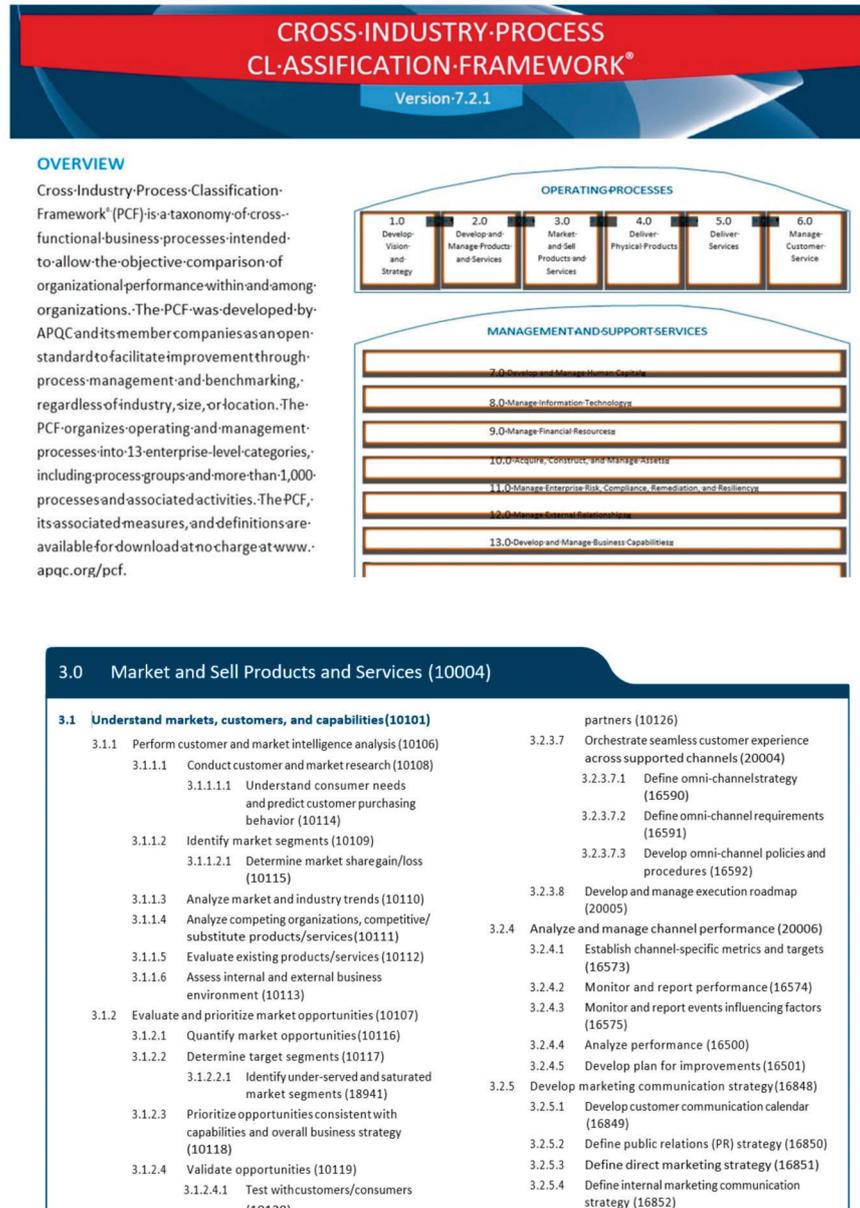


Figure 1. APQC cross-industry process classification framework.

A first version of the business processes for maritime container shipping companies has been generated by tailoring these cross-industry business processes, taking advantage of the following assets:

- The most relevant handbooks on maritime economics [16,17];
- Published research that includes parts of the business processes for a maritime container shipping company [9,18];
- The UN Convention on International Multimodal Transport [19].

The content validation of this model was performed by using an inter-judge validation process. The experts that participated in this validation were:

- Three Spanish Maritime Transportation Shipping Companies, including the participation of C-level executives and vice-presidents from these companies.
- A Spanish logistics management company.
- An expert in the sector from Universidad Politécnica de Madrid (UPM).

The content validation of this business process model was performed by using an inter-judge validation process. This method has been extensively used specially for validating survey questions. This work makes use of it, extending the concept of content validation beyond the one related with questions from a survey. The quantification of the agreement was calculated using the content validity ratio (CVR) developed by Lawshe [20]:

$$CVR = (ne - N/2)/(N/2)$$

where ne = number of judges indicating the question as “essential” (in this research, ne = number of judges indicating the modification of the BPMo as “essential”); and N = total number of judges (in this work, $N = 5$).

Lawshe considered the values of CVR included in Table 1 as the ones necessary for item validation.

Table 1. Minimum values of CVR.

Number of Judges	CVR Min. Value
5–7	0.99
8	0.85
9	0.78
10	0.62
11	0.59
12	0.56
13	0.54
14	0.51
15	0.49
20	0.42
25	0.337
30	0.33
35	0.31
40	0.29

Therefore, the method required the agreement amongst judges on the validity and clarity of the model.

The next step was building the list of digital application; three sources have been used for building such a list:

- Applications coming from academic research. These are the ones coming from the aforementioned *Sensors* paper from the team of this research.
- Applications that are already available in the market. This list has been built using the newsletter from the market. Some of them are: www.maritime-executive.com, www.vpoglobal.com, www.thedigitalship.com, www.dnv.com, www.shippingandfrieghtresource.com, and www.wartsila.com. It includes not only market-available applications but also others that are inspired by market-available ones. The range of dates for this analysis has been between March 2018 and September 2021.
- Applications coming from other industries. The search for these applications was completed via the internet between June 2021 and September 2021.

The list of applications was confronted to the aforementioned BPMo for maritime container shipping companies in order to qualify the impact on each process, the requirements for the implementation of the app, and the Key Performance Indicators (KPIs) that will measure the impact of the implementation.

The digitalization of the processes implies their redesign. The tool for quantifying the impact of this redesign is the devil’s quadrangle [21]. This framework evaluates the impact

using the four performance dimensions for processes: costs, time, quality, and flexibility. In this research, the impact has been quantified using the following criteria:

- Costs: this performance dimension is broken down into two sub-dimensions:
 - Implementation costs, which accounts the costs for deploying the application in the company. It has these values:
 - Low (equal to 2) for applications that require a low investment for their deployment.
 - Medium (equal to 1) for applications that require a medium investment for their deployment.
 - High (equal to 0) for applications that require a high investment for their deployment.

The aforementioned values are comparatively weighted (i.e., the values low, medium, and high are relative to the rest of the applications). The comparative analysis situated the applications in one of the three aforementioned tertiles (i.e., low, medium, and high).

- Execution cost, which evaluates the return of investment (ROI). It has these values:
 - Low (equal to 2) for applications with an ROI in less than 2 months.
 - Medium (equal to 1) for applications with an ROI in 2–12 months.
 - High (equal to 0) for applications that need more than 12 months for their ROI.

The final value of the performance indicator is obtained by arithmetic media of the two sub-dimensions.

- Time is also broken down into two sub-dimensions:
 - Implementation time, which accounts the time needed for deploying the application in the company. It has these values:
 - Low (equal to 2) for applications that can be deployed in less than 6 months.
 - (equal to 1) for applications that can be deployed in 6–18 months.
 - (equal to 0) for applications that need more than 18 months for their deployment.

- Execution time, which evaluates the savings in time for the processes' execution. It has these values:

- High (equal to 2) for applications with a high decrease on processes' execution time.
- Medium (equal to 1) for applications with a medium decrease on processes' execution time.
- Low (equal to 0) for applications with a small decrease on processes' execution time.

The aforementioned values are comparatively weighted (i.e., the values' categorization as low, medium, or high is relative to the rest of the applications). The final value of the performance indicator is obtained by arithmetic media of the two sub-dimensions.

- Quality, that evaluates the reliability added to the processes by the application. It has the following values:
 - High (equal to 2) for applications with a high increase on the processes' reliability.
 - Medium (equal to 1) for applications with a medium increase on the processes' reliability.
 - Low (equal to 0) for applications with a small increase on the processes' reliability.

The aforementioned values are comparatively weighted (i.e., the values' categorization as low, medium, or high is relative to the rest of the applications).

- Flexibility, the performance indicator that evaluates the flexibility that the application has on the company’s processes. It has the following values:
 - High (equal to 2) for applications with a high increase on the processes’ flexibility.
 - Medium (equal to 1) for applications with a medium increase on the processes’ flexibility.
 - Low (equal to 0) for applications a small increase on the processes’ flexibility.
 The aforementioned values are comparatively weighted (i.e., the values’ categorization as low, medium, is high is relative to the rest of the applications).

The “ideal” application is the one that maximizes the four performance indicators, and therefore, the impact on the company is considered positive. That application will achieve a total score of 8 (i.e., a score of 2 in each of the four performance dimensions for processes).

A data sheet was developed for each of these applications, which has the information from Figure 2.

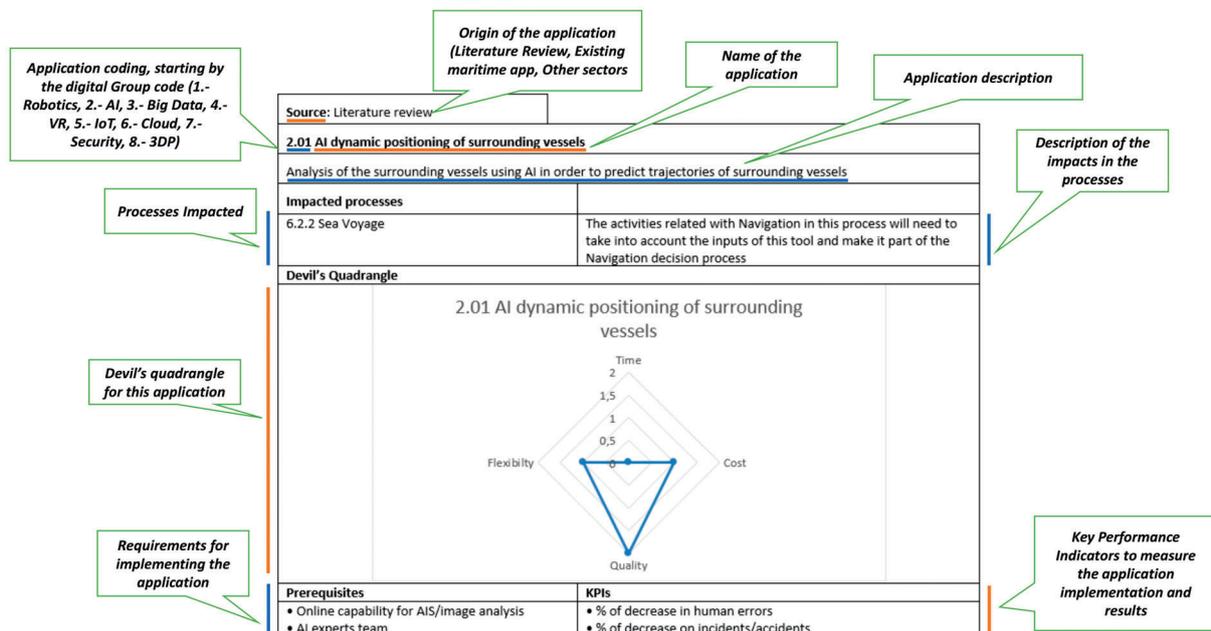


Figure 2. Application data sheet sample.

The impacts on the processes have been analyzed assuming that the application considered is the only one that has been deployed, i.e., that it has been deployed stand alone. The combined deployment of several applications will require a review on the impacts. This same consideration applies to the KPIs: in case of deployment of several applications, the KPIs should be reviewed and confirmed.

3. Results and Discussion

The validated BPMo for maritime container shipping companies can be found as additional material of this paper. The application of the methodology from Section 3 resulted in a total of 46 application data sheets that contain the results of the research. These results are the applications data sheets, they have the impacts in the BPMo, and they can be found in Appendix A of the present work.

Regarding the impacts on the processes, a total of 147 impacts have been found. The processes highly impacted by different applications are shown in Table 2.

Table 2. Processes with more than two impacts.

Process	Number of Impacts
6.2.2 Sea Voyage	15
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	9
6.1.5 Charge Ship	7
6.3.4 Unload Ship	7
6.2.1 Unberth Ship	6
6.2.6 Berth Ship	6
7.3.1 BIS Analysis of Operational Data for Maintenance	5
2.2.1 Analyze Competitors' Routes	4
6.1.3 Prepare Stowage Plan	4
6.3.2 Prepare Ship Unloading Plan	4
9.8.5 Define Safety Framework (Goals, KPIs, Training, Drills, etc.)	4
6.1.2 Manage Departure Customs and Rest of Departure Paperwork	3
6.3.1 Manage Arrival Customs and Rest of Arrival Paperwork	3
9.3.3 Manage Employees Training	3
9.4.3 Manage Training on Board	3

The processes with higher impacts are within the operations process categories domain. The reason is that these processes are the ones that produce the wealth of the company, so these are the ones subject to higher investments.

There are six new processes that need to be added to the business process model for different applications:

- 3.3.1 BIS analysis of containers' capabilities. This process is added to the process group "3.3 Acquire/Rent Containers" for application "5.01 Container tracking". The introduction of this application recommends a process to group the tasks and activities related with the different use of a container and the technological capabilities from it in an IoT environment.
- 4.2.2 BIS analysis of liner terms based on AI analysis of client information. This process is added to process group "4.2 Analyze and Define Liner Terms" for application "2.05 Client offering optimization via AI analysis of client information".
- 6.2.1 BIS start equipment monitorization. This process is added to process group "6.2 Depart, Sea Voyage and Berth" for application "5.02 Optimization of equipment usage".
- 6.2.6 BIS end equipment monitorization. This process is also added to process group "6.2 Depart, Sea Voyage and Berth" for the same reason, the application "5.02 Optimization of equipment usage".
- 7.3.1 BIS analysis of operational data for maintenance. This process is also added to process group "7.3 Ship Maintenance" for applications "2.06 Analysis of engine parameters to anticipate issues", "2.11 Optimizing maintenance process using digital twin and AI", "2.22 Using AI to reduce emissions", "3.02 Big data analysis for energy efficiency", and "3.03 Analysis of data on consumption and emissions for bunkering selection".
- 7.3.1 BIS capture and analysis of ship structure image. This process is added to process group "7.3 Ship Maintenance" for application "2.13 Analysis of ship structure images to anticipate issues".

On the opposite side, there are three processes that will need to be decommissioned due to the introduction of different applications:

- 6.2.4 Technical support at shore; the introduction of applications "1.01 UV-controlling system" or "1.02 Autonomous vessels" makes it unnecessary.

- 11.5.2 Define development plan; the introduction of application “6.02 Use of SaaS via cloud” makes it unnecessary.
- 11.5.3 Develop and test solution for the same reason mentioned when talking of decommissioning process 11.5.2.

The total number of KPIs used is 51. These KPIs measure the performance of the introduction of the 46 digital applications; this means that on average, there is more than one KPI needed for measuring the performance of the introduction of one application, which is reasonable. Actually, there are only two applications that require only one KPI for tracking their performance: “6.01 Cloud/edge platform” and “7.01 Enhanced cybersecurity” can be measured using KPIs “Percentage of reduction of operational cost” and “Percentage of improvement on cyberattacks prevented”, respectively. These two applications tracked only with one KPI are the only ones that are service platforms for the entire company.

It has just been said that the majority of the applications are measured using more than one KPI; actually, the 46 applications required a total of 105 KPIs to measure their performance. Since the number of unique KPIs is 51, not 105, this means that many of them are used in more than one application. A total of 11 KPIs are used more than twice; they are used 53 times out of the mentioned 105 (Table 3), so 11 KPIs can measure more than 50% of what needs to be measured to quantify applications’ performance, which is a good number since with this relatively small number of KPIs, a company can track most of their improvements coming from all the digital applications.

Table 3. KPIs used more than twice for measuring application performance.

KPI	Number of Applications that Use This KPI to Measure Their Performance
% improvement on ratio cost using old process/cost using new process	8
% of decrease in human errors	8
Number of days of improvement in the decision process	8
% decrease on annual maintenance hours	5
% of decrease on safety incidents	5
% of reduction on costs of fuel consumption	4
% decrease on mechanical failures	3
% of decrease on incidents/accidents	3
% of improvement on customer satisfaction	3
% of improvement on end-of-year financial results	3
% of reduction on training costs	3
TOTAL	53

Section 2 explained how the impact of the introduction of any of these applications in the BPMo has been quantified using the four performance dimensions for processes: Costs, Time, Quality, and Flexibility. Applications have been grouped into three tertiles in order to analyze the results of this quantified impact. The three tertiles are not always equal in size, since being strict on the balance between the three tertiles would have forced the separation in different tertiles of applications that have the same value on a performance dimension. This happens since there are performance dimensions such as costs or time that are made of two sub-dimensions (see Section 2); and the consolidated impact score is calculated from the four performance dimensions.

Table 4 contains the applications that are in the top tertile applying that evaluation.

Table 4. List of applications with higher consolidated impact score.

Application	Impact Score
5.01 Container tracking	8
2.06 Analysis of engine parameters to anticipate issues	7
7.02 Cargo documents management	7
2.04 Route optimization via AI analysis of client information	6.5
4.03 VR for maintenance	6.5
2.07 Route optimization via AI analysis of operational information	6
2.14 Optimizing ship's operations via AI analysis of operational information	6
2.11 Optimizing maintenance process using digital twin and AI	5.5
2.22 Using AI to reduce emissions	5.5
3.02 Big data analysis for energy efficiency	5.5
3.03 Analysis of data on consumption and emissions for bunkering selection	5.5
6.03 Use of eLearning via cloud	5.5
8.01 Spare parts using 3DP	5.5
1.03 Digital twin for AV controlling and maintenance	5
1.04 Use of robots in complex/hazardous tasks	5
2.13 Analysis of ship structure images to anticipate issues	5
2.21 Using AI to enhance navigation safety	5
3.06 Big data for ship speed controlling	5
6.02 Use of SaaS via cloud	5

Not surprisingly, application “5.01 Container tracking” is leading the score given the following:

- Regarding costs, it is not too expensive to implement, and the return of investment is high, since it enhances containers’ delivery process.
- Looking at time, it is also optimal in terms of implementation time (there are many market applications for a quick implementation), and it saves considerable time for tracking the containers.
- The quality of the process increases as the applications are error-prone compared with the manual process.
- The flexibility of the affected processes increases considerably compared to the manual tracking.

On the other side of the list in Table 5, the applications that are in the bottom tertile can be found.

Table 5. List of applications with lower consolidated impact score.

Application	Impact Score
2.01 AI dynamic positioning of surrounding vessels	3
3.04 ISPS security levels	3
5.02 Optimization of equipment usage	3
5.03 Digital twin for training purposes	3
1.01 UV controlling system	2.5
1.02 Autonomous vessels	2
3.05 Big data for ship renewal	2
7.01 Enhanced cybersecurity	2
2.15 AI applied to cybersecurity	1
2.17 AI applied to competitors tracking and monitoring	1
2.18 AI applied to business partners tracking and monitoring	1
2.19 AI applied to providers tracking and monitoring	1
2.20 AI applied to three parties route prediction	1
2.16 AI applied to data management and clean	0.5
6.01 Cloud/Edge platform	0.5

The applications from Table 5 do not necessarily fall into applications that should not be implemented or that should be discarded. What these 15 applications from Table 5 have in common is that they are the lowest when compared with the 46 applications; this should not prevent companies from the implementation of any of them, it is just that they need to know these have more costs or require more time for their deployment and for benefits realization. Actually, there are two of them that are service platforms for the rest: “6.01 Cloud/Edge platform” and “7.01 Enhanced cybersecurity”.

The analysis can be taken to a level below the ones conducted so far by looking at the results obtained in each performance dimension. When looking at the list of top applications on time performance dimension (Table 6), there is one application that is top when looking at time performance but is not only not included in the top list for consolidated impact but is in the bottom side, so it is included in Table 5: it is “3.04 ISPS security levels”. The reason is that this application does not increase substantially the flexibility or the quality of the affected processes compared with the rest of the 46 applications.

Table 6. Top applications on Time performance dimension.

Application	Impact Score
4.01 VR for training	2
4.03 VR for maintenance	2
5.01 Container tracking	2
6.02 Use of SaaS via cloud	2
6.03 Use of eLearning via cloud	2
7.02 Cargo documents management	2
7.04 Electronic logbook	2
1.04 Use of robots in complex/hazardous tasks	1.5
2.04 Route optimization via AI analysis of client information	1.5
2.06 Analysis of engine parameters to anticipate issues	1.5
2.07 Route optimization via AI analysis of operational information	1.5
2.10 Fleet dimensioning optimization	1.5
2.12 Conversational virtual assistance for helping seafarers in day-to-day activities	1.5
2.13 Analysis of ship structure images to anticipate issues	1.5
3.02 Big data analysis for energy efficiency	1.5
3.03 Analysis of data on consumption and emissions for bunkering selection	1.5
3.04 ISPS security levels	1.5
3.06 Big data for ship speed controlling	1.5
8.01 Spare parts using 3DP	1.5

One application is found in the opposite situation: being in the list of top performers when looking at the consolidated (Table 4); it is at the bottom side when looking at the time performance dimension (Table 7). This is the case of “2.14 Optimizing ship’s operations via AI analysis of operational information”. This application scores 6/8 in the consolidated impact score given it is outstanding when compared to others in the Flexibility and Quality provided to the affected processes, and its availability in the current market makes it almost optimal when looking at Costs.

Table 7. Bottom applications on Time performance dimension.

Application	Impact Score
2.02 Assessment of ship risks using fuzzy logic	0.5
2.08 Process optimization and reengineering using AI	0.5
2.09 Freight rate optimization	0.5
2.14 Optimizing ship’s operations via AI analysis of operational information	0.5
5.02 Optimization of equipment usage	0.5
1.01 UV controlling system	0
1.02 Autonomous vessels	0
2.01 AI dynamic positioning of surrounding vessels	0
2.15 AI applied to cybersecurity	0
2.16 AI applied to data management and clean	0
2.17 AI applied to competitors tracking and monitoring	0
2.18 AI applied to business partners tracking and monitoring	0
2.19 AI applied to providers tracking and monitoring	0
2.20 AI applied to 3 parties route prediction	0
6.01 Cloud/Edge platform	0

Moving to Costs, two applications are at the top for this performance dimension (Table 8) and at the bottom when looking at the consolidated score (Table 5). These are “3.04 ISPS security levels” and “5.02 Optimization of equipment usage”. The first one was found in the same situation when looking at Time, and the reason is the same: it does not increase substantially the flexibility or the quality of the affected processes compared with the rest of the 46 applications. Regarding “5.02 Optimization of equipment usage”, it does not increase flexibility or quality, and it is also low when looking at Time, since it is not available on the market yet.

As it happened when analyzing Time, one application is found in the bottom list from Costs (Table 9) and at the top when looking at the consolidated score (Table 4); this is “2.11 Optimizing maintenance process using digital twin and AI” given it is outstanding when compared to others in the Flexibility and Quality provided to the affected processes, but the cost of a digital twin makes it go down in the list when looking only at this performance dimension.

The next performance dimension to be looked into is Quality. In it, there are two applications that are in the top for this performance dimension (Table 10), whereas they are part of the list of bottom applications in the consolidated score (Table 5). These are “1.01 UV controlling system” and “1.02 Autonomous vessels”, which have a very high impact on quality improvement for the affected processes but perform very low in the rest of the

variables (high costs, high time of ROI and implementation, and without a substantial impact in flexibility compared to the others).

Table 8. Top applications on Cost performance dimension.

Application	Impact Score
2.04 Route optimization via AI analysis of client information	2
3.02 Big data analysis for energy efficiency	2
3.03 Analysis of data on consumption and emissions for bunkering selection	2
5.01 Container tracking	2
7.02 Cargo documents management	2
7.04 Electronic logbook	2
8.01 Spare parts using 3DP	2
1.04 Use of robots in complex/hazardous tasks	1.5
2.02 Assessment of ship risks using fuzzy logic	1.5
2.06 Analysis of engine parameters to anticipate issues	1.5
2.07 Route optimization via AI analysis of operational information	1.5
2.13 Analysis of ship structure images to anticipate issues	1.5
2.14 Optimizing ship's operations via AI analysis of operational information	1.5
2.22 Using AI to reduce emissions	1.5
3.01 Big data algorithm for collision avoidance	1.5
3.04 ISPS security levels	1.5
3.06 Big data for ship speed controlling	1.5
4.03 VR for maintenance	1.5
5.02 Optimization of equipment usage	1.5
6.03 Use of eLearning via cloud	1.5

Table 9. Bottom applications on Cost performance dimension.

Application	Impact Score
1.01 UV controlling system	0.5
2.03 Pricing market prediction	0.5
2.05 Client offering optimization via AI analysis of client information	0.5
2.11 Optimizing maintenance process using digital twin and AI	0.5
2.16 AI applied to data management and clean	0.5
6.01 Cloud/Edge platform	0.5
7.03 Blockchain-based Incoterms	0.5
1.02 Autonomous vessels	0
1.03 Digital twin for AV controlling and maintenance	0
2.08 Process optimization and reengineering using AI	0
2.10 Fleet dimensioning optimization	0
2.17 AI applied to competitors tracking and monitoring	0
2.18 AI applied to business partners tracking and monitoring	0
2.19 AI applied to providers tracking and monitoring	0
2.20 AI applied to 3 parties route prediction	0
3.05 Big data for ship renewal	0
5.03 Digital twin for training purposes	0

Table 10. Top applications on Quality performance dimension.

Application	Impact Score
1.01 UV controlling system	2
1.02 Autonomous vessels	2
1.03 Digital twin for AV controlling and maintenance	2
1.04 Use of robots in complex/hazardous tasks	2
2.06 Analysis of engine parameters to anticipate issues	2
2.11 Optimizing maintenance process using digital twin and AI	2
2.14 Optimizing ship's operations via AI analysis of operational information	2
2.21 Using AI to enhance navigation safety	2
2.22 Using AI to reduce emissions	2
3.02 Big data analysis for energy efficiency	2
3.03 Analysis of data on consumption and emissions for bunkering selection	2
4.03 VR for maintenance	2
5.01 Container tracking	2
7.02 Cargo documents management	2
7.03 Blockchain-based Incoterms	2

Comparing Table 11 (bottom applications for Quality performance dimension) and Table 4 (top consolidated score), applications “6.02 Use of SaaS via cloud”, “6.03 Use of eLearning via cloud” and “8.01 Spare parts using 3DP” are in both lists due to the same reason: they do not increase substantially the quality of the affected processes when compared to others, whereas they perform well on the rest of the performance dimensions.

Table 11. Bottom applications on Quality performance dimension.

Application	Impact Score
2.02 Assessment of ship risks using fuzzy logic	0
2.15 AI applied to cybersecurity	0
2.16 AI applied to data management and clean	0
2.17 AI applied to competitors tracking and monitoring	0
2.18 AI applied to business partners tracking and monitoring	0
2.19 AI applied to providers tracking and monitoring	0
2.20 AI applied to 3 parties route prediction	0
3.04 ISPS security levels	0
4.01 VR for training	0
6.01 Cloud/Edge platform	0
6.02 Use of SaaS via cloud	0
6.03 Use of eLearning via cloud	0
7.01 Enhanced cybersecurity	0
7.04 Electronic logbook	0
8.01 Spare parts using 3DP	0

Moving to the last performance dimension, Flexibility, comparing Table 12 (top performers in Flexibility) and Table 5 (bottom in consolidated score), there is no application in both lists.

Table 12. Top applications on Flexibility performance dimension.

Application	Impact Score
1.03 Digital twin for AV controlling and maintenance	2
2.02 Assessment of ship risks using fuzzy logic	2
2.03 Pricing market prediction	2
2.04 Route optimization via AI analysis of client information	2
2.05 Client offering optimization via AI analysis of client information	2
2.06 Analysis of engine parameters to anticipate issues	2
2.07 Route optimization via AI analysis of operational information	2
2.08 Process optimization and reengineering using AI	2
2.09 Freight rate optimization	2
2.10 Fleet dimensioning optimization	2
2.11 Optimizing maintenance process using digital twin and AI	2
2.14 Optimizing ship's operations via AI analysis of operational information	2
5.01 Container tracking	2
6.02 Use of SaaS via cloud	2
6.03 Use of eLearning via cloud	2
8.01 Spare parts using 3DP	2

However, doing the same exercise with bottom applications in Flexibility (Table 13) and top performers in consolidated score (Table 4), there are three applications in both lists: "1.04 Use of robots in complex/hazardous tasks", "3.02 Big data analysis for energy efficiency", and "3.03 Analysis of data on consumption and emissions for bunkering selection", all of them outperforming in the rest of the performance dimensions.

Table 13. Bottom applications on Flexibility performance dimension.

Application	Impact Score
1.01 UV controlling system	0
1.02 Autonomous vessels	0
1.04 Use of robots in complex/hazardous tasks	0
2.12 Conversational virtual assistance for helping seafarers in day-to-day activities	0
2.15 AI applied to cybersecurity	0
2.16 AI applied to data management and clean	0
3.01 Big data algorithm for collision avoidance	0
3.02 Big data analysis for energy efficiency	0
3.03 Analysis of data on consumption and emissions for bunkering selection	0
3.04 ISPS security levels	0
3.05 Big data for ship Renewal	0
5.02 Optimization of equipment usage	0
6.01 Cloud/Edge platform	0
7.01 Enhanced cybersecurity	0
7.04 Electronic logbook	0

To finalize the analysis of results, we identified the 10 applications that can be named as "quick wins". These are applications that, given their optimal results on the Time performance dimension and good results on the Costs performance dimension, could be considered as the starting point for digitalizing a company. A company starting its

digitalization with these could obtain a sense of what digitalization is and learn lessons of the implementation project, which will be value for going to the next step.

The list has been obtained by sorting the results of the score of the devil's quadrant first by those performing better on Time, then on Costs, and finally on consolidated global score. The list is in Table 14.

Table 14. Quick-win applications.

Application	Time	Costs	Global Impact
5.01 Container tracking	2	2	8
7.02 Cargo documents management	2	2	7
7.04 Electronic logbook	2	2	4
4.03 VR for maintenance	2	1.5	6.5
6.03 Use of eLearning via cloud	2	1.5	5.5
6.02 Use of SaaS via cloud	2	1	5
4.01 VR for training	2	1	4
2.04 Route optimization via AI analysis of client information	1.5	2	6.5
3.02 Big data analysis for energy efficiency	1.5	2	5.5
3.03 Analysis of data on consumption and emissions for bunkering selection	1.5	2	5.5

The majority of these are in Table 4 (List of applications with higher consolidated impact score); they are applications that are top performers in the consolidated impact score. The exceptions are "7.04 Electronic logbook" and "4.01 VR for training". These two do not score as high as others when looking at the consolidated score but can be good candidates for testing the benefits of digitalization in one company, given their ease of implementation.

Summing up the analysis of the results, the main outcomes are as follows:

- The processes with higher impacts are within the Operations process categories domain. The reason is that these processes are the ones that produce the wealth of the company, so these are the ones subject to higher investments. The one more frequently impacted is "6.2.2 Sea Voyage"; this will be impacted by 32.6% of the applications.
- There are six new processes that will be necessary when implementing some applications from AI or IoT domains. These processes are from the Strategy, Infrastructure & Products and from the Operations process categories domains.
- On the other side, there are three processes that will need to be decommissioned when implementing two applications (one from the Cloud digital domain and one from the Robotics one). They are within the Operations and the Enterprise Management process categories domains.
- The KPIs needed for measuring the performance of the digitalization of the BPMo are 51, though 11 of them can measure more than 50% of what is necessary for tracking the outcomes of the digitalization.
- The quantification of the impacts performed with the devil's quadrant gives a perspective on how the digitalization can benefit a company for implementing an application, but it does not necessary imply that applications in the bottom of the list should not be implemented; the decision of going for one application or another should be made by the company looking at its priorities and needs. There are some conclusions though coming from the results of this analysis:
 - Applications "5.01 Container tracking", "2.06 Analysis of engine parameters to anticipate issues", and "7.02 Cargo documents management" are at the top of the list of the consolidated impact score. These applications are market available which, together with the nature of the application, makes the Time and Costs performance dimensions better when compared to others. They are also in the top of the list in Quality.
 - Applications "2.16 AI applied to data management and clean" and "6.01 Cloud/Edge platform" are at the bottom of the list, though especially the last one is necessary for others to work (i.e., it is a prerequisite for implementing a number of other applications).

- There are applications that are at the top when looking at the consolidated score but at the bottom when looking at one performance dimension. This is the case for “2.14 Optimizing ship’s operations via AI analysis of operational information” (bottom in Time but top in Flexibility and Quality and almost optimal in Costs), and it is also the case for “2.11 Optimizing maintenance process using digital twin and AI” (bottom in Costs but top in Flexibility and Quality and average in Time). This happens also with “6.02 Use of SaaS via cloud”, “6.03 Use of eLearning via cloud”, and “8.01 Spare parts using 3DP” (low in Quality but much better in the rest of performance dimensions), and with “1.04 Use of robots in complex/hazardous tasks”, “3.02 Big data analysis for energy efficiency”, and “3.03 Analysis of data on consumption and emissions for bunkering selection” (same situation just described but with Flexibility rather than Quality).
- The opposite also happens: applications that are at the bottom when looking at consolidated score are at the top for one performance dimension. This is the case of “3.04 ISPS security levels” (top in Time and Costs but bottom in Flexibility and Quality), “5.02 Optimization of equipment usage” (top in Costs but bottom or almost at the bottom in the rest). This happens also with “1.01 UV controlling system” and “1.02 Autonomous vessels” (top in Quality and low in the rest).
- A list of 10 applications has been identified as “quick wins” applications that can be the starting point for digitalizing a company given their optimal results on the Time performance dimension and good results on the Costs performance dimension.

4. Conclusions

This work analyzes the impact of digitalization in a part of the maritime transport industry, the maritime containers shipping companies. This research has been conducted in order to help the digitalization of this industry, in particular in the aforementioned companies: digitalization in today’s world is required for remaining competitive.

The analysis of the introduction of digital applications in the Business Process Model of maritime containers shipping companies shows that digitalization is feasible for these companies and can be completed at different paces. Each company should make a specific and detailed plan for digitalization, according to their needs and environment. They can leverage the work presented here on the applications and the KPIs that should measure the implementation of any of these applications.

Companies can also benefit from the identification of the applications named in this work as “quick wins”; these applications can be a sandbox that can be used to test the benefits of digitalization and learn how to best execute the deployment customized to the needs of the company. Application “5.01 Container tracking” is in the top of the list of these “quick wins” given its optimal behavior when looking at the four performance dimensions for processes (Time, Costs, Quality, and Flexibility).

The impact of digitalization is high when trying to deploy all the applications at the same time in a big bang approach. Such an approach is not advisable not only given the high investment it requires but also due to the risks that such a huge effort poses for a company. Companies should consider the impacts in their processes and the applications’ prerequisites documented for each application in Section 3 of this work. They should also review their existing balanced scorecard incorporating the application’s KPIs documented in the aforementioned section. The KPIs defined are 51, but with 11 of them, a company can track the majority of the impacts of an application deployment.

A relevant outcome of the analysis of the results of the impacts in processes is that the Operational process categories domain is the one with higher impacts. This is a consequence of the applications trying to impact the processes that generate the company’s incomes. Looking at the rest of the process categories domains, there is one process that stands out

from the rest, “Analyze Competitors Routes”. This process from the Strategy, Infrastructure, and Products process categories domain is impacted by four different applications given the importance that the market and the research is given to a company’s strategy.

Digitalizing a company imposes changes in their processes and the definition of new processes as well as the decommissioning of others. In other words, digitalization will change the way a company operates. This is something that must be taken into account when defining the deployment plan of the applications, educating their personnel in the new way of doing things and the benefits that this will bring.

Digitalization has many impacts in the company’s operations but a plan well defined, in which the impacts and prerequisites are detailed and where a number of KPIs is included to track the deployment’s performance, is the key for success. This work covers these aspects in order to allow a successful digitalization.

Author Contributions: P.-L.S.-G. designed the methodology and applied it to obtain the results; D.D.-G. and L.R.N.-R. analyzed the results and provided feedback on the reporting; P.-L.S.-G. wrote the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research is under consideration for funding by Fundación Marqués de Suanzes and by Soermar Chair from Universidad Politécnica de Madrid.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A Application Data Sheets

Source: Literature review	
1.01 UV controlling system	
System for controlling the unmanned vessel. There are different types: straight line patch, non-linear line-of-sight, approach controller, etc.	
Impacted processes	

Figure A1. Cont.

2.3.4 Analyze Ports Windows and Infrastructure	The analysis should include the port facilities for unmanned vessels as well as the local regulations on them.
6.1.1 Receive and Process Charge Documentation	The documentation will be received by the ground crew.
6.1.2 Manage Departure Customs and Rest of Departure Paperwork	These activities will rely on ground staff.
6.1.3 Prepare Stowage Plan	The stowage plan will be prepared by the ground crew.
6.1.4 Complete Departure Preparation	Departure will be split into two types of staff: administrative staff that will take care of paperwork and seafarers that will take care of reviewing the readiness of the vessel for the navigation.
6.1.5 Charge Ship	These activities will be very similar to the current ones. There will be some adaptations to the fact that the ship is unmanned; for example, the captain and the bridge crew will be on land controlling the process instead of in the bridge
6.2.1 Unberth Ship	The process will change in all its content given that the control will be off the ship.
6.2.2 Sea Voyage	The process will change in all its content given that the control will be off the ship.
6.2.3 Logistics Coordination at Shore	Its content will remain the same but the players will no longer be on board.
6.2.4 Technical Support at Shore	Process will be decommissioned.
6.2.5 Prepare for Port of Arrival Activities	The interactions will be split among vessel, crew on land, and third parties.
6.2.6 Berth ship	The process will change in all its content given that the control will be off the ship.
6.3.1 Manage Arrival Customs and Rest of Arrival Paperwork	Minor changes coming from the fact that crew will be off board.
6.3.2 Prepare Ship Unloading Plan	These activities will be very similar to the current ones. There will be some adaptations to the fact that the ship is unmanned; for example, the captain and the bridge crew will be on land controlling the process instead of in the bridge.
6.3.3 Prepare for Shore Logistics	These activities will rely on ground staff.
6.3.4 Unload Ship	These activities will be very similar to the current ones. There will be some adaptations to the fact that the ship is unmanned; for example, the captain and the bridge crew will be on land controlling the process instead of in the bridge.
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	The maintenance will need to be completed remotely and will rely on sensors information as well as on automatic tools. There is a need of having resources for boarding the ship for some high-priority maintenance activities.
9.3.1 Manage Staffing Needs	There needs to be a new profile of ground employees that will have a crew profile given the high number of crew activities.
9.4.1 Manage Boarding Process	The crew will no longer board, but the activities will need to be completed with this sole exception.
9.4.4 Manage Disembark Process	The crew will no longer board, but the activities will need to be completed with this sole exception.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • IoT integration • Autonomous controls in the ships • Cloud/Edge network in the ships 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Crew cost decrease • Percentage of improvement in ratio space for cargo space for crew • Percentage of increase in ship utilization

Figure A1. UV controlling system data sheet.

Source: Existing maritime app	
1.02 Autonomous vessels	
Navigation directly controlled from earth, reducing cost and human errors	
Impacted processes	
6.1.1 Receive and Process Charge Documentation	The documentation will be received by the ground crew.
6.1.2 Manage Departure Customs and Rest of Departure Paperwork	These activities will rely on ground staff.
6.1.3 Prepare Stowage Plan	The stowage plan will be prepared by the ground crew.
6.1.4 Complete Departure Preparation	Departure will be split in two types of staff: administrative staff that will take care of paperwork, and seafarers that will take care of reviewing the readiness of the vessel for the navigation
6.1.5 Charge Ship	These activities will be completed very similar to the current one. There will be some adaptations to the fact that the ship is unmanned; for example, the captain and the bridge crew will be on land controlling the process instead of in the bridge.
6.2.1 Unberth Ship	The process will change in all its content given that the control will be off the ship.
6.2.2 Sea Voyage	The process will change in all its content given that the control will be off the ship.
6.2.3 Logistics Coordination at Shore	Its content will remain the same but the players will no longer be on board.
6.2.4 Technical Support at Shore	The process will be decommissioned.
6.2.5 Prepare for Port of Arrival Activities	The interactions will be split among the vessel, crew on land, and third parties.
6.2.6 Berth Ship	The process will change in all its content given that the control will be off the ship.
6.3.1 Manage Arrival Customs and Rest of Arrival Paperwork	Minor changes coming from the fact that the crew will be off board.
6.3.2 Prepare Ship Unloading Plan	These activities will be completed very similar to the current one. There will be some adaptations to the fact that the ship is unmanned; for example, the captain and the bridge crew will be on land controlling the process instead of in the bridge.
6.3.3 Prepare for Shore Logistics	These activities will rely on ground staff.
6.3.4 Unload Ship	These activities will be completed very similar to the current one. There will be some adaptations to the fact that the ship is unmanned; for example, the captain and the bridge crew will be on land controlling the process instead of in the bridge
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • IoT integration • Autonomous controls in the ships • Cloud/Edge network in the ships 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Crew cost decrease • Percentage of improvement in ratio space for cargo space for crew • Percentage of increase in ship utilization

Figure A2. Autonomous vessels data sheet.

Source: Existing maritime app	
1.03 Digital twin for AV controlling and maintenance	
Leverage in the digital twin to better control UV and to help in maintenance	
Impacted processes	
6.1.3 Prepare Stowage Plan	The process needs to integrate the digital twin in its activities (e.g., simulations of stowage in the digital twin).
6.1.5 Charge Ship	The process needs to integrate the digital twin in its activities, mainly on troubleshooting actions.

Figure A3. Cont.

6.2.1 Unberth Ship	The process needs to integrate the digital twin in its activities, mainly on troubleshooting actions.
6.2.2 Sea Voyage	The process needs to integrate the digital twin in its activities, mainly on troubleshooting actions.
6.2.6 Berth Ship	The process needs to integrate the digital twin in its activities, mainly on troubleshooting actions.
6.3.2 Prepare Ship Unloading Plan	The process needs to integrate the digital twin in its activities (e.g., simulations of unloading in the digital twin).
6.3.4 Unload Ship	The process needs to integrate the digital twin in its activities, mainly on troubleshooting actions.
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	The maintenance activities will now leverage on the digital twin for the preparation and troubleshooting.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Digital twin • IoT integration • Autonomous controls in the ships • Could/Edge network in the ships 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Percentage of improvement on maintenance costs

Figure A3. Digital twin for AV controlling and maintenance data sheet.

Source: Existing maritime app	
1.04 Use of robots in complex/hazardous tasks	
Make use of robots to avoid the exposure of the crew to risks (checks of cargo, underwater surveys, antifouling, etc.)	
Impacted processes	
6.1.3 Prepare Stowage Plan	The preparation will need to take into account the readiness of these robots.
6.1.5 Charge Ship	The activities will need to change to integrate the robots into them.
6.3.2 Prepare Ship Unloading Plan	The preparation will need to take into account the readiness of these robots.
6.3.4 Unload Ship	The activities will need to change to integrate the robots into them.
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	The plan will need to make sure robots are available for the activities. The execution will need to change some of the activities that will now be executed by robots.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Robots • Could/Edge network in the ships 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Percentage of increase in efficiency on these tasks

Figure A4. Use of robots in complex/hazardous tasks data sheet.

Source: Literature review	
2.01 AI dynamic positioning of surrounding vessels	
Analysis of the surrounding vessels using AI in order to predict the trajectories of the surrounding vessels	
Impacted processes	
6.2.2 Sea Voyage	The activities related with navigation in this process will need to take into account the inputs of this tool and make it part of the navigation decision process.
Devil's quadrangle	
Prerequisites	
<ul style="list-style-type: none"> • Online capability for AIS/image analysis • AI experts team 	KPIs
	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Percentage of decrease on incidents/accidents

Figure A5. AI dynamic positioning of surrounding vessels data sheet.

Source: Literature review	
Other industry app	
2.02 Assessment of ship risks using fuzzy logic	
Analysis of ships previous to the acquisition for determining risks associated with it. Applies both to new and existing ships (prediction of future markets, among others)	
Impacted processes	
3.2.1 Analyze Freights Evolution and Ships Demand	The activities of this process will be updated with a new one: freight and ship demand analysis using AI techniques.
3.2.2 Analyze Ships Pricing (chartering, new vs. secondhand)	The activities of this process will be updated with a new one: ships pricing analysis using AI techniques.
Devil's quadrangle	
Prerequisites	
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	KPIs
	<ul style="list-style-type: none"> • Number of days of improvement in the decision process • Percentage of improvement on ratio cost using old process/cost using new process

Figure A6. Assessment of ship risks using fuzzy logic data sheet.

Source: Literature review	
Other industry app	
2.03 Pricing market prediction	
Using AI techniques, identify market trends on pricing	
Impacted processes	
4.2.3 Obtain Optimal Liner Terms	The liner terms will be now obtained looking to AI analysis results as well as to the legacy activities.

Figure A7. Cont.

Devil's quadrangle	
<p>2.03 Pricing market prediction</p> <p>The radar chart shows scores for four metrics: Time (top), Cost (right), Quality (bottom), and Flexibility (left). The scale ranges from 0 to 2. The scores are approximately: Time: 1.1, Cost: 0.8, Quality: 0.7, Flexibility: 1.0.</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Number of days of improvement in the decision process • Percentage of improvement on ratio cost using old process/cost using new process

Figure A7. Pricing market prediction data sheet.

Source: Literature review Other industry app	
2.04 Route optimization via AI analysis of client information	
Analysis of routes and containers used by a client to reinforce or decrease routes	
Impacted processes	
2.5.2 Analyze Route Operational Results	The activities of this process will be updated with a new one: route operational analysis using AI techniques.
Devil's quadrangle	
<p>2.04 Route optimization via AI analysis of client information</p> <p>The radar chart shows scores for four metrics: Time (top), Cost (right), Quality (bottom), and Flexibility (left). The scale ranges from 0 to 2. The scores are approximately: Time: 1.5, Cost: 1.0, Quality: 0.8, Flexibility: 1.0.</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of decrease on unattended demand on time due to demand peaks • Percentage of decrease on overcapacity due to demand valleys • Number of optimized routes • Percentage of improvement on ratio cost using old process/cost using new process

Figure A8. Route optimization via AI analysis of client information data sheet.

Source: Existing maritime app	
2.05 Client offering optimization via AI analysis of client information	
Analysis of routes and containers used by a client for clients offering customization	
Impacted processes	
4.2.2 BIS Analysis of Liner Terms Based on AI Analysis of Client Information	New process that needs to be added to the model in order to include the activities derived from this application.
Devil's quadrangle	

Figure A9. Cont.

<p>2.05 Client offering optimization via AI analysis of client information</p> 	
<p>Prerequisites</p> <ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<p>KPIs</p> <ul style="list-style-type: none"> • Percentage of decrease on unattended demand on time due to demand peaks • Percentage of decrease on overcapacity due to demand valleys • Percentage of improvement on customer satisfaction • Percentage of improvement on ratio cost using old process/cost using new process

Figure A9. Client offering optimization via AI analysis of client information data sheet.

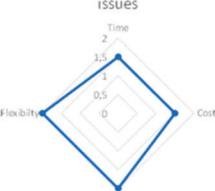
<p>Source: Existing maritime app</p>	
<p>2.06 Analysis of engine parameters to anticipate issues</p>	
<p>The parameters from the engine (consumption, performance, rpm, etc.) will be analyzed using AI techniques to anticipate potential issues</p>	
<p>Impacted processes</p>	
<p>7.3.1 BIS Analysis of Operational Data for Maintenance</p>	<p>New process that needs to be added to the model in order to include the activities derived from this application.</p>
<p>Devil's quadrangle</p>	
<p>2.06 Analysis of engine parameters to anticipate issues</p> 	
<p>Prerequisites</p> <ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<p>KPIs</p> <ul style="list-style-type: none"> • Percentage of decrease on annual maintenance hours • Percentage decrease on mechanical failures

Figure A10. Analysis of engine parameters to anticipate issues data sheet.

<p>Source: Existing maritime app</p>	
<p>2.07 Route optimization via AI analysis of operational information</p>	
<p>Analysis of routes using AI on operational data (weather, sea conditions, traffic, online port conditions (costs, bunkering, delays, etc.), changes of crews, etc.) to optimize routes</p>	
<p>Impacted processes</p>	
<p>2.5.2 Analyze Route Operational Results</p>	<p>The activities of this process will be updated with a new one: route operational analysis using AI techniques.</p>
<p>Devil's quadrangle</p>	
<p>2.07 Route optimization via AI analysis of operational information</p> 	

Figure A11. Cont.

Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Crew cost decrease • Percentage of improvement in ratio space for cargo-space for crew • Percentage of increase in ship utilization

Figure A11. Route optimization via AI analysis of operational information data sheet.

Source: Existing maritime app	
2.08 Process optimization and reengineering using AI	
Analysis based on AI of the outcomes from every process in the organization to optimize and redesign them	
Impacted processes	
12.2.3 Manage Business Processes Performance	The activities of this process will be updated with a new one: business process analysis using AI techniques.
12.2.4 Improve Business Processes Model	The activities of this process will be updated with a new one: business process improvement options analysis using AI techniques.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Data Warehouse for process performance metrics (PPDW) • Analysis of PPWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of reduction on route operational costs • Percentage of improvement on ratio efficiency using old process/ efficiency using new process • Percentage of improvement on customer satisfaction

Figure A12. Process optimization and reengineering using AI data sheet.

Source: Existing maritime app	
2.09 Freight rate optimization	
Analysis of internal and external data using AI techniques for determine the optimal freight rate	
Impacted processes	
4.2.3 Obtain Optimal Liner Terms	The liner terms will be now obtained looking to AI analysis results as well as to the legacy activities.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of improvement in ratio space for cargo-space for crew • Percentage of improvement on end-of-year financial results

Figure A13. Freight rate optimization data sheet.

Source: Existing maritime app	
2.10 Fleet dimensioning optimization	
Analysis of internal and external data using AI techniques for determine the optimal use of the existing fleet as well as improving fleet dimensioning process	
Impacted processes	
3.1.1 Obtain Fleet Operational Data	This process will change since the operational data will be treated using AI techniques in order to propose options regarding the fleet.
3.1.2 Obtain Route Operational Data	This process will change since the operational data will be treated using AI techniques in order to propose options regarding the routes.
3.1.3 Obtain Customers Demand Data	This process will change since the operational data will be treated using AI techniques in order to forecast customer demand.
3.1.4 Design Fleet Deployment Plan	The deployment plan options will go through the AI tool to help in the decision of the best option.
3.1.5 Maintain Fleet Deployment Plan	Results will be compared to the AI tools one in order to tune both (fleet deployment plan and the AI tool).
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of improvement in ratio space for cargo space for crew • Percentage of improvement on end-of-year financial results

Figure A14. Fleet dimensioning optimization data sheet.

Source: Existing maritime app	
2.11 Optimizing maintenance process using digital twin and AI	
Using data form the digital twin, an AI framework can help in the preventive maintenance process (for example, using the historical data to predict when it will fail) as well as in the optimization of the maintenance process (for example, making sure spares are available when needed, minimizing maintenance time, etc.)	
Impacted processes	
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	The maintenance activities will now leverage not only on the digital twin for the preparation and troubleshooting but also on the inputs gathered from sensors that will be analyzed using AI techniques.
7.3.1 BIS Analysis of Operational Data for Maintenance	This is a new process that needs to be added to the model in order to include the activities derived from this application.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage decrease on annual maintenance hours • Percentage decrease on mechanical failures

Figure A15. Optimizing maintenance process using digital twin and AI data sheet.

Source: Existing maritime app	
2.12 Conversational virtual assistance for helping seafarers in day-to-day activities	
Development of a virtual assistance (Alexa type) that combines the use of AI techniques to help seafarers in daily activities (reporting of weather conditions, recommendation of routes, gathering data of engine, etc.)	
Impacted processes	
6.1.5 Charge Ship	The process needs to integrate the AI virtual assistance in its activities not only on troubleshooting actions but also for providing inputs and executing orders on demand.
6.2.1 Unberth Ship	The process needs to integrate the AI virtual assistance in its activities not only on troubleshooting actions but also for providing inputs and executing orders on demand.
6.2.2 Sea Voyage	The process needs to integrate the AI virtual assistance in its activities not only on troubleshooting actions but also for providing inputs and executing orders on demand.
6.2.6 Berth Ship	The process needs to integrate the AI virtual assistance in its activities not only on troubleshooting actions but also for providing inputs and executing orders on demand.
6.3.4 Unload Ship	The process needs to integrate the AI virtual assistance in its activities not only on troubleshooting actions but also for providing inputs and executing orders on demand.
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	The main impact will be in the execution of the maintenance, since this tool will improve the technician diagnosis and maintenance.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Development of an IoT network on board • Analysis of IoT data using AI techniques • AI experts team • Integration with a conversational virtual assistance • Development of an AI human interaction framework 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Percentage of reduction on costs of the activities (including crew costs and SW/HW costs)

Figure A16. Conversational virtual assistance for helping seafarers in day-to-day activities data sheet.

Source: Existing maritime app	
2.13 Analysis of ship structure images to anticipate issues	
The images from ships (hull, hatches, cranes, etc.) will be analyzed using AI techniques to anticipate potential issues (corrosion, coating issues, welding problems, etc.)	
Impacted processes	
7.3.1 BIS Capture and Analysis of Ship Structure Image	This is a new process that needs to be added to the model in order to include the activities derived from this application.
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	The maintenance activities will now leverage on the images for the preparation and troubleshooting that will be analyzed using AI techniques.
Devil's quadrangle	

Figure A17. Cont.

<p>2.13 Analysis of ship structure images to anticipate issues</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Data exchange between ship and land-based data center • Analysis of IoT data using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage decrease on annual maintenance hours • Percentage decrease on mechanical failures

Figure A17. Analysis of ship structure images to anticipate issues data sheet.

Source: Existing maritime app	
2.14 Optimizing ship's operations via AI analysis of operational information	
Analysis of vessel operational data (engine parameters, load distribution, consumptions, sea traffic, etc.) to maximize its performance	
Impacted processes	
6.0.1 Capture and analysis of ship's operational data	This is a new process that needs to be added to the model in order to include the activities derived from this application.
6.1.5 Charge Ship	The activities of the process will be adapted to process the inputs coming from level 3 process 6.0.1.
6.2.1 Unberth Ship	The activities of the process will be adapted to process the inputs coming from level 3 process 6.0.1.
6.2.2 Sea Voyage	The activities of the process will be adapted to process the inputs coming from level 3 process 6.0.1.
6.2.6 Berth Ship	The activities of the process will be adapted to process the inputs coming from level 3 process 6.0.1.
6.3.4 Unload Ship	The activities of the process will be adapted to process the inputs coming from level 3 process 6.0.1.
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	The main impact will be in the execution of the maintenance, since this tool will improve the technician's diagnosis and maintenance.
Devil's quadrangle	
<p>2.14 Optimizing ship's operations via AI analysis of operational information</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Data exchange between ship and land-based data center • Analysis of IoT data using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Percentage of reduction on costs of fuel consumption

Figure A18. Optimizing ship's operations via AI analysis of operational information data sheet.

Source: Existing maritime app	
2.15 AI applied to cybersecurity	
Use of neuronal networks to detect and prevent cyberattacks	
Impacted processes	

Figure A19. Cont.

11.3.3 Control IT Security and Risks	The activities will be highly impacted, since the integration of the AI technology will require a full review of cybersecurity procedures.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> Development and training of a neuronal network using internal and external data on cyberattacks 	<ul style="list-style-type: none"> Percentage of improvement on cyberattacks prevented

Figure A19. AI applied to cybersecurity data sheet.

Source: Existing maritime app	
2.16 AI applied to data management and clean	
Use of AI to enhance dashboards, enhance data quality, detect patterns on information at the DWH, populate DWH with relevant information	
Impacted processes	
2.1.4 Generate Global Economy Situation Model	The activities of this process will need to apply AI to the data obtained here for the final goal of the process group.
2.3.7 Generate Ports Analysis Report	The activities of this process will need to apply AI to the data obtained here for the final goal of the process group.
3.1.4 Design Fleet Deployment Plan	The activities of this process will need to apply AI to the data obtained here for the final goal of the process group.
5.3.2 Populate and Manage Customer's Information DWH	The population of the DWH will strongly leverage on the AI application for enhancing the data quality.
9.6.2 Manage Employees Data	The activities of this process will need to apply AI to the data obtained here for the final goal of the process group.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> Create a Big Data Warehouse (BDWH) Analysis of BDWH using AI techniques AI experts team 	<ul style="list-style-type: none"> Number of days of improvement in the decision process Percentage of reduction of data inaccuracy Percentage of reduction of failure on data population Percentage of maintenance resources costs reductions

Figure A20. AI applied to data management and clean data sheet.

Source: Existing maritime app	
2.17 AI applied to competitors tracking and monitoring	
Use of AI to monitor actions and performance of competitors	
Impacted processes	
2.2.1 Analyze Competitors Routes	The activities of this process will be adapted to accommodate the use of AI in such a way that information will flow constantly instead of the current batches of information.
2.2.2 Analyze Competitors Finance	The activities of this process will be adapted to accommodate the use of AI in such a way that information will flow constantly instead of the current batches of information.
2.2.3 Analyze Competitors Sales	The activities of this process will be adapted to accommodate the use of AI in such a way that information will flow constantly instead of the current batches of information.
2.2.4 Analyze Competitors Clients	The activities of this process will be adapted to accommodate the use of AI in such a way that information will flow constantly instead of the current batches of information.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Number of days of improvement in the decision process • Percentage reduction in customers' churn • Percentage of increase on new customers/business

Figure A21. AI applied to competitors tracking and monitoring data sheet.

Source: Existing maritime app	
2.18 AI applied to business partners tracking and monitoring	
Use of AI to monitor actions and performance of business partners	
Impacted processes	
4.3.2 Design Alliance Programs and Plans for Managing Relationships	The activities of this process will be adapted to accommodate the use of AI in such a way that information will flow constantly instead of the current batches of information.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Number of days of improvement in the decision process • Percentage of increase on partners' efficiency

Figure A22. AI applied to business partners tracking and monitoring data sheet.

Source: Existing maritime app	
2.19 AI applied to providers tracking and monitoring	
Analysis performance by provider using AI techniques to better define the relationship	
Impacted processes	
10.5.2 Manage Procurement	The activities of this process will be adapted to accommodate the use of AI in such a way that information will flow constantly instead of the current batches of information.
Devil's quadrangle	
Prerequisites	
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	KPIs
	<ul style="list-style-type: none"> • Number of days of improvement in the decision process • Percentage of increase on providers' efficiency

Figure A23. AI applied to providers tracking and monitoring data sheet.

Source: Existing maritime app	
2.20 AI applied to 3 parties route prediction	
Use of AI to monitor and predict route of other vessels around	
Impacted processes	
6.2.2 Sea Voyage	The activities related with navigation in this process will need to take into account the inputs of this tool and make it part of the navigation decision process, acting as a backup for AIS when not available.
Devil's quadrangle	
Prerequisites	
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	KPIs
	<ul style="list-style-type: none"> • Percentage of decrease on incidents/accidents • Percentage of reduction of cost coming incidents/accidents

Figure A24. AI applied to 3 parties route prediction data sheet.

Source: Existing maritime app	
2.21 Using AI to enhance navigation safety	
Analysis on vessel data to anticipate dangerous scenarios based on AI	
Impacted processes	
6.2.2 Sea Voyage	The activities related with navigation in this process will need to take into account the inputs of this tool and make it part of the navigation decision process.
Devil's quadrangle	

Figure A25. Cont.

<p>2.21 Using AI to enhance navigation safety</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of decrease on safety incidents • Percentage of reduction of cost coming safety incidents

Figure A25. Using AI to enhance navigation safety data sheet.

Source: Existing maritime app	
2.22 Using AI to reduce emissions	
Analysis on vessel data to reduce GHG emissions. Capturing data from vessel, weather, and route to reduce ship's emissions	
Impacted processes	
6.2.2 Sea Voyage	The activities related with navigation in this process will need to take into account the inputs of this tool and make it part of the navigation decision process.
7.3.1 BIS Analysis of Operational Data for Maintenance	This is a new process that, in this case, will provide inputs to the tool for helping in the decision-making process.
Devil's quadrangle	
<p>2.22 Using AI to reduce emissions</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Analysis of BDWH using AI techniques • AI experts team 	<ul style="list-style-type: none"> • Percentage of decrease on annual maintenance hours • Percentage of reduction of cost coming from penalties

Figure A26. Using AI to reduce emissions data sheet.

Source: Literature review	
3.01 Big data algorithm for collision avoidance	
Big data analysis to calculate the safe distance of approach of a ship under the head-on situation, the crossing situation, and the overtaking situation, calculating the risk-degree of collision of ships and determining the degree of immediate danger of ships for avoidance of shipwreck	
Impacted processes	
6.2.2 Sea Voyage	The activities related with navigation in this process will need to take into account the inputs of this tool and make it part of the navigation decision process.
Devil's quadrangle	

Figure A27. Cont.

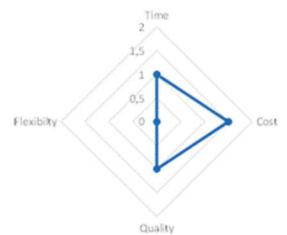
<p>3.01 Big data algorithm for collision avoidance</p> 	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Online capability for AIS/image analysis • Big Data experts 	<ul style="list-style-type: none"> • Percentage of decrease in human errors • Percentage of decrease on incidents/accidents • Percentage of reduction of cost related to incidents/accidents

Figure A27. Big data algorithm for collision avoidance data sheet.

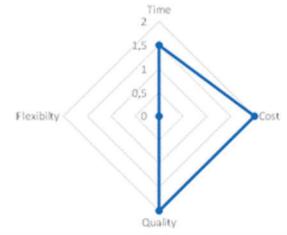
Source: Literature review	
3.02 Big data analysis for energy efficiency	
The energy efficiency plan is adjusted via big data analysis of the different ship parameters; the analysis supports the decision-making process	
Impacted processes	
6.2.2 Sea Voyage	The activities related with navigation in this process will need to take into account the inputs of this tool and make it part of the navigation decision process.
7.3.1 BIS Analysis of Operational Data for Maintenance	This is a new process that, in this case, will provide inputs to the tool for helping in maximizing efficiency.
Devil's quadrangle	
<p>3.02 Big data analysis for energy efficiency</p> 	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Online capability for AIS/image analysis • Big Data experts 	<ul style="list-style-type: none"> • Percentage of reduction on costs of fuel consumption • Percentage of CII improvement

Figure A28. Big data analysis for energy-efficiency data sheet.

Source: Existing maritime app	
3.03 Analysis of data on consumption and emissions for bunkering selection	
Data on consumption, NOx, SOx, engine failures, etc. will be stored and analyzed in order to evaluate the effect of bunkering providers in ship performance	
Impacted processes	
7.3.1 BIS Analysis of Operational Data for Maintenance	This is a new process that, in this case, it will provide inputs to the tool for helping in reducing costs and improve ship efficiency, both relative to enhance bunkering
Devil's quadrangle	

Figure A29. Cont.

<p>3.03 Analysis of data on consumption and emissions for bunkering selection</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Create a Big Data Warehouse (BDWH) • Big Data experts 	<ul style="list-style-type: none"> • Percentage of reduction on maintenance costs • Percentage of reduction on costs of fuel consumption

Figure A29. Analysis of data on consumption and emissions for bunkering selection data sheet.

Source: Existing maritime app	
3.04 ISPS security levels	
Gather data from different sources (news updates, radio communications, homeland security alerts, etc.) to better-quality risks related to ISPS security levels	
Impacted processes	
9.8.5 Define Safety Framework (Goals, KPIs, Training, Drills, etc.)	The inputs from previous tasks will enter into the DWH for a big data analysis in order to anticipate security issues.
9.8.6 Deploy and Monitor Safety Framework	The inputs from previous tasks will enter into the DWH for a big data analysis in order to anticipate security issues.
Devil's quadrangle	
<p>3.04 ISPS security levels</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Integration with 3rd parties DB for data collection • Big Data experts 	<ul style="list-style-type: none"> • Percentage of decrease on security incidents • Percentage of reduction of cost coming security incidents

Figure A30. ISPS security levels data sheet.

Source: Existing maritime app	
3.05 Big Data for Ship Renewal	
Gather data from different sources (market insights, supply/demand, changes on routes, etc.) to decide the best type of vessel for the market that is coming	
Impacted processes	
3.2.1 Analyze Freight Evolution and Ship Demand	The activities of this process will be adapted to accommodate the use of big data techniques in such a way that information will flow constantly instead of the current batches of information.
3.2.2 Analyze Ship Pricing (chartering, new vs. secondhand)	The activities of this process will be adapted to accommodate the use of big data techniques in such a way that information will flow constantly instead of the current batches of information.
3.2.3 Analyze Ship Characteristics (ship size, stability in different demand conditions, etc.)	The activities of this process will be adapted to accommodate the use of big data techniques in such a way that information will flow constantly instead of the current batches of information.
3.2.4 Analyze Fleet Productivity	The activities of this process will be adapted to accommodate the use of big data techniques in such a way that information will flow constantly instead of the current batches of information.
Devil's quadrangle	

Figure A31. Cont.

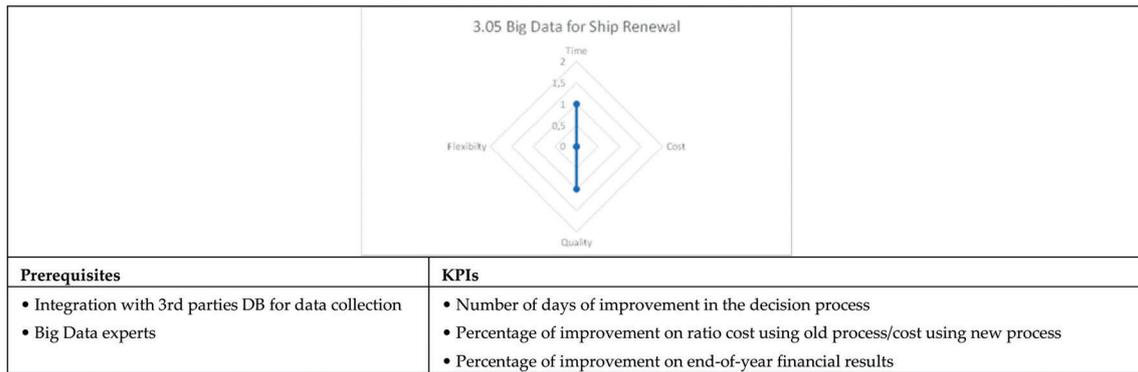


Figure A31. Big data for ship renewal data sheet.

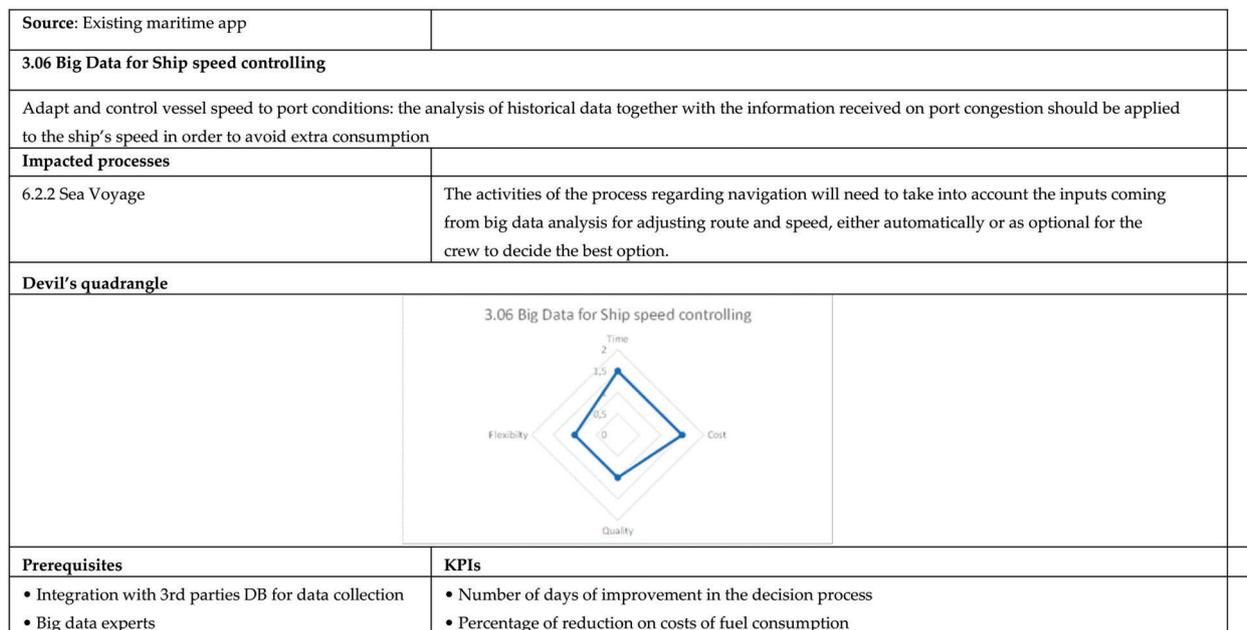


Figure A32. Big data for ship speed controlling data sheet.

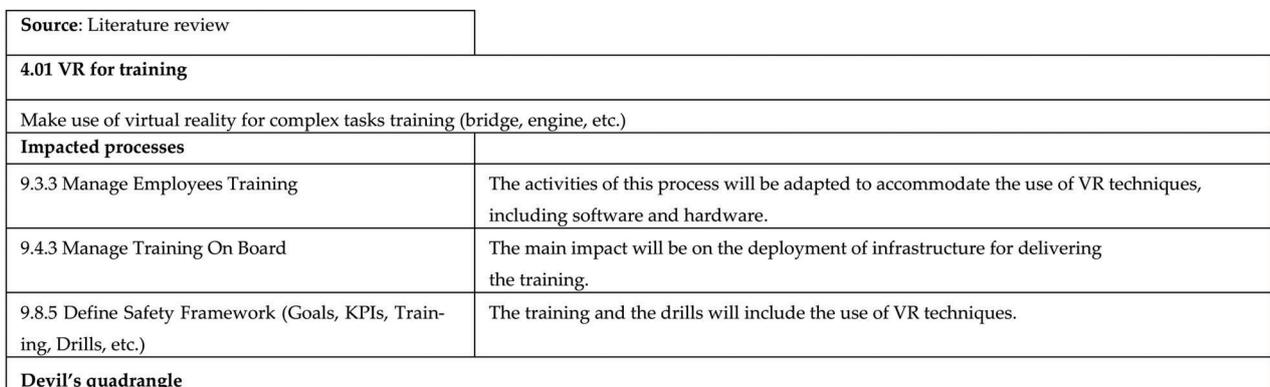


Figure A33. Cont.

Prerequisites	KPIs
<ul style="list-style-type: none"> • VR platform 	<ul style="list-style-type: none"> • Percentage of decrease on safety incidents • Percentage of reduction on training costs

Figure A33. VR for training data sheet.

Source: Literature review Existing maritime app	
4.02 VR as navigation aid	
Deploy VR solution for helping captain on decision-making process during navigation (access to port, navigation in poor weather conditions, etc.)	
Impacted processes	
6.2.1 Unberth Ship	The activities of this process will be adapted to accommodate the use of VR techniques so that the decisions from the crew will be based on the inputs from this tool.
6.2.2 Sea Voyage	The activities of this process will be adapted to accommodate the use of VR techniques so that the decisions from the crew will be based on the inputs from this tool.
6.2.6 Berth Ship	The activities of this process will be adapted to accommodate the use of VR techniques so that the decisions from the crew will be based on the inputs from this tool.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • VR platform 	<ul style="list-style-type: none"> • Percentage of decrease on safety incidents • Percentage of reduction of costs related to safety incidents

Figure A34. VR as navigation aid data sheet.

Source: Existing maritime app	
4.03 VR for maintenance	
Integrate VR in on-board maintenance processes, including views of the manuals and inputs from landed personnel such as vendors	
Impacted processes	
7.3.1 Plan and Execute Ship Daily Maintenance and Periodical Crew Exercises	Manuals and schemas will now be loaded in the VR tool, allowing quick access to them.
Devil's quadrangle	

Figure A35. Cont.

Prerequisites	KPIs
<ul style="list-style-type: none"> • VR platform 	<ul style="list-style-type: none"> • Percentage of decrease on safety incidents • Percentage of reduction of maintenance cost

Figure A35. VR for maintenance data sheet.

Source: Literature review Existing maritime app	
5.01 Container tracking	
Implement IoT into containers for tracking purposes	
Impacted processes	
3.3.1 BIS Analyze Containers Capabilities	New process that will include the qualification of containers for these services.
5.1.1 Contact Customer	The contact will move from a user contact to a service contact.
6.1.5 Charge Ship	All the container charging/discharging information will be through the IoT platform.
6.2.2 Sea Voyage	The IoT platform will be integrated into this process so that it can provide to the customers a reliable information of time of arrival.
6.3.4 Unload Ship	All the container charging/discharging information will be through the IoT platform.
8.2.1 Sign Multimodal Transportation Merchant Reception Documents	All the container reception/delivery information in a multimodal transport will be through the IoT platform.
8.2.2 Execute Maritime Leg	The IoT platform will be integrated into this process so that it can provide to the customers a reliable information of time of arrival.
8.2.3 Sign Multimodal Transportation Merchant Delivery Documents	All the container reception/delivery information in a multimodal transport will be through the IoT platform.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • IoT integration into containers • Cloud/Edge network in the ships 	<ul style="list-style-type: none"> • Percentage of improvement on customer satisfaction • Percentage of increase in containers' utilization

Figure A36. Container tracking data sheet.

Source: Existing maritime app	
5.02 Optimization of equipment usage	
Through IoT monitorization, optimize the use of equipment in order to avoid unnecessary energy consumption	
Impacted processes	
6.2.1 BIS Start Equipment Monitorization	This is a new process that will trigger the monitorization of the ship equipment and take the actions for optimizing their usage.
6.2.6 BIS End Equipment Monitorization	This is a new process that will end the monitorization of the ship equipment.

Figure A37. Cont.

Devil's quadrangle	
<p>5.02 Optimization of equipment usage</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • IoT integration into equipment • Edge network in the ships 	<ul style="list-style-type: none"> • Percentage of reduction of energy cost of equipment • Percentage of increase on equipment' lifetime

Figure A37. Optimization of equipment data sheet.

Source: Existing maritime app	
5.03 Digital twin for training purposes	
Leverage in the digital twin to develop advanced training to selected crew members	
Impacted processes	
9.3.3 Manage Employees Training	The activities of this process will be adapted to accommodate the use of the digital twin.
9.4.3 Manage Training On Board	The activities of this process will be adapted to accommodate the use of the digital twin.
9.8.5 Define Safety Framework (Goals, KPIs, Training, Drills, etc.)	The activities of this process will be adapted to accommodate the use of the digital twin.
Devil's quadrangle	
<p>5.03 Digital twin for training purposes</p>	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Digital twin • Development of training sessions based on twin 	<ul style="list-style-type: none"> • Percentage of decrease on safety incidents • Percentage of reduction on training costs

Figure A38. Digital twin for training purposes data sheet.

Source: Literature review Existing maritime app	
6.01 Cloud/Edge platform	
Use of a cloud/edge environment to capture and send/receive information used by the rest of the applications and digital domains	
Impacted processes	
11.6.3 Plan and Manage IT Infrastructure	The management of the infrastructure will need to include new activities in order to interconnect the company's IT management system with the one from the cloud provider.
Devil's quadrangle	

Figure A39. Cont.

Prerequisites	KPIs
<ul style="list-style-type: none"> • Development of the cloud/edge platform • Communications infrastructure • Establish cloud policies (security, configuration management, etc.) • Automated data replication infrastructure • Automated data workflow 	<ul style="list-style-type: none"> • Percentage of reduction of operational cost

Figure A39. Cloud/Edge platform data sheet.

Source: Existing maritime app	
6.02 Use of SaaS via cloud	
Reduce the TCO in SW by the use of SaaS platforms	
Impacted processes	
11.5.2 Define Development Plan	Decommissioned. This process will not be required for the SaaS since it will be completed by the vendor.
11.5.3 Develop and Test Solution	Decommissioned. This process will not be required for the SaaS since it will be completed by the vendor.
Devil's quadrangle	
Prerequisites	KPIs
	<ul style="list-style-type: none"> • Percentage of reduction of CAPEX –OPEX ratio • Percentage of reduction of cost coming of SW maintenance

Figure A40. Use of SaaS via cloud data sheet.

Source: Existing maritime app	
6.03 Use of eLearning via cloud	
Develop training sessions that can be downloaded by the workforce and move the online learning sessions to the cloud	
Impacted processes	
9.3.3 Manage Employees Training	The activities of this process will now include the ones for loading updated training in the cloud.
9.4.3 Manage Training On Board	The activities of this process will now include the ones for loading updated training in the cloud.
9.8.5 Define Safety Framework (Goals, KPIs, Training, Drills, etc.)	The KPIs will be updated to include performance management KPIs on these courses.
Devil's quadrangle	

Figure A41. Cont.

Prerequisites	KPIs
<ul style="list-style-type: none"> • Development of the eLearning platform or negotiate with an eLearning provider • HW for the workforce 	<ul style="list-style-type: none"> • Percentage of reduction in training costs • Percentage of increase in training on time

Figure A41. Use of eLearning via cloud data sheet.

Source: Existing maritime app	
7.01 Enhanced cybersecurity	
Deploy a cybersecurity layer so that it can provide support to the model. Add to periodical drills the ones on cybersecurity	
Impacted processes	
11.3.1 Define IT Security and Risks Strategy	The process in itself will not change much, but the resources allocated will now be of a higher expertise and with the tools to monitor and enable constant improvement.
11.3.2 Define IT Resilience Strategy	The process in itself will not change much, but the resources allocated will now be of a higher expertise and with the tools to monitor and enable constant improvement.
11.3.3 Control IT Security and Risks	The process in itself will not change much, but the resources allocated will now be of a higher expertise and with the tools to monitor and enable constant improvement.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> • Identification of critical systems • Identification of vulnerabilities (The IMO has identified below systems on board ships as particularly vulnerable: (1) Bridge systems; (2) Cargo handling and management systems; (3) Propulsion and machinery management and power control systems; (4) Access control systems; (5) Passenger servicing and management systems; (6) Passenger facing public networks; (7) Administrative and crew welfare systems; and (8) Communication systems) • Cybersecurity experts 	<ul style="list-style-type: none"> • Percentage of improvement on cyberattacks prevented

Figure A42. Enhanced cybersecurity data sheet.

Source: Existing maritime app	
7.02 Cargo documents management	
Management of cargo documentation (certificates, contracts, transmission, etc.) using blockchain	
Impacted processes	
6.1.2 Manage Departure Customs and Rest of Departure Paperwork	The activities of this process will change, replacing manual activities with blockchain ones.

Figure A43. Cont.

6.3.1 Manage Arrival Customs and Rest of Arrival Paperwork	The activities of this process will change, replacing manual activities with blockchain ones.
8.2.1 Sign Multimodal Transportation Merchant Reception Documents	The activities of this process will change, replacing manual activities with blockchain ones.
8.2.3 Sign Multimodal Transportation Merchant Delivery Documents	The activities of this process will change, replacing manual activities with blockchain ones.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> Blockchain network 	<ul style="list-style-type: none"> Percentage of decrease on number of cargo documentation issues Percentage of decrease on operational cost coming from cargo management

Figure A43. Cargo documents management data sheet.

Source: Existing maritime app	
7.03 Blockchain-based Incoterms	
Management of Incoterms implementation using blockchain	
Impacted processes	
5.2.1 Define Contractual Agreement Type	The activities of this process will change, replacing manual activities with blockchain ones.
Devil's quadrangle	
Prerequisites	KPIs
<ul style="list-style-type: none"> Blockchain network 	<ul style="list-style-type: none"> Percentage of decrease on number of Incoterms documentation issues Percentage of decrease on operational cost coming from Incoterms management

Figure A44. Blockchain-based Incoterms data sheet.

Source: Existing maritime app	
7.04 Electronic logbook	
Securitized logbooks on board using blockchain (maritime, machine, fuel, ballast water, etc.)	
Impacted processes	
6.2.2 Sea Voyage	The process needs to integrate the logbook in its activities, mainly on collecting and reporting information from the voyage.
13.2.2 Obtain Ship Initial Certificates	The activities within the process should include the certification using this electronic logbook.
13.2.4 Maintain Certificates	The activities within the process should include the certification using this electronic logbook.
Devil's quadrangle	

Figure A45. Cont.

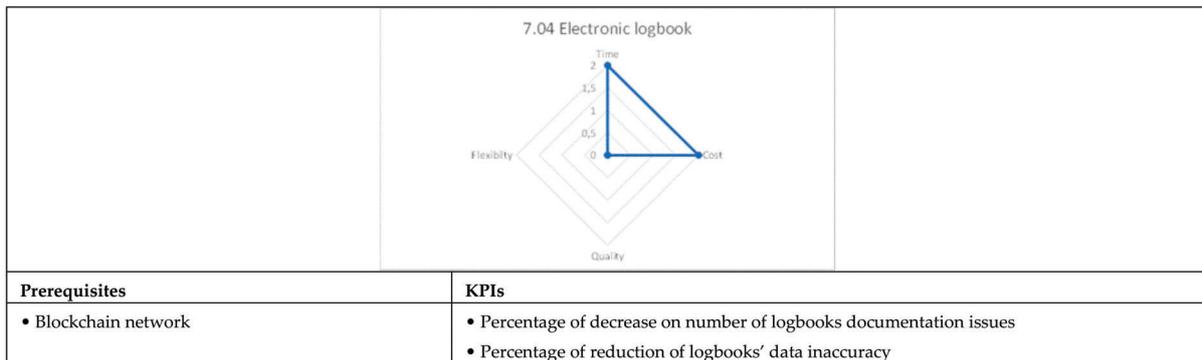


Figure A45. Electronic logbook data sheet.

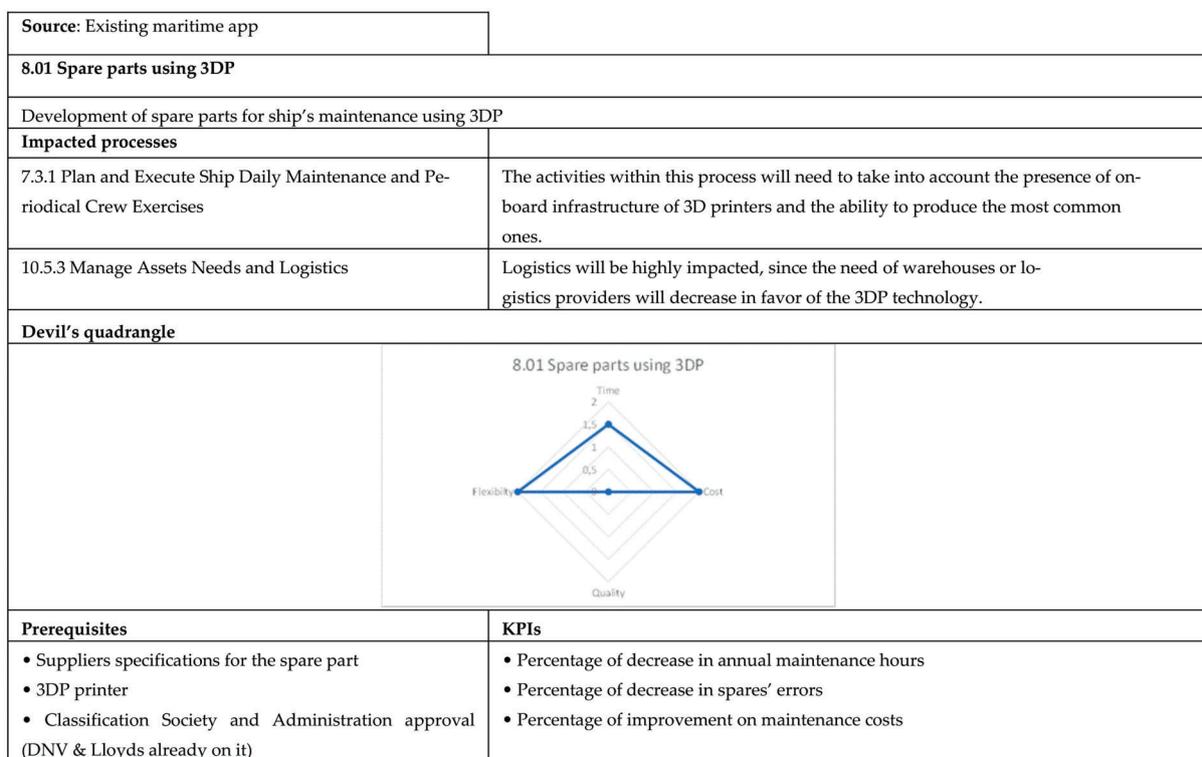


Figure A46. Spare parts using 3DP data sheet.

References

1. United Nations Conference on Trade and Development. Review of Maritime Transport 2021. *UNCTAD/RMT/2021*. Available online: https://unctad.org/system/files/official-document/rmt2021_en_0.pdf (accessed on 2 February 2022).
2. Gartner Group. Gartner Predicts the Future of Supply Chain Technology. Available online: <https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-supply-chain-technology> (accessed on 2 February 2022).
3. Munim, Z.H.; Dushenko, M.; Jaramillo Jimenez, V.; Shakil, M.H.; Imset, M. Big data and artificial intelligence in the maritime industry: A bibliometric review and future research directions. *Marit. Policy Manag.* **2020**, *47*, 577–597. [CrossRef]
4. Plaza-Hernández, M.; Gil-González, A.B.; Rodríguez-González, S.; Prieto-Tejedor, J.; Corchado-Rodríguez, J.M. Integration of IoT Technologies in the Maritime Industry. In Proceedings of the International Symposium on Distributed Computing and Artificial Intelligence, L'Aquila, Italy, 16–19 June 2020; Volume 1242. [CrossRef]
5. Kapidani, N.; Bauk, S.; Davidson, I.E. Digitalization in Developing Maritime Business Environments towards Ensuring Sustainability. *Sustainability* **2020**, *12*, 9235. [CrossRef]
6. Kapnissis, G.; Leligou, E.-E.; Vaggelas, G. Blockchain Challenges in Maritime Industry: An Empirical Investigation of the Willingness and the Main Drivers of Adoption from the Hellenic Shipping Industry. *Open J. Appl. Sci.* **2020**, *10*, 779–790. [CrossRef]

7. Tijan, E.; Jovic, M.; Aksentijevic, S.; Pucihar, A. Digital transformation in the maritime transport sector. *Technol. Forecast. Soc. Change* **2021**, *170*, 120879. [[CrossRef](#)]
8. Dumas, M.; La Rosa, M.; Mandling, J.; Reijers, H.A. Introduction to Business Process Management. In *Fundamentals of Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 1–31.
9. Lyridis, D.V.; Fyrvik, T.; Kapetanios, G.N.; Ventikos, N.; Anaxagorou, P.; Uthaug, E.; Psaraftis, H.N. Optimizing shipping company operations using business process modelling. *Marit. Policy Manag.* **2006**, *32:4*, 403–420. [[CrossRef](#)]
10. Elbert, R.; Pontow, H.; Benlain, A. The role of inter-organizational information systems in maritime transport chains. *Electron. Mark.* **2017**, *27*, 157–173. [[CrossRef](#)]
11. Cimino, M.G.C.A.; Palumbo, F.; Vaglini, G.; Ferro, E.; Celandroni, N.; La Rosa, D. Evaluating the impact of smart technologies in harbor's logistics via BPMN modeling and simulation. *Inf. Technol. Manag.* **2017**, *18*, 223–239. [[CrossRef](#)]
12. Nikitakos, N.; Lambrou, M.A. Chapter 12 Digital Shipping: The Greek Experience. *Res. Transp. Econ.* **2007**, *21*, 383–417. [[CrossRef](#)]
13. Trkman, P. The critical success factors of business process management. *Int. J. Inf. Manag.* **2010**, *30*, 125–134. [[CrossRef](#)]
14. American Productivity & Quality Centre. APQC Process Classification Framework (PCF)—Cross Industry—PDF Version 7.2.1. 2018. Available online: <https://www.apqc.org/resource-library/resource-listing/apqc-process-classification-framework-pcf-cross-industry-pdf-8> (accessed on 30 January 2020).
15. American Productivity & Quality Centre. APQC Process Classification Framework (PCF)—Cross Industry—Excel Version 7.2.1. 2019. Available online: <https://www.apqc.org/resource-library/resource-listing/apqc-process-classification-framework-pcf-cross-industry-excel-7> (accessed on 30 January 2020).
16. Stopford, M. *Maritime Economics*; Routledge: Oxon, UK, 2009.
17. Song, D.W.; Panayides, P.M. *Maritime Logistics*; Kogan Page Limited: London, UK, 2015.
18. Tran, N.K.; Haasis, H.-D.; Buer, T. Container shipping route design incorporating the costs of shipping, inland/feeder transport, inventory and CO2 emission. *Marit. Econ. Logist.* **2016**, *19*, 667–694. [[CrossRef](#)]
19. United Nations Conference on Trade and Development. United Nations Conference on a Convention on International Multimodal Transport. *UNCTAD 1981*. Available online: https://unctad.org/en/PublicationsLibrary/tdmtconf17_en.pdf (accessed on 30 January 2020).
20. Lawshe, C.H. A quantitative approach to content validation. *Pers. Psychol.* **1975**, *28*, 563–575. [[CrossRef](#)]
21. Dumas, M.; La Rosa, M.; Mandling, J.; Reijers, H.A. Process Redesign. In *Fundamentals of Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 253–296.

Article

A Helping Human Hand: Relevant Scenarios for the Remote Operation of Highly Automated Vehicles in Public Transport

Carmen Kettwich *, Andreas Schrank, Hüseyin Avsar and Michael Oehl

German Aerospace Center (DLR), Institute of Transportation Systems, Lilienthalplatz 7,
38108 Braunschweig, Germany; andreas.schrank@dlr.de (A.S.); huseyin.avsar@dlr.de (H.A.);
michael.oehl@dlr.de (M.O.)

* Correspondence: carmen.kettwich@dlr.de

Abstract: Remote operation bears the potential to roll out highly automated vehicles (AVs, SAE Level 4) more safely and quickly. Moreover, legal regulations on highly automated driving, e.g., the current law on highly automated driving (SAE Level 4) in Germany, permit a remote supervisor to monitor and intervene in driving operations remotely in lieu of a safety operator on board AVs. In order to derive requirements for safe and effective remote driving and remote assistance of AVs and to create suitable human-centered design solutions for human-machine interfaces (HMIs) that serve this purpose, a set of 74 core scenarios that are likely to occur in public transport AVs under remote operation was compiled. The scenarios were collected in several projects on the remote operation of AVs across a variety of contexts including interviews with and observations of control center staff, video analyses from naturalistic road events, and interviews with safety operators of AVs. A hierarchical system that is based on interactions of central actors was used to structure the scenarios. The set explicates relevant cases in remote operation, which may help improve workplaces for remote operation both by combatting human factors issues such as distraction and fatigue, and by boosting usability, user experience, trust, and acceptance. As the catalogue of scenarios is not exhaustive, scenarios may be added as knowledge of the remote operation of AVs progresses. Further research is needed to validate and adapt the scenarios to specific conceptualizations of remote operations.

Keywords: human-machine interaction; scenarios; use cases; remote operation; highly automated vehicles; user-centered design; remote assistance; remote driving

Citation: Kettwich, C.; Schrank, A.; Avsar, H.; Oehl, M. A Helping Human Hand: Relevant Scenarios for the Remote Operation of Highly Automated Vehicles in Public Transport. *Appl. Sci.* **2022**, *12*, 4350. <https://doi.org/10.3390/app12094350>

Academic Editors:

Giovanni Randazzo,

Anselme Muzirafuti, Dimitrios

S. Paraforos and Stefania Lanza

Received: 18 March 2022

Accepted: 11 April 2022

Published: 25 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As mobility demands grow while calls for more sustainable and environment-friendly travel options are becoming more vocal, the transportation sector is facing dramatic changes. Not only has there been a wave of technological innovations in driving automation, electromobility, and mobile network bandwidth [1–5]. There is also a previously unseen boom in innovative means of public transport such as ride-sharing and other flexible on-demand mobility solutions that may be serious alternatives to individual mass mobility [6,7]

However, even though automation technologies have demonstrated sharp improvements, there are still plenty of cases where a vehicle's automation might be overwhelmed. Complex, multilayered sets of quickly evolving traffic situations particularly in urban mixed traffic environments present extreme challenges to highly automated vehicles (AVs). Some of these cannot be resolved by the automation alone as they exceed the vehicle's Operational Design Domain, or ODD, i.e., the defined conditions in which the AV can operate autonomously in a safe way [8]. Also, a human on-board operator is not mandatory to serve as a fallback solution for highly automated vehicles (Level of Automation Four according to SAE's terminology [8]). Instead, remote operation by a human operator may be a viable solution to enable automated driving without specifying every possible ODD, which is the requirement of fully automated driving according to SAE Level 5. This

endeavor is likely not be fulfilled for a very long time, if ever. From the perspective of the substitution model, remote operation can be conceptualized as a substitute for the primary controller, in this case, the driving automation [9]. The remote operator could observe automated driving operations and intervene when the AV's driving automation capabilities are exceeded. This approach is becoming increasingly feasible as computers' processing speed and capacity have shown a sharp incline and high-bandwidth low-latency communication technologies with the option to prioritize certain kinds of data such as 5G have been widely rolled out [10].

In order to identify critical situations where remote operation could be used as an effective and efficient approach to support or resume highly automated driving, an extensive collection of scenarios with relevance to remote operation has been collected using a multi-method approach (see Section 2). This collection will help notice and address challenges in the practical use of remote operation solutions and support the human-centered design process of interfaces for remotely operating vehicles.

1.1. Regulation, Standardization, and Conceptualization of Remote Operation of AVs

In addition to technological leaps forward, legal environments have become more favorable toward using remote operations on public roads as well. For instance, the German Road Traffic Act ("Straßenverkehrsgesetz") has been modified last year so it now explicitly permits AV of Level 4 on German roads—as long as they are monitored and controlled, if necessary, by a human operator coined "Technical Supervisor" ("Technische Aufsicht") [11]. This supervisor can be either on board the vehicle or at another location. Thus, remote vehicle operations are now legally feasible on German roads. Also, in the UK, Sweden, Japan and a few US states, laws and regulations that require remote supervision of highly automated cars without an on-board driver have been passed [12]. Moreover, driverless operation of vehicles on public roads is now possible in major European countries like France and the UK [13] as well as at least 41 states of the US [14].

The standardization of remote operations has also seen tremendous steps forward. The latest update of the SAE's Taxonomy for Driving Automation Systems [8] includes two conceptualizations of remote vehicle operations: Remote Driving and Remote Assistance. In *Remote Driving* (RD), a human operator is executing "real-time performance of [. . .] the DDT (i.e., dynamic driving task such as braking, steering, or accelerating)" ([8] p. 19). Thus, remote driving resembles the conventional way of driving a vehicle: by initiating direct low-level driving maneuvers, including lateral and longitudinal motion control, right when the situation requires them. The Remote Driver, who is the actor to execute Remote Driving, may overrule the vehicle automation's driving tasks.

In contrast, *Remote Assistance* (RA) is defined as an "event-driven provision, by a remotely located human, of information or advice to an ADS (i.e., automated driving system) equipped vehicle in driverless operation in order to facilitate trip continuation when the ADS encounters a situation it cannot manage" ([8] p. 18). Thus, the Remote Assistant supports the vehicle by providing high-level guidance on how to deal with certain situations that are not part of the automation's ODD. This advice is provided well before the challenging situation and must not be time-critical. In addition, the automation must be capable of processing the high-level information provided and translate it into direct driving maneuvers. Examples for guidance range from simple "giving clearance" cases in which the vehicle requests an assessment of the situation from the Remote Assistant on how to proceed in a certain situation, e.g., when it is uncertain whether an identified object is an actual obstacle, to more demanding "setting trajectories or waypoints" cases that require the Remote Assistant to determine a pathway or waypoints of a pathway that the AV follows to circumvent an obstacle.

1.2. Real-World Tests of Remote Operation

The remote operation of AVs is currently being tested in a variety of real-world laboratories. These labs usually are collaborations between research-focused and industrial

partners which examine the feasibility of novel technologies in a real-world setting, aiming at maximizing ecological validity. Thanks to their incorporation into naturalistic settings within the intended context of use, real-world labs offer invaluable insights into the in situ application of devices or systems that have previously only been investigated in higher controlled oftentimes experimental, yet less realistic environments. Thus, situations and phenomena are more likely to occur that may not have been observed in a more controlled setting. Since they demonstrate the interplay of the technology with users and other actors in a less standardized environment, they yield scenarios with a larger external validity, i.e., transferring them to other (real-world) contexts may be facilitated.

The German Aerospace Center (DLR) is involved in real-world labs that use both RD and RA for remote operation. Figure 1 displays examples of vehicles that are remotely operated within these projects. Regarding RD, the modular electric AV concept “U-Shift” that caters to different urban mobility demands, including transportation of people and goods, encompasses Remote Drivers in its system architecture [15]. Pertaining to RA, DLR is engaged in urban mobility projects that provide last-mile shuttle services from major hubs of transportation, e.g., train stations, to the final destination. In the “Hamburg Electric Autonomous Transportation”, or “HEAT”, project, a self-driving minibus ran along a fixed route through the Harbor District of the city incorporated in the public transport provider “Hochbahn”’s network [16]. “The Real-World Lab Hamburg (“Reallabor Hamburg”)” provided on-demand service from a suburban railway hub to nearby neighborhoods that could be booked via a cellphone application [17]. The Berlin-based “KIS’M” project aims at demonstrating an AI-based system for connected mobility and at examining the interaction between human operators in the control center with passengers of remote-controlled AVs [18], utilizing an RA approach.



Figure 1. Remotely operated vehicles in investigated projects. (a) Modular vehicle concept “U-Shift”. Reprinted with permission from Ref. [19]. 2022, DLR; (b) Shuttle “EasyMile” used in real-world transportation laboratory “Reallabor Hamburg”.

1.3. Rationale and Objectives

Since the remote operation of AVs has not been widely rolled out so far, there is limited knowledge about concrete use cases and scenarios that are most relevant to it. However, being aware of events that may occur during remote operation is pivotal for several reasons: (1) It enables an ecologically valid determination of ODD thresholds for a vehicle’s automation, (2) helps bridge those thresholds, and (3) feeds into the derivation of requirements for the task and workplace design regarding the remote operation (both RD and RA) by a Technical Supervisor.

Therefore, a method of approximation via adjacent roles and workplaces needs to be taken. This includes the study of today’s already existing control centers for public transport as they execute tasks of monitoring and resolving disturbances that are comparable to those of remote operators. Further, gaining insights into workplaces of operators on board of already in real-world laboratories operating highly automated shuttles may be helpful as well. In this vein, control center staff has been interviewed about their expectations

on remote-operators' tasks [20] and has been confronted with a first prototype for remote operation [21]; a study on on-board operators' tasks and human-machine interfaces (HMIs) is currently carried out [22].

In spite of the lack of research opportunities regarding actual remote operators, there is an urgent need for an initial compilation of use cases and scenarios in remotely operating AVs. This is an important element of the user-centered design process as user requirements are derived from them. As presented in Figure 2, initially in this process the authors of this paper followed, and observations and expert interviews in a context that is similar to the future context of use are conducted. From their results, both potential tasks and potential scenarios are derived. This paper focuses on deriving potential scenarios. Next, both tasks that the users will have to execute and scenarios they will be exposed to are used to compile user requirements. These, in turn, need to be addressed while designing the prototype of the remote-control workplace. Whether they were met or not is subsequently evaluated in user studies.

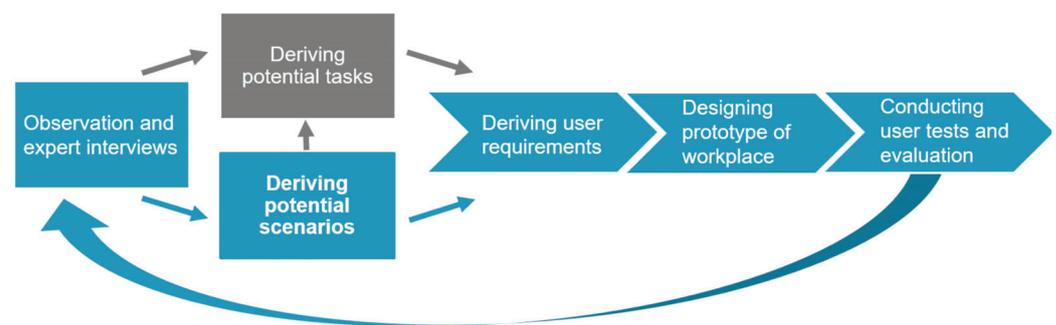


Figure 2. Empirical application of the user-centered design process by the authors. Observations and expert interviews serve as a basis both for deriving potential tasks and, which is the focus of this paper, on deriving potential scenarios. Both help derive user requirements that in turn inform the design of a workplace prototype, which will be validated by conducting user tests and evaluations. These results feed back to investigations of the context of use that has been initially investigated.

This procedure adapts the user-centered design process depicted in Figure 3 as specified by ISO (Section 7) [23]. First, the context of use needs to be understood and specified (box “Understanding and specifying the context of use”). There are different approaches to do so: One way is describing the context of use, of which tasks of the users are an essential characteristic. This approach is being pursued by the authors across several real-world laboratories for future mobility in Germany where they investigate tasks and human-machine interfaces of highly automated shuttle buses that are supervised by a human operator on board the vehicle [22]. Since on-board operators execute tasks similar to remote operators, this approach serves as an approximation to the tasks of remote operators that barely exist in urban road traffic to this date.

Another way is the specification of “as-is scenarios”. This is the core objective of this paper. It contains an extensive list of so-called “Is Scenarios”, as defined below. Thus, compiling scenarios and defining tasks are *parallel* steps that are both based on the empirical data, from sources such as interviews and observations. In a subsequent step, user requirements can be derived from both scenarios and tasks (“Specifying user requirements”). Eventually, these will be used to generate design solutions (“Producing design solutions”). This approach of basing the design on requirements that were factually articulated by potential users helps designing interactions in a user-friendly way that may increase safety, efficacy, ease of use, and prevent task overload, fatigue, and a lack of situational awareness that may increase the risk for accidents. The interactions, in turn, will facilitate the specification of HMIs and, eventually, help design workplaces for remote operation both in research and industrial application.

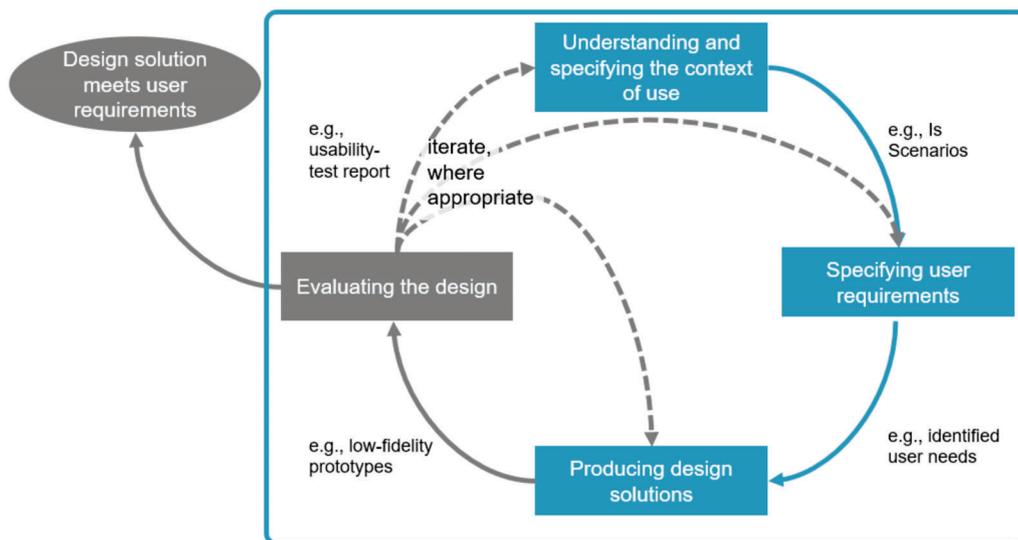


Figure 3. The user-centered design process is adapted from ISO [23]. Scenarios are a central element of understanding and specifying the context of use and determine user requirements that in turn help produce design solutions. Adapted from Ref. [22], 2019, ISO.

In addition to considerations of HMI design, compiling use cases and scenarios is also essential for creating a framework that can be used as a basis for interdisciplinary dialogue between engineers, computer scientists, mobility researchers, human factors specialists, and decision-makers on how to conceptualize and further develop remote operation. Furthermore, it may also be used in driving automation and transportation research (see Section 4).

This paper proposes an initial catalogue of use cases and scenarios in which remote operation, operationalized either as Remote Driving or Remote Assistance at SAE Level 4, supports vehicle automation. It is highlighted that this catalogue is a living conceptual document. As it contains statements from a limited number of sources, it does not claim to be exhaustive.

2. Materials and Methods

The following section will outline the process of user-centered design in which scenarios for remote control are a central element. Second, the process of collecting scenarios will be described before a system for systematically structuring the scenarios will be proposed.

2.1. Process of Collecting Scenarios

The scenarios that have been collected both from control centers and the operation of highly automated vehicles (see Figure 4).

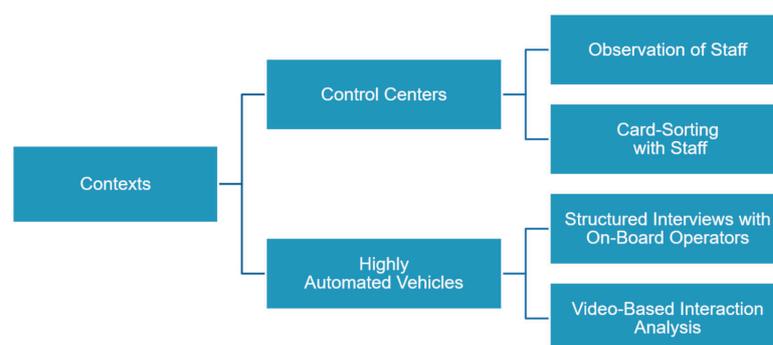


Figure 4. Contexts from which scenarios were extracted and methods used in the process.

2.1.1. Control Centers

To date, many real-world labs have tested automated shuttles with on-board operators in order to expand public transport services. AVs in real-world labs are usually operated with the assistance of a steward on board the shuttle who, among other things, monitors the traffic situation and the technical vehicle status and, if necessary, intervenes and initiates appropriate measures depending on the situation.

Remotely operated shuttles *without* onboard operators, however, will differ fundamentally regarding their interactions. Instead of interacting with the AV's on-board operators, the control center will interact directly with the AV while the tasks of the onboard operator will be shifted to the control center. However, only crude concepts and isolated prototypes for control centers to monitor, supervise, and, if necessary, remote-control AVs without on-board operators exist at the moment. Thus, in a first step, the roles and activities of today's control centers in public transport were analyzed by means of observation and interviewing. Participatory observation and expert interviews with control center staff in Hamburg and Braunschweig, Germany, helped examine the working equipment, tasks, roles, and collaborations in a control center for teleoperation in public transport in general. More importantly, these methods yielded scenarios with potential relevance for remote operation.

First, *observation* was used as a tool to collect essential data. It includes the description of behavioral and temporal patterns, the consequences for control center staff and their environment, as well as the spatial relationship of the control center employee with other people. Observations were characterized by the following attributes:

- *Time sampling*: Control center staff's behaviors were observed during a fixed time interval by researchers.
- *Unstructured observation*: Observations were conducted in a holistic way. The researcher entered the field with some general ideas of what might be salient, but not of what specifically will be observed, i.e., without using any pre-determined objectives, schedules, or variables (cf. [24]).
- *Naturalistic setting*: The process involved observing and studying the spontaneous behavior of the control center employees in their natural environment.

Second, based on these observations, several expert workshops yielded a set of categories which were subsequently used for two card-sorting studies. In a first study, interdisciplinary traffic researchers clustered the categories, assigned concrete task sets to them, and finalized them. In a second study, expert interviews were conducted using *the card-sorting approach* [20] to identify tasks and roles in future control centers.

2.1.2. Highly Automated Vehicles

Structured in-depth *interviews* were carried out with three onboard operators of automated shuttles (SAE Level Four [13]) integrated into Hamburg's public transport system as part of the HEAT project [11]. These interviews focused on control center tasks, disruptions, work experience, current and future workplace design [25]. From the disruptions mentioned, scenarios were derived.

Videoclips from the EU CityMobil2 project [1,3] were *analyzed* focusing on the interaction of AVs with other road users to generate scenarios. The main objective of CityMobil2 was to implement different demonstrations of AVs in five European cities as a part of local public transport [12]. Before the analysis, literature research and workshops with four traffic experts led to a set of categories that were applied to analyze the videos of AVs on the road. The categories were chosen to represent the events of interest as exhaustively as possible. The categories were mutually exclusive, precisely defined and their wording was simplistic. They included interactions with vehicles, pedestrians, cyclists, and infrastructure. In accordance with this categorization scheme, naturalistic video clips from the AV demonstrations were evaluated. They showed highly automated shuttles from the cities of La Rochelle, France, and Trikala, Greece, and were recorded as part of the EU project CityMobil2. The videoclips were shot independently from the raters who categorized

them and therefore not influenced by them. In order to categorize them, the videos were analyzed regarding incidents, the main events were noted down and grouped regarding the interactions of the AV with other road users, including vulnerable road users (VRUs), and the infrastructure, e.g., traffic lights. These interaction categorizations served as a basis for generating scenarios.

2.2. System of Structuring Scenarios

In order to highlight similarities among scenarios, a hierarchical structure with three levels is proposed. It consists, from top to bottom, of use case clusters (UCCs), use cases (UCs), and scenarios.

The terminology for these terms is based on Ulbrich et al. [26] and Wilbrink et al. [27]. It is illustrated in Figure 5 and will be defined in the following paragraphs.

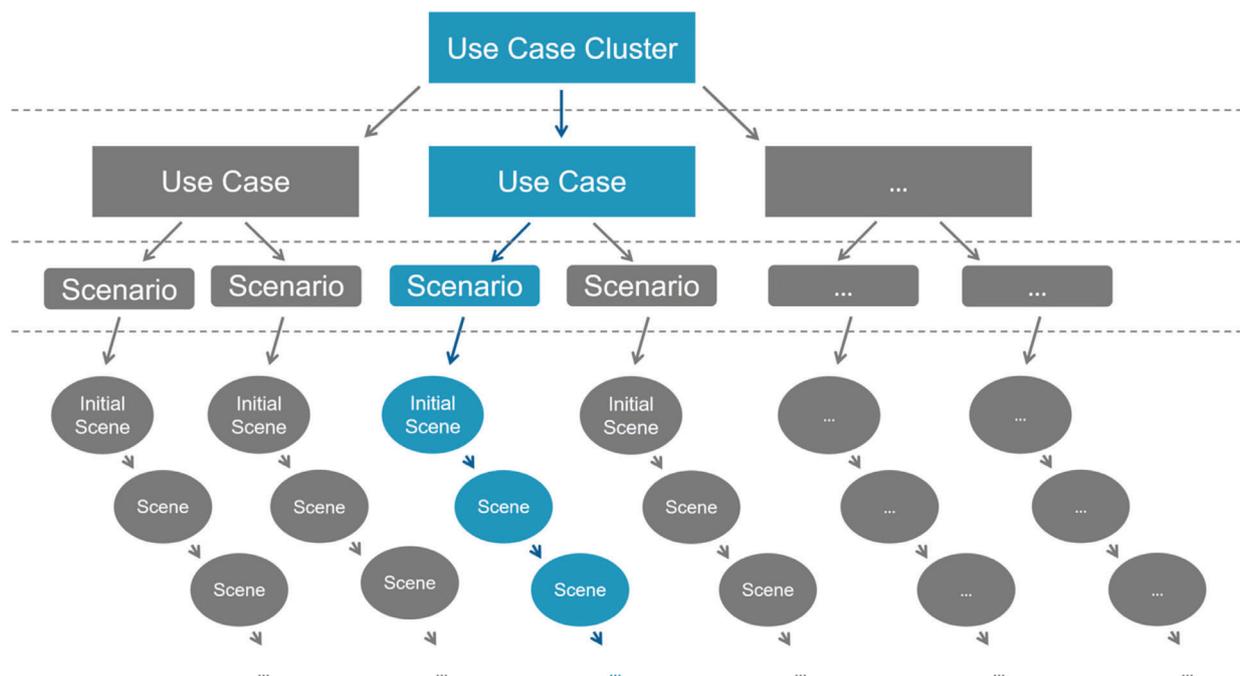


Figure 5. Relations between use case cluster (UCC), use case (UC), scenario, and scene. Adapted with permission from Ref. [27]. 2018, interACT.

A *scene* describes a snapshot of the environment. It includes a scenery (e.g., lane networks, stationary elements, and environmental conditions), dynamic elements (e.g., dynamic objects' states and attributes), self-representations of actors, and observers (e.g., actors' and observers' states and attributes, skills, and abilities) as well as the relationships between those entities. A *scenario* is defined as a temporal development of different scenes within a sequence of scenes. In order to characterize this temporal development, events and actions, as well as objectives, might be specified. Unlike a scene, a scenario describes a period of time. Scenarios start with an initial scene and can be visualized using interaction diagrams (cf. Figure 6). A *use case* is a functional description of a technical system and its behavior for a specific use. Use cases can comprise numerous different scenarios, but a scenario can only contain a certain number of scenes arranged in a certain order. A *use case cluster* comprises similar use cases.

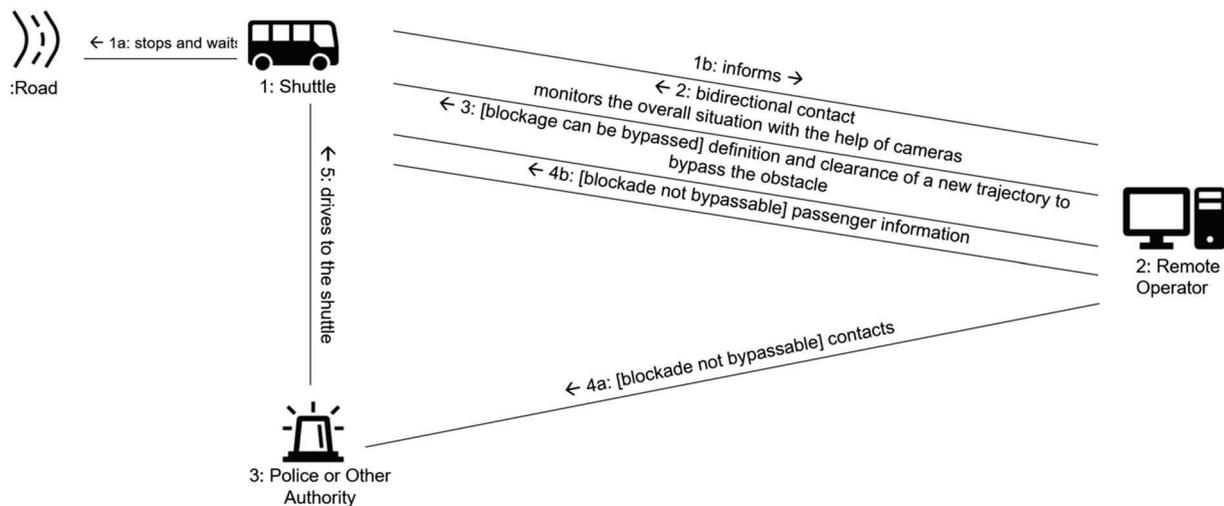


Figure 6. Interaction diagram for the scenario with the cause “Vehicles parked in second row”. The lane is blocked, e.g., due to vehicles parked in second row. This causes delays or disturbs the onward journey of the shuttle. The shuttle waits due to the blocked lane (1a). The shuttle informs the remote operator that no further travel is possible due to an obstacle (1b). The remote operator is in bidirectional contact with the passengers and switches to the shuttle’s cameras to get an overall view (2). The remote operator finds out whether it is possible to bypass the obstacle. The remote operator sets a new trajectory, e.g., by setting waypoints, selecting a trajectory, or steers the AV manually around the obstacle, gives clearance, and permits the AV to continue (3). If this is the case, the shuttle can continue its journey. If a bypass cannot take place due to e.g., constructional conditions, then the remote operator contacts the police or another authority (4a). The passengers are proactively informed about the further procedure and possible delays (4b). The police or another authority drive to the shuttle and solve the blockade (5).

For research on human-machine interaction, the focus on singular static scenes does not suffice to describe processes of interaction. On the other side, the system-based level applied in use cases is too abstract to pay enough attention to these processes. Therefore, the focus of this paper will be on scenarios. Particularly, it will list and categorize various scenarios in a uniform structure to lay the groundwork for a scaffolding of scenarios pertaining to remotely operating AVs.

The scenarios presented in this paper were compiled based on interviews and observations in control centers of public transport (see previous section). They are inspired by Geis and Tesch’s notion of *Is Scenarios* [28]. These are events or chains of events occurring in a naturalistic setting. They are characterized by a “narrative, textual description of actions that a certain user applies in order to attend to one or several tasks (*translated by the authors*)” [28] (p. 71). They describe components of the context of use in interplay with the perspective of the interviewed or observed person. According to Dzida and Freitag [29], *Is Scenarios* are the central source for identifying demands and deriving user requirements. Even though tasks may be included in an *Is Scenario*, they are not in the focus—unlike in *Use Scenarios*, which describe the implementation of tasks by the user. Rather, as *Is Scenarios* investigate the interrelations between tasks, the relevance of specific resources is elucidated. This, in turn, may facilitate the identification of previously concealed demands. Furthermore, *Is Scenarios* may help surface the intertwining of various actors.

As processes of interaction are vital to understand what is happening in urban mixed traffic settings including remote-controlled HAVs, they are used here to provide a scaffolding on the uppermost level, i.e., the level of use case clusters. On this top level, *actors* play a significant role. This paper defines “actor” in accordance with the United Modeling Language. Thus, an actor “specifies a role played by a user or any other system that interacts with the subject” [27] (p. 586). It emphasizes that actors are not limited to human users

such as remote operators: “Actors may represent roles played by human users, external hardware, or other subjects” [27] (p. 586).

Following this definition, the following actors are used in the compilation of scenarios:

- *Remote Operator*, who may be both *Remote Driver* or *Remote Assistant* in SAE’s [8] terminology,
- *Highly Automated Vehicles* at SAE Level 4,
- *Passengers*,
- *Infrastructure*, e.g., crosswalks, traffic lights, or intelligent road-side infrastructures (including road-site units, i.e., sensors along the road that scan the traffic in their immediate surroundings and submit this information to a traffic management center, cf. [28]), and
- *Other Actors*, e.g., emergency services, control center staff, and other road users.

All of these five actors may be interacting with any other actor in a given scenario within a specific context that influences them. Figure 7 includes the most relevant actors and their interrelations.

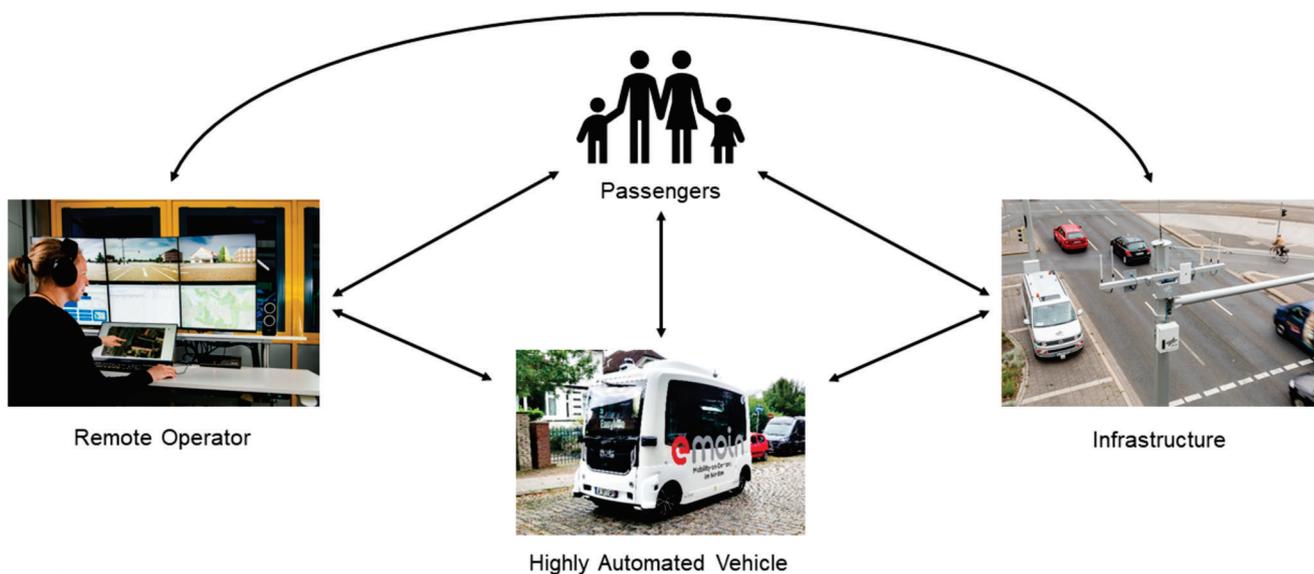


Figure 7. Most relevant actors interacting in the compiled scenarios on remote operation of highly automated vehicles. The remote operator, the highly automated vehicle, the passengers, and the infrastructure collaborate with each other to complete the driving task at SAE Level 4. Adapted with permission from Ref. [30]. 2022, DLR (CC BY-NC-ND 3.0).

On the second-to-top level, scenarios are grouped into use cases. Here, a use case is defined as a functional description of a technical system and its behavior for a specific use. Use cases can comprise numerous different scenarios, but a scenario can only contain a certain number of scenes arranged in a certain order [14].

On the third-to-top level, scenarios are listed. In order to facilitate comparability, every scenario is structured in a chain of cause, event, and consequence, as Figure 7 represents. The sequence consists of the following elements: a cause that the event is attributed to, the central event, and the consequences that arise from it.

A generic template is proposed that is used for every scenario:

Due to <Cause>, <Event> takes place. This results in <Consequence>.

The event and its consequence, in turn, lead to certain measures that are required to resolve the event. These required measures, however, are not part of the presented catalogue of scenarios. Adding them would be beyond the scope of this paper since its focus is on events in remote operation, their causes, and consequences. The required measures will be addressed in future publications.

It is important to note that the scenario does not take place in ignorance of the concrete context in which it happens. Rather, all its stages are embedded in this context (see Figure 8). Additionally, the actors that interact throughout the scenario represent another level of analysis that accompanies the chain of cause, event, and consequence, and might also be involved in the measures required for resolving the event.

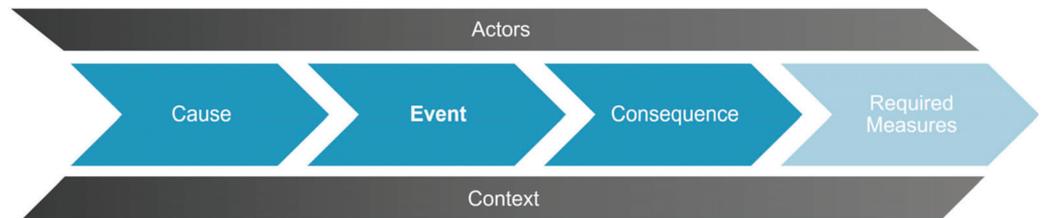


Figure 8. The chain of cause, event, and consequence provides the scaffolding for each scenario. In future iterations of the scenario catalogue, required measures may be added.

3. Results

The following section outlines the structure of the scenario catalogue, provides the entire compilation of scenarios, and concludes with an exemplary scenario and its related interaction diagram.

3.1. Structure and Catalogue of Scenarios

Figure 9 presents an organigram of the structure of the compiled scenarios of scenarios, organized in use case clusters and use cases. The *central interactions* can be considered the main body of the scenario collection. They are structured in a way that enables an interaction of one actor with any of the remaining ones. In addition, use case clusters (UCCs) regarding the remote operator’s state, contextual factors, and technical malfunctions related to *peripheral factors* that are not directly based on interactions.

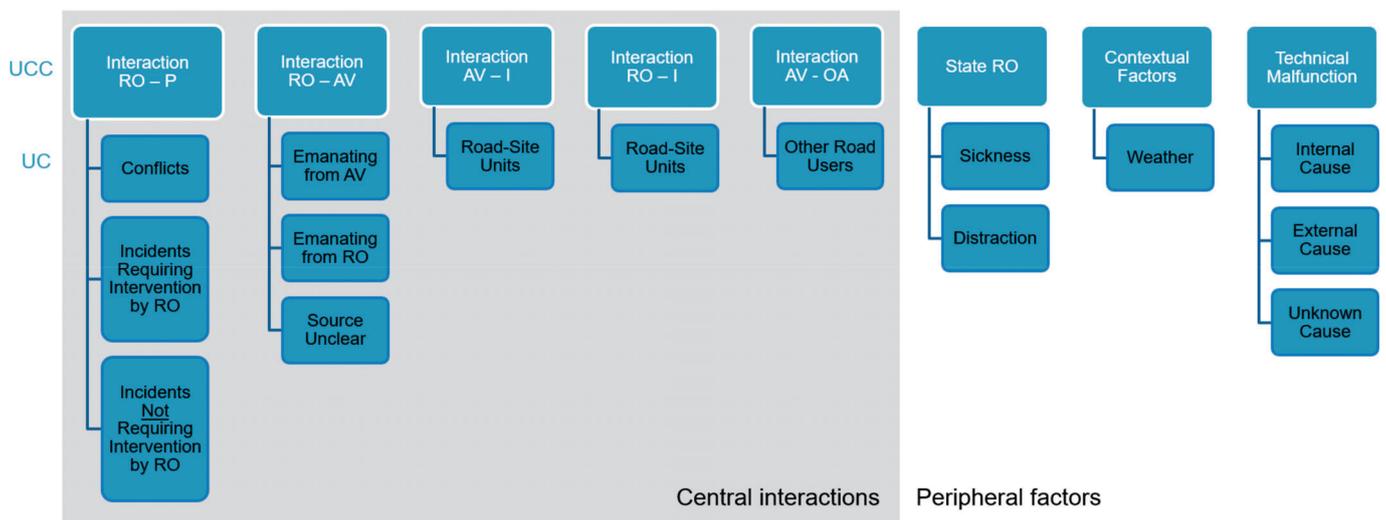


Figure 9. Structure of use case clusters (UCCs, top row) and use cases (UCs, rows below top row). The core of the scenario collection is made up of interactions between different actors (central interactions). In addition, UCCs regarding the remote operator’s state, contextual factors, and technical malfunctions related to peripheral factors that are not directly based on interactions. RO = Remote Operator, AV = Highly Automated Vehicle, P = Passengers, I = Infrastructure, OA = Other Actors.

Table 1 is a comprehensive list of scenarios in remote operation compiled. It is structured as follows: the overarching classification categories use case cluster and use case, the defining elements of the scenario consisting of cause, event, and consequence, a column for every of the five actor categories, and the mode of remote operation.

Table 1. Comprehensive list of scenarios in remote operation compiled. A classificatory number indicates the use case cluster (UCC), use case (UC) and Is Scenario (Sc) in the following way: <N_{UC}> . <N_{UC}> . <N_{UC}> . × indicates that an option applies for a given scenario.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors						Mode of ROn ¹				
						RO _I	AV _I	P ¹	I ¹	QA _I	RD _I	RA _I	RO _I	RD _I	RA _I	
1 Interaction RO with Passengers	1.1 Conflicts	Vandalism	AV ¹ is damaged (inside and/or outside)	AV cannot continue its ride.	1.1.1 Due to vandalism, the AV is damaged (inside and/or outside). This results in the AV being unable to continue its ride.	X	X					X	X	X		
			Passenger abuses intercom	Control center employee is unnecessarily distracted	1.1.2 Due to an undefined cause, a passenger abuses the intercom. This leads to unnecessary distraction of the control center staff.	X		X					X	X	X	
		NA ¹	Passenger does not want to get off	Delay	1.1.3 Due to an undefined cause, a passenger does not want to get off the AV. This leads to a delay of the AV.	X		X					X	X	X	
		NA	Disputes between passengers	AV cannot continue its journey	1.1.4 Due to an undefined cause, there are disputes between the passengers. This leads to the AV not being able to continue its journey.	X		X					X	X	X	
	1.2 Incidents requiring intervention by RO	Mobility-impaired passenger would like to use AV	Mobility-impaired passenger receives support from the control center for using the AV	Mobility-impaired passenger receives support from the control center for using the AV	Mobility-impaired passenger uses AV	1.2.1 Due to a request of a mobility-impaired passenger, they receive support to use the AV. This leads to the mobility-impaired passenger being able to use the AV.	X	X	X				X	X	X	
				Control center contacts passengers	Passengers is able to appropriately react to emergency	1.2.2 Due to an emergency, the control center contacts the passengers. This leads to the passengers becoming enabled to appropriately react to the emergency.	X		X					X	X	X
		Emergency	RO opens the door between regular stops	Passengers can leave the AV between regular stops	Passengers can leave the AV between regular stops	1.2.3 Due to an emergency, the doors of the AV are opened by the RO between regular stops. This leads to the fact that passengers can leave the AV between regular stops.	X	X	X				X	X	X	
				Passenger needs immediate medical treatment	Onward journey might be delayed	1.2.4 Due to a medical emergency in the AV, a passenger needs immediate medical treatment. This might delay the onward journey.	X		X					X	X	X
		NA	Passenger/RO finds ownerless suitcase	Passenger/RO finds ownerless suitcase	Possible security risk	1.2.5 Due to an undefined cause, a passenger or the RO finds an ownerless suitcase. This may lead to a security risk.	X		X					X	X	X
				AV stops	Onward journey of the AV delayed, restricted or not possible	1.2.6 Due to a technical malfunction, the AV stops. This leads to the fact that the onward journey of the AV is delayed, restricted or not possible.	X		X					X	X	X

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors					Mode of ROn ¹		
						RO _I	AV _I	P ¹	I ¹	O _A	RD _I	RA _I	
		NA	Passenger would like to exit between regular stops	Onward journey delayed	1.2.7 Due to an undefined cause, a passenger wants to exit the AV between regular stops. This leads to a delay in the onward journey of the AV.	X		X				X	
		Bulky objects inside the AV	Doors blocked	Doors do not close, AV cannot continue its journey, onward journey delayed	1.2.8 Due to bulky objects inside the AV, the doors are blocked. This leads to the doors not closing and the AV not being able to continue its journey. The onward journey of the AV is delayed.	X	X	X				X	X
		AV crowded	Passengers block the door	Doors do not close, AV cannot continue its journey, or onward journey is delayed	1.2.9 Due to the AV being crowded, passengers block the door. This leads to the doors not closing and the AV not being able to continue its journey or the onward journey being delayed.		X	X				X	X
1.3 Incidents not requiring intervention by RO		Spilled liquids, scattered food	AV's interior dirty	Passenger satisfaction declines	1.2.10 Due to spilled liquids or scattered food in the interior of the AV, the AV's interior is dirty. This leads to declining passenger satisfaction.		X	X				X	X
		Emergency	Passenger applies emergency brake	AV stops, door opens, passengers can get off	1.3.1 Due to an emergency, a passenger applies the emergency brake. This leads to the doors opening and the passengers being able to get off.		X	X				X	X
	2	2.1 Emanating from AV	Sensors detect obstacle	AV cannot continue its journey/onward journey is delayed, AV waits for clearance from RO until it can resume ride	2.1.1 Due to obstacles detected by sensors of the AV, the AV stops. This leads to the fact that the AV cannot continue its journey or the journey being delayed. The AV waits for the RO to give clearance so it can resume its ride.	X	X					X	X
Interaction RO with AV		Sensors detect obstacle (e.g., other road users, parked vehicles, trees, foliage, animals), technical malfunction	AV stops	AV cannot continue its journey/onward journey is delayed, AV waits for the RO to give clearance, or select a trajectory or set waypoints so that AV can resume ride	2.1.2 Due to obstacles detected by sensors of the AV, the AV stops. This leads to the fact that continuing the journey is not possible or delayed. The AV waits for the RO to give clearance, select a trajectory or set waypoints so that it can resume its ride.	X	X				X	X	

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors					Mode of ROn ¹		
						RO _I	AV _I	P ₁	I ₁	O _A	RD _I	RA _I	
		Technical malfunction	RO receives an error/malfunction message	Resolving the error/malfunction	2.1.3 Due to a technical malfunction, the RO receives an error/malfunction message. This leads to the RO resolving the error/malfunction.	X	X					X	X
		AV is partially defective but still operable	AV stops	AV cannot resume ride/is delayed. AV waits for the RO to give clearance, select a trajectory or set waypoints so that AV can resume ride. Resuming ride only possible for AV until a safe stopping point with restrictions	2.1.4 Due to the AV being partially defective but still operable, the AV stops. This leads to the fact that resuming the ride is not possible for the time being and is delayed. The AV waits for the RO to give clearance, select a trajectory or set waypoints so that the AV can continue its ride to a safe stopping point.	X	X						X
		Objects or parts of the AV on fire	Fire alarm activated	AV stops, cannot continue its journey	2.1.5 Due to objects or parts of the AV being on fire, the fire alarm is activated. This results in the AV stopping and not being able to continue.		X	X				X	X
		Smoke emitted by passengers from cigarette, e-shisha, etc.	Fire alarm activated	AV cannot continue its journey or onward journey is delayed	2.1.6 Due to passengers emitting smoke from cigarettes, e-shishas, etc. in the AV, the fire alarm is activated. This results in the AV stopping and not being able to continue or the onward journey being delayed.		X	X				X	X
		Technical malfunction	RO brakes too hard	Wheels block, critical situation or accident may occur	2.1.7 Due to a technical malfunction of the AV, the RO brakes too hard. This makes the wheels block, a critical situation or accident may occur.		X	X					X
		Unforeseen situation (traffic jam, route change, dispatching error, etc.)	AV battery's state of charge low	AV cannot reach the planned charging station or AV cannot complete the roundtrip as planned	2.1.8 Due to an unforeseen situation (traffic jam, route change, dispatching error, etc.), the battery has a low state of charge. This leads to AV not being able to reach the planned charging station or to complete the roundtrip as planned.		X	X				X	X
		AV conducted roundtrip with high energy consumption, e.g., air conditioning, traffic jam	State of charge not sufficient for another roundtrip	AV needs to be charged earlier than planned	2.1.9 Due to a roundtrip with high energy consumption, the state of charge is not sufficient for another roundtrip. This leads to the fact that AV needs to be charged earlier than planned.		X	X				X	X

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors					Mode of ROn ¹		
						RO _I	AV _I	P ₁	I ₁	O _A	RD _I	RA _I	
	Technical malfunction		Charging not possible	AV cannot continue its journey	2.1.10 Due to a technical malfunction, charging the AV's battery is not possible. This leads to the AV not being able to continue its journey.	X	X		X			X	X
	Demand for charging stations exceeds supply		No free charging stations	Charging not possible	2.1.11 Due to the demand for charging stations exceeding the supply, the AV cannot find a free charging station. This leads to the fact that the AV cannot be charged.	X	X		X			X	X
	Allocated charging station is out of order		Charging not possible	AV cannot continue its journey	2.1.12 Due to the allocated charging station being out of order, charging is not possible. This leads to the fact that the AV cannot continue its journey.	X	X		X		X	X	X
2.2 Emanating from RO	RO's misbehavior		RO is inattentive or distracted, does not recognize street signs with traffic rules	AV overrides RO's input because it is no longer compliant with traffic rules	2.2.1 Due to misbehavior of the RO during direct control, the inattentive or distracted RO does not recognize the street signs. This leads to the AV overriding the RO's input because it is no longer compliant with the traffic rules.	X	X						X
	Accident on driving route		Closure of parts of the route	AV cannot continue route as planned (with or without stopping), planned route is adjusted	2.2.2 Due to an accident on the AV's route, there is a closure of parts of the AV's planned route. This means that the AV cannot continue its route as planned (with or without stopping). The route must be adjusted.	X	X						X
	Major event, protest, construction site, etc.		Change of route	Changed order of stops; stops may be omitted, RO contacts passengers	2.2.3 Due to a major event, protest, construction site, etc., there is a change in the route. This leads to a change in the order of the stops and, if necessary, stops might be skipped. The RO contacts the passengers and lets them know about the changes.	X	X		X				X
	RO's misbehavior		RO brakes too hard	Wheels block, accident	2.2.4 Due to the misbehavior of the RO, the AV brakes too hard. This leads to the wheels blocking and the risk of a critical situation or an accident.	X	X						X
	Distraction RO		RO accidentally goes off-track	Critical situation or accident	2.2.5 Due to a distraction of the RO during remote driving, the AV goes off-track. This leads to a critical situation or an accident.	X	X						X

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors						Mode of ROn ¹				
						RO _I	AV _I	P ₁	I ₁	O _A	RD _I	RA _I	RO _I	RD _I	RA _I	
2.3 Source unclear		Technical malfunction, latency too high	RO accidentally goes off-track	Critical situation or accident	2.2.6 Due to a technical malfunction and a too-high latency in direct control, the AV accidentally goes off-track. This leads to a critical situation or an accident.	X	X							X		
		AV is urgently needed	No complete charging possible	Range of the AV is restricted	2.2.7 Due to an urgent need of the AV, it is not possible to fully charge the AV's battery. This leads to a limited range of the AV.	X	X							X	X	
		Soiling of the AV's interior camera	Camera images from the AV's interior cannot be recognized by RO	RO may not notice passenger interactions (e.g., disputes)	2.3.1 Due to a soiled camera inside the AV, the RO cannot see the camera images inside the AV. This leads to the RO not noticing passenger interactions (e.g., disputes).	X	X							X	X	X
		Soiling of the sensors	Sensors limited in their functionality	AV cannot continue its journey or onward journey delayed	2.3.2 Due to soiled sensors on the AV, the functionality of the sensors is limited. This leads to the AV not being able to continue its journey or in a delayed onward journey.	X	X							X	X	X
		Connectivity issues between control center and AV	RO loses connection to AV	RO cannot establish contact with AV	2.3.3 Due to connectivity issues between the control center and the AV, the RO loses connection to the AV. This results in RO no longer being able to contact the AV.	X	X							X	X	X
3 Interaction AV with infrastructure	3.1 Road-site units	Connectivity issues of audio stream from AV to control center	Strong noise interference	RO cannot understand passengers	2.3.4 Due to connectivity issues of the audio stream between the control center and the AV, there is strong noise interference, resulting in the RO not being able to understand the passengers.	X	X							X	X	
			Desired camera image is not transmitted in sufficient quality	AV can only continue journey with restrictions	2.3.5 Due to connectivity issues of the video stream between the control center and the AV, the desired camera image is not transmitted in sufficient quality. This leads to the fact that the AV can only continue its journey to a limited extent.	X	X							X	X	
			Road-site units support AV while maneuvering through unmapped construction site	AV can continue its journey without stopping	3.1.1 Due to an unmapped construction site, road-site units support the AV while maneuvering through the unmapped area. This leads to the AV being able to continue its journey without stopping.										X	X

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors					Mode of ROn ¹			
						RO _I	AV _I	P ¹	I ¹	O _A _I	RD _I	RA _I		
4 Interaction RO with in- frustructure	4.1 Road-site units	Road infrastructure is disrupted	Unstable/disturbed connection from the AV to the road-site units	RO is informed and forwards error so that it can be fixed as quickly as possible	4.1.1 Due to a disruption of the road infrastructure, the connection from the AV to the road-site units is unstable or disturbed. This leads to RO being informed about the unstable or disturbed connection. The RO forwards the fault to the relevant actors as quickly as possible so that the faults can be fixed as soon as possible.	X	X					X	X	X
5 Interaction AV with other actors	5.1 Other road users	Vehicle or VRU is occluded by object or outside the focus area	AV reacts abruptly to suddenly appearing vehicle or VRU	Critical situation or accident	5.1.1 Due to the occlusion of a vehicle or VRU ¹ by another object or because a vehicle or VRU is outside of the focus area, the AV reacts abruptly to the suddenly appearing vehicle or VRU. This leads to a critical situation or an accident.		X			X				
		Several vehicles reach an intersection without a right-of-way sign at the same time and congest it	Deadlock occurs at an intersection without a right-of-way sign (left-yields-to-right rule)	RO gives clearance	5.1.2 Due to vehicles reaching the intersection from all directions at the same time, a deadlock occurs. This leads to the RO giving clearance and permitting the AV to proceed if the other road users do not resolve this situation beforehand.	X	X			X				X
		VRU or animal suddenly or unexpectedly crosses the intersection	AV performs emergency braking maneuver	Critical situation or accident	5.1.3 Due to a suddenly crossing or unexpected VRU or animal, the AV has to perform an emergency braking. This leads to a critical traffic situation or an accident.	X	X			X			X	
		VRU's misbehavior or technical malfunction	AV collides with VRU	AV cannot continue its journey	5.1.4 Due to a misbehavior of the VRU or a technical malfunction, the AV collides with a VRU. This leads to the AV being unable to continue its journey.	X	X			X			X	X
	Vehicles parked in second row	Pathway blocked		Continuation of ride delayed; RO assesses situation and guides AV around the obstacles	5.1.5 Due to vehicles parked in the second row, the AV's pathway is blocked. This leads to a delay in the onward travel of the AV. The RO gives clearance to the AV to continue, sets a new pathway, e.g., by setting waypoints or selecting a pathway, or steers the AV manually around the obstacles.	X	X			X			X	X
	Accident		AV stops	AV cannot continue its journey	5.1.6 Due to an accident, the AV stops. This leads to the AV not being able to continue its journey.	X	X			X			X	X

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors					Mode of ROn ¹		
						RO _I	AV _I	P ¹	I ¹	O _A _I	RD _I	RA _I	
6 State RO	6.1 Sickness	Driver's distraction, difficult weather conditions, etc.	Collision of another vehicle with AV	Damage to the AV by another vehicle, AV cannot continue its journey, damage must be fixed and passengers must be informed	5.1.7 Due to the driver's distraction, difficult weather conditions, etc., another vehicle collides with the AV. This leads to the AV being damaged by the other vehicle and being unable to continue its journey. The RO conducts procedure as with accidents in general (i.e., contacts passengers, informs blue light organization, organizes a substitute vehicle if necessary, documentation, etc.).	X	X	X	X	X	X	X	X
						Major event, protest, construction site, etc.	Traffic jam	Onward journey is delayed	5.1.8 Due to a major event, protest, construction site, etc., a traffic jam occurs. This leads to a delay in the onward journey of the AV.	X	X	X	X
6 State RO	6.2 Distraction	NA	RO sick	RO cannot fulfill supervisory duty	6.1.1 Due to an undefined cause, the RO is sick. This leads to the RO being unable to fulfill their supervisory duties or to perform their tasks.	X	X						X
						Fatigue, stress, distraction in general, etc.	RO distracted	RO cannot fulfill supervisory duty	6.1.2 Due to fatigue, stress, distraction in general, etc., the RO is distracted. This leads to the RO being unable to fulfill their supervisory duties or to perform their tasks.	X	X		
7 Contextual Factors	7.1 Weather conditions	Difficult weather conditions	Overload or soiling of the sensors on the vehicle	AV cannot continue its journey	7.1.1 Due to difficult weather conditions, the sensors on the AV are overloaded or soiled. This leads to the fact that they cannot detect the environment accurately so the AV cannot continue its journey.	X	X						X
						Difficult weather conditions, such as icy roads, fog, heavy rain, or snowfall	AV goes off-track, no other parties involved	Critical situation or accident, onward journey is delayed, restricted, or not possible	7.1.2 Due to difficult weather conditions, such as icy roads, fog, heavy rain, or snowfall, the AV goes off-track. There are no other parties involved. This leads to a critical situation or accident and to the fact that the continuation of the AV's journey is delayed, restricted, or not possible.		X		
						Difficult weather conditions, such as icy roads, fog, heavy rain, or snowfall	AV comes off the road, there are other parties involved	Accident with internal and external damage, onward journey is delayed, restricted, or not possible	7.1.3 Due to difficult weather conditions, such as icy roads, fog, heavy rain, or snowfall, the AV goes off-track. There are other people involved. This leads to an accident with internal and external damage and the fact that the continuation of the AV's journey is delayed, restricted, or not possible.		X		X

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors					Mode of ROn ¹				
						RO _I	AV _I	P ₁	I ₁	O _A	RD _I	RA _I			
		Flood, fire, etc.	Traffic routing changed, road closed, speed limit changed	Planned route changed, planned stops skipped	7.1.4 Due to a flood, fire, etc., for example, the traffic routing was changed, a road was closed, or the speed limit was changed. This leads to the planned route being changed and planned stops being skipped.	X	X							X	
8 Technical malfunction	8.1 Internal cause	Technical malfunction	Doors do not close	AV cannot continue its journey	8.1.1 Due to a technical malfunction, the doors of the AV do not close. This leads to the AV not being able to continue its journey.		X	X							
		Technical malfunction	Doors do not open	Passengers can neither get on nor off	8.1.2 Due to a technical malfunction, the doors of the AV do not open. This leads to passengers not being able to get on or off and the AV not being able to continue its journey.		X	X							
		Leakage in tank or pipes	AV loses liquid	AV cannot continue its journey; may obstruct other road users	8.1.3 Due to a leakage in the tank or pipes, the AV loses liquid. This leads to the fact that it is not possible for the AV to continue its journey and other road users being obstructed.		X				X				
		GPS receiver malfunction or connectivity issues	AV cannot be localized	Monitoring and controlling of the AV by RO not possible	8.1.4 Due to a defective GPS receiver or connectivity issues, the AV cannot be localized. This leads to the fact that monitoring and controlling the AV by the RO is not possible.	X	X					X		X	
		Technical malfunction	AV's fire detector disturbed	Passengers' safety cannot be guaranteed	8.1.5 Due to a malfunction, the AV's fire detector is disturbed. This leads to the fact that the safety of the passengers cannot be guaranteed.	X	X						X		X
		Technical malfunction	Air conditioning out of order	Overheating of the passenger cabin; reduction of comfort or impairments of passengers' health	8.1.6 Due to a technical malfunction, the air conditioning in the AV is out of order. This leads to an overheating of the passenger cabin and to a reduction in comfort or impairments of passengers' health.		X	X							
		Technical malfunction	Signal and/or brake light fails	RO receives an error message	8.1.7 Due to a technical malfunction, a signal and/or brake light of the AV fails. This leads to the RO receiving an error message.	X	X							X	X

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors					Mode of ROn ¹			
						RO _I	AV _I	P ₁	I ₁	O _A	RD _I	RA _I		
	Technical malfunction	Technical malfunction	Headlight cleaning system defective	Limited visibility of the AV	8.1.8 Due to a technical malfunction, the headlight cleaning system of the AV is defect. This leads to limited visibility of the AV.		X							
	Technical malfunction	Technical malfunction	Problems with thermal management	Danger of overheating of the vehicle's battery or its interior	8.1.9 Due to a technical malfunction, the AV has problems with the thermal management. This leads to an overheating of the vehicle's battery or its interior.		X		X					
	Failure of parts of the sensors	MRM ¹		AV cannot continue its journey and may endanger passengers and other road users due to AV's location on the road	8.1.10 Due to a partial failure of the AV's sensor system, the AV performs an MRM. This leads to the AV being unable to continue its journey and possibly endangering passengers and other road users due to AV's location on the road.		X		X					
	Interface between AV and control center disturbed	Interface between AV and control center disturbed	Target and actual direction of travel do not align	AV drives on the wrong pathway	8.1.11 Due to a malfunction in the interface between the AV and the control center, the AV's planned and actual driving directions do not align. This leads to the AV driving on the wrong pathway.		X		X					X
	Interface between AV and control center disturbed	Interface between AV and control center disturbed	Target and actual speed do not align	AV drives at the wrong speed	8.1.12 Due to a malfunction in the interface between the AV and the control center, the desired and actual speed of the AV do not align. This leads to the AV driving at the wrong speed. The AV decides to adapt the speed according to the traffic regulations.		X		X					X
8.2 External cause	Spiky objects on the roadway	Spiky objects on the roadway	Damaged tire	AV can continue to drive if necessary but with increased risk of accident; onward journey is delayed	8.2.1 Due to spiky objects on the road, a tire of the AV is damaged. This leads to the AV being able to continue driving if necessary but with increased risk of accident. The onward journey is delayed.						X			
	Collision with an obstacle: vehicle hits curb or boundary of construction site	Collision with an obstacle: vehicle hits curb or boundary of construction site	Damage to the AV, possibly further damage	AV cannot continue its journey	8.2.2 Due to a collision with an obstacle, such as a curb or a boundary of a construction site, the AV is damaged. Further damage may occur. This leads to the AV not being able to continue its journey.						X			
8.3 Unknown cause	NA	NA	AV stops	AV cannot continue its journey; onward journey is delayed	8.3.1 Due to an undefined cause, the AV stops. This leads to the fact that the AV cannot continue its journey or the onward journey is delayed.						X			

Table 1. Cont.

Use Case Cluster	Use Case	Cause	Event	Consequence	Description ² (Is Scenario)	Actors						Mode of ROn ¹		
						RO	AV	P	I	OA	RD	RA	I	I
		Technical malfunction	AV cannot be controlled remotely	AV may endanger passengers and other road users	8.3.2 Due to a technical malfunction, the AV cannot be controlled remotely. This leads to the AV not being able to continue driving or the onward journey being delayed. Passengers and other road users may be endangered.	X	X						X	X
		Connectivity issues	RO cannot connect to AV	AV no longer complies with all safety regulations, may endanger passengers and other road users	8.3.3 Due to connectivity issues, the RO cannot connect to the AV. This leads to the fact that the AV no longer complies with all safety regulations. Passengers and other road users may be endangered.	X	X						X	X
		Technical malfunction or transmission error	RO fails to identify the road signs	Onward journey is delayed	8.3.4 Due to a technical malfunction or transmission error, the RO does not recognize the road signs. This leads to the onward journey being delayed.	X	X							X
		Poor object identification	Unclear detection situation	AV cannot continue its journey, AV waits for the RO to give clearance so it can continue journey	8.3.5 Due to poor object identification, the detection situation of the AV is unclear. This leads to the AV not being able to continue its journey. The AV waits for the RO to give clearance so it can continue its journey.	X	X							

¹ NA = Not specified, AV = Highly Automated Vehicle (SAE Level 4), I = Infrastructure, MRM = Minimal Risk Maneuver, OA = Other Actors, P = Passengers, RA = Remote Assistance, RD = Remote Driving, RO = Remote Operator, ROn = Remote Operation, VRU = Vulnerable Road User (e.g., pedestrian, cyclist). ² According to standardized template.

Overall, 74 scenarios were compiled. These were subsumed under 15 use cases, which in turn were grouped into eight use case clusters, five of which are central interactions, and the remaining three are regarded as peripheral factors. Both use cases and use case clusters are stated in Figure 8. A sequential number (N) was given to each use case cluster (UCC), use case (UC), and Is Scenario (Sc), in the following fashion:

$$\langle N_{UCC} \rangle . \langle N_{UC} \rangle . \langle N_{Sc} \rangle$$

From left to right, Table 1 includes the following columns: Use case cluster, Use case, Cause, Event, Consequence, and the descriptive Is Scenario. In the section “Actors”, each actor involved is checked with an “X” if involved. If not involved, the respective column remains blank. The same system is used to indicate the Mode of Remote Operation. Here, an “X” indicated that the scenario is valid for Remote Driving, Remote Operation, or both.

3.2. Example for Scenario “Vehicles Parked in Second Row”

An example for a scenario is from the Use Case Cluster “Interaction AV with Other Actors”, Use Case “Other Road Users”. The Causes in this scenario are “Vehicles parked in second row”. This leads to the Event “Driveway blocked” which in turn triggers the following Consequence: “Continuation of ride delayed; RO needs to assess situation and intervene”. The following Actors are involved in this scenario: Remote Operator, Highly Automated Vehicle, and Other Actors. The scenario may occur both in Remote Driving and Remote Assistance. Finally, the full description of an Is Scenario reads as follows:

“Due to vehicles parked in the second row, the AV’s lane is blocked. This leads to a delay in the further travel of the AV. The RO allows the AV to continue, sets a new trajectory, e.g., by setting waypoints or selecting a pathway, or steers the AV manually around the obstacle.”

Figure 6 above shows an interaction diagram for this scenario. It displays the main actors in this scenario. Every scenario in Table 1 can be translated into an interaction diagram like this.

4. Discussion

This paper proposes an initial draft for a catalogue of scenarios that might help when creating design solutions for the remote operation of AVs. It is published preliminarily with a remaining need for evaluation, validation, and modification in future iterations. Even though the scenarios presented here originate from diverse sources, particularly real-world labs, idiosyncrasies of other contexts of use will need to be considered by revalidating and extending the catalogue.

In spite of these constraints, the catalogue fulfills several purposes. First, it serves as a starting point for *designing novel interfaces* for remote-controlling highly automated vehicles—and has, in fact, already done so. Based on these scenarios, the authors of this paper designed an HMI for a workplace for remote operation in an online study with experts employed by control centers in public transport [21]. Incorporating the results from the evaluation study, a workplace for remotely assisting AVs has been set up at DLR’s Braunschweig premises.

Second, the catalogue is suitable to *test and validate HMIs* in teleoperations of means of public transport. For instance, the workplace described above will be tested and validated using the catalogue of scenarios presented here. Thanks to the workplace’s integration in realistic road simulations and DLR’s fleet of highly automated vehicles [31], a validation study with high ecological validity will be carried out. Using both quantitative, performance-based indicators and qualitative interview and questionnaire methodology, a group of experts from public transportation facilities and associations and other potential users of a remote operation workplace was exposed to a selection of the scenarios presented here [32]. Hence, a bidirectional relationship is established between the scenarios and the workplace: Not only will the study guide the process of improving the workplace for

the needs of the remote operator to execute their tasks safely, effectively, efficiently, while ensuring an optimal task load, keeping the operator in the loop and preventing fatigue and monotony. It will also help reassess and refine the compiled scenarios.

Third, *Operational Design Domains (ODDs) may be derived* from the scenarios. Thus, the catalogue will help specify the context and boundaries of safe teleoperation and create if-then contingencies between a certain contextual factor and its adequate driving mode. Hence, the remote operation can be incorporated in a wider framework of highly automated driving that encompasses different modes of operation, e.g., relying on input from the driving automation [33] and the infrastructure [34,35] in addition to remote operation. This multimodal, holistic approach is conceptualized within the framework of Managed Automated Driving [36].

Fourth, the catalogue will help to *identify the most safety-relevant scenarios* for teleoperation in public transport. This will be done, for example, by conducting a study that makes experts and operators in public transportation review the scenarios and have them rate (1) the probability of a scenario's occurrence and (2) its criticality for safe remote operation. A resulting priority classification will help understand the most pressing safety-relevant scenarios, among others, and provide a pathway to address them effectively. Subsequently, they will be used to derive user requirements to further *improve the existing HMI* of DLR's prototypical remote operation workstation. In addition, they will also help *create adaptive HMIs* solutions that consider the remote operator's current state and adapt the interface to accommodate it. This approach is pursued within the European Union's Horizon 2020 project "Hi-Drive", among others, and may be suitable to defragment the transition between various operational contexts [37].

Fifth, the identified user requirements may also facilitate creating a checklist for assessing the quality of a remote operator's workstation from a Human Factors perspective. Critically, *guidelines for selecting and training remote operators* could be interpolated from these requirements. This is highly relevant to rolling out highly automated vehicles since the role of having a Technical Supervisor to remotely monitor AVs and intervene, if necessary, i.e., a remote operator, was put forward as a requirement for SAE Level Four driving operations by several legislators on the German and European level [11,38]. Thus, remote operation is likely to be an inevitable prerequisite for highly automated driving.

Finally, the collection of scenarios may also be of importance for *mobility research in adjacent disciplines*. For instance, it could contribute to IT mobility services that require a holistic user-centered view on future mobility systems, embedded in a network of various means of transport, and feed data into a comprehensive mobility data space that is used to exchange mobility data. This is currently examined by the project "GAIA-X 4 ROMS" that aims at supporting and remote-operating automated and networked mobility services [39].

In spite of its numerous benefits, the catalogue of scenarios comes with certain limitations. Particularly, it must be noted that the catalogue does not claim to be exhaustive in several regards. First, not all the interactions between the actors proposed are addressed in the catalogue. Second, not every use case cluster or use case is considered at the same level of depth and detail which affects the balance of scenarios across use case clusters and use cases. The focus of this catalogue is on the projects and contexts that have been presented initially in this paper. Filling the gaps of the presented framework and coming up with more use case clusters, use cases, and scenarios is a quest for further research and inevitably depends on the context of use, the vehicles investigated, and their level of automation, as well as on the mode of remote operation and its concrete conceptualization. Also, categorizing the scenarios' consequences, e.g., by severity or impact, is left to future empirical investigations. Appropriate categorization systematics may be part of prospective iterations of this scenario catalogue, which is considered a living document that is permanently updated as research progresses. The same is true for deriving required measures for each scenario. Third, the categories used for analyzing the scenarios might not be mutually exclusive in certain cases. For instance, the listed consequences of some of the scenarios may already contain required measures even though deriving measures will be

the subsequent step in the user-centered design process. This is because for comprehending these scenarios, indicating the required measures is inherently necessary. Moreover, only if the scenario is understood correctly, adequate design solutions can be derived from it. Further research is needed to evaluate and modify the categories used in this framework, if necessary, as well as to disentangle remaining unclarities in the categories assigned.

At any rate, the remote operation of vehicles is likely to become a vital element of highly automated driving systems. Analyzing typical scenarios that may occur when remotely operating highly automated vehicles is a first but essential step towards enabling remote operation while designing with human needs at the center of attention. While the presented scenarios focus on the notion of controlling one vehicle at a time, feasible remote-operation solutions may need to be able to supervise several vehicles simultaneously. Also, communication of the remote operator with other road users is a research area that has not been investigated yet to the knowledge of the authors. Relying on external HMI solutions for highly automated vehicles (e.g., [40,41]) may be a feasible approach.

All in all, the presented catalogue of scenarios is an important milestone for bringing highly automated vehicles onto the roads as it is a precursor for testing and validating them under realistic conditions. By being able to tackle the scenarios presented, the remote operation will significantly improve the operation of AVs and therefore bring us a small but vital step closer to fully automated mobility. And even in the case that fully automated driving (SAE Level 5) may be achieved one day, certain scenarios in fully automated public transport, such as passenger emergencies or vehicle malfunctions, may always require human support. Hence, the remote operation may be more than a preliminary technology to bridge the gap to a certain level of automated driving. It may be a long-lasting alternative to human operators on-site without compromising on the unique and sometimes irreplaceable abilities and skills of humans.

Author Contributions: Conceptualization, C.K., A.S. and M.O.; methodology, C.K., A.S., H.A. and M.O.; formal analysis; C.K., A.S. and H.A.; investigation, C.K. and A.S.; data curation, A.S., C.K. and H.A.; writing—original draft preparation, A.S. and C.K.; writing—review and editing, C.K., A.S. and M.O.; visualization, A.S. and C.K.; supervision, M.O.; project administration, C.K.; funding acquisition, M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry for Digital and Transport via the project “RealLab Hamburg Digital Mobility” which goes back to the German National Platform for the Future of Mobility (NPM).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be made available on request to the authors.

Acknowledgments: We thank the public transport organizations “Braunschweiger Verkehrs-GmbH” in Braunschweig, Germany, “Hamburger Hochbahn AG” and “Verkehrsbetriebe Hamburg-Holstein GmbH” as well as “Hamburg University of Technology” in Hamburg, Germany for the kind collaboration.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Camacho, F.; Cárdenas, C.; Muñoz, D. Emerging technologies and research challenges for intelligent transportation systems: 5G, HetNets, and SDN. *Int. J. Interact. Des. Manuf.* **2018**, *12*, 327–335. [[CrossRef](#)]
2. Ullah, H.; Gopalakrishnan Nair, N.; Moore, A.; Nugent, C.; Muschamp, P.; Cuevas, M. 5G Communication: An Overview of Vehicle-to-Everything, Drones, and Healthcare Use-Cases. *IEEE Access* **2019**, *7*, 37251–37268. [[CrossRef](#)]
3. Van Mierlo, J.; Bercibar, M.; El Baghdadi, M.; De Cauwer, C.; Messagie, M.; Coosemans, T.; Jacobs, V.A.; Hegazy, O. Beyond the State of the Art of Electric Vehicles: A Fact-Based Paper of the Current and Prospective Electric Vehicle Technologies. *World Electr. Veh. J.* **2021**, *12*, 20. [[CrossRef](#)]
4. Meyer, G.; Beiker, S. (Eds.) *Road Vehicle Automation*; Springer International Publishing: Cham, Switzerland, 2014; ISBN 978-3-319-05989-1.

5. Arena, F.; Pau, G.; Severino, A. An Overview on the Current Status and Future Perspectives of Smart Cars. *Infrastructures* **2020**, *5*, 53. [CrossRef]
6. Pavone, M. Autonomous Mobility. In *Autonomes Fahren*; Maurer, M., Gerdes, J.C., Lenz, B., Winner, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 399–416, ISBN 978-3-662-45853-2.
7. Machado, C.; de Salles Hue, N.; Berssaneti, F.; Quintanilha, J. An Overview of Shared Mobility. *Sustainability* **2018**, *10*, 4342. [CrossRef]
8. Society of Automotive Engineers. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for on-Road Motor Vehicles*; SAE: Washington, DC, USA, 2021. Available online: https://www.sae.org/standards/content/j3016_202104 (accessed on 2 July 2021).
9. Shi, E.; Frey, A.T. Theoretical Substitution Model for Teleoperation. In *Automatisiertes Fahren 2021*; Bertram, T., Ed.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2021; pp. 69–81, ISBN 978-3-658-34753-6.
10. Sag, A. The State of 5G in Early 2022. *Forbes* [Online]. Available online: <https://www.forbes.com/sites/moorinsights/2022/01/18/the-state-of-5g-in-early-2022/?sh=1247e9125652> (accessed on 20 January 2022).
11. Gesetz zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes—Gesetz zum Autonomen Fahren. 2021. Available online: <https://www.bundesrat.de/SharedDocs/beratungsvorgaenge/2021/0401-0500/0430-21.html> (accessed on 17 March 2022).
12. Rosenzweig, A. Regulators Know Teleoperation Is Key for Self-Driving Vehicles to Succeed. Available online: <https://venturebeat.com/2021/06/17/regulators-know-teleoperation-is-a-must-have-for-self-driving-vehicles-to-succeed/> (accessed on 20 January 2022).
13. The Knowledge Base on Connected and Automated Driving. France Takes Lead on Allowing Automated Driving on Public Roads. Available online: <https://www.connectedautomateddriving.eu/blog/france-takes-lead-on-allowing-automated-driving-on-public-roads/> (accessed on 20 January 2022).
14. National Conference of State Legislatures. Autonomous Vehicles | Self-Driving Vehicles Enacted Legislation. Available online: <https://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx> (accessed on 20 January 2022).
15. Weimer, J. *U-Shift—A Novel on-the-Road Modular Vehicle Concept*; German Aerospace Center: Stuttgart, Germany, 2019.
16. Hamburger Hochbahn. The Future Is Driverless: Be Part of the Hochbahn Research and Development Project Heat. Available online: https://www.hochbahn.de/hochbahn/hamburg/en/home/projects/expansion_and_projects/project_heat (accessed on 29 June 2021).
17. RealLab Hamburg. Autonomes Fahren. Available online: <https://reallab-hamburg.de/projekte/autonomes-fahren/> (accessed on 25 March 2021).
18. Freie Universität Berlin. KIS-M—KI-basiertes System für Vernetzte Mobilität. Available online: <https://www.mi.fu-berlin.de/inf/groups/ag-ki/Projects/KIS-M/index.html> (accessed on 21 January 2022).
19. DLR. U-Shift: Information on the Vehicle Concept. Available online: <https://verkehrsforschung.dlr.de/en/projects/u-shift> (accessed on 17 March 2022).
20. Kettwich, C.; Dreßler, A. Requirements of Future Control Centers in Public Transport. In Proceedings of the AutomotiveUI '20: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Washington, DC, USA, 21–22 September 2020; ACM: New York, NY, USA, 2020; pp. 69–73.
21. Kettwich, C.; Schrank, A.; Oehl, M. Teleoperation of Highly Automated Vehicles in Public Transport: User-Centered Design of a Human-Machine Interface for Remote-Operation and Its Expert Usability Evaluation. *MTI* **2021**, *5*, 26. [CrossRef]
22. Kettwich, C.; Schrank, A.; Oehl, M. Supervising the Automation: Tasks and Human-Machine Interfaces for on-Board Operators of Highly Automated Shuttles in Real-World Laboratories across Germany. In Proceedings of the Annual Meeting of HFES Europe Chapter, Torino, Italy, 20–22 April 2022.
23. International Organization for Standardization. *ISO Standard No. 9421-210:2019. Ergonomics of Human-System Interaction—Part 210: Human-Centered Design for Interactive Systems*; International Organization for Standardization: Geneva, Switzerland, 2019.
24. The Sage Encyclopedia of Qualitative Research Methods. Given, L., Given, L.M., Eds.; Sage: Los Angeles, CA, USA, 2008; ISBN 9781412941631.
25. Dreßler, A. User-needs-based design of public transport with autonomous vehicles: User-centered research in the project HEAT. In Proceedings of the 2021 SIP-Adus Workshop, Human Factors Breakout Session, Online, 29 October 2021.
26. Ulbrich, S.; Menzel, T.; Reschka, A.; Schuldt, F.; Maurer, M. Defining and Substantiating the Terms Scene, Situation and Scenario for Automated Driving. In Proceedings of the IEEE International Annual Conference on Intelligent Transportation Systems, Las Palmas, Spain, 15–18 September 2015; pp. 982–988.
27. Wilbrink, M.; Schieben, A.; Markowski, R.; Weber, F.; Gehb, T.; Ruenz, J.; Tango, F.; Kaup, M.; Jan-Henning, W.; Portouli, V.; et al. interACT D1.1. Definition of interACT Use Cases and Scenarios. 2018. Available online: https://www.interact-roadautomation.eu/wp-content/uploads/interACT_WP1_D1.1_UseCases_Scenarios_1.1_approved_UploadWebsite.pdf (accessed on 19 January 2022).
28. Geis, T.; Tesch, G. *Basiswissen Usability und User Experience: Aus- und Weiterbildung zum UXQB® Certified Professional for Usability and User Experience (CPUX)—Foundation Level (CPUX-F)*; dpunkt.verlag: Heidelberg, Germany, 2019; ISBN 9783960886297.

29. Dzida, W.; Freitag, R. Making use of scenarios for validating analysis and design. *IEEE Trans. Softw. Eng.* **1998**, *24*, 1182–1196. [CrossRef]
30. DLR. Anwendungsplattform Intelligente Mobilität. Available online: <https://www.dlr.de/ts/aim#gallery/25291> (accessed on 22 March 2022).
31. German Aerospace Center. AIM vehicle Fleet. Available online: <https://www.dlr.de/content/en/research-facilities/aim-vehicle-fleet.html> (accessed on 11 February 2022).
32. Schrank, A.; Kettwich, C.; Heß, S.; Oehl, M. Highly automated yet highly controlled: A case study of HAVs' on-board operators' workplaces across three real-world laboratories. In Proceedings of the 64th Conference of Experimental Psychologists (TeaP), Online, 20–23 March 2022.
33. Eclipse Foundation. Automated Driving Open Research (ADORE). Available online: <https://projects.eclipse.org/proposals/eclipse-automated-driving-open-research-adore> (accessed on 22 February 2022).
34. Lapoehn, S.; Heß, D.; Böker, C.; Böhme, H.; Schindler, J. Infrastructure-aided Automated Driving in Highly Dynamic Urban Environments. In Proceedings of the ITS World Congress 2021, Hamburg, Germany, 11–15 September 2021.
35. TransAID. Transition Areas for Infrastructure-Assisted Driving. Available online: <https://www.transaid.eu/> (accessed on 22 February 2022).
36. Weimer, J.; Ulrich, C.; Conzelmann, M.; Fleck, T.; Zofka, M.R.; Grünhäuser, M. Managed automated driving: A new way for safe and economic automation. In Proceedings of the 27th ITS World Congress, Hamburg, Germany, 11–15 October 2022.
37. Hi-Drive. Hi-Drive: Addressing Challenges towards the Deployment of Higher Automation. Available online: <https://www.hi-drive.eu/> (accessed on 11 February 2022).
38. Verband deutscher Verkehrsunternehmen. Stellungnahme zum Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes. Available online: <https://www.bundestag.de/resource/blob/838594/305010163d3d1daf817adbdec8483507/19-15-489-C-data.pdf> (accessed on 11 February 2022).
39. Lenz, K. *Data-Centric Solutions for Future Mobility: GAIA-X 4 Future Mobility—Five Projects Presented*; DLR: Cologne, Germany, 2022; Available online: <https://fgvt.htwsaar.de/site/en/gaia-x-4ams-2021-2024/> (accessed on 17 March 2022).
40. Schieben, A.; Wilbrink, M.; Kettwich, C.; Dodiya, J.; Oehl, M.; Weber, F.; Sorokin, L.; Lee, Y.M.; Madigan, R.; Merat, N.; et al. *Testing External HMI Designs for Automated Vehicles—An Overview on User Study Results from the EU Project interACT*; Tagung Automatisiertes Fahren: Munich, Germany, 2019.
41. Lau, M.; Le, D.H.; Oehl, M. Design of External Human-Machine Interfaces for Different Vehicle Types. In *Contributions to the 63rd Tagung Experimentell arbeitender Psychologen*; Huckauf, A., Baumann, M., Ernst, M., Herbert, C., Kiefer, M., Sauter, M., Eds.; Pabst Science: Lengerich, Germany, 2021.

Article

The Influence of Sentiments of Economic Agents on Pedestrians and Vehicle Crossings along the US–Mexico Border

René Cabral, Francisco García-Flores * and Eduardo Saucedo

EGADE Business School, Tecnológico de Monterrey, Rufino Tamayo and Eugenio Garza Lagüera, San Pedro Garza García 66269, Mexico; rcabral@tec.mx (R.C.); eduardo.saucedo@tec.mx (E.S.)

* Correspondence: f.garciaf@tec.mx

Abstract: This study aimed to investigate the impact of people's sentiments toward border crossings on personal vehicle and pedestrian crossings along the US–Mexico border. This study focused on regional factors and employed data derived from Google Trends as a proxy for people's sentiments. Monthly data from the first quarter of 2004 to February 2020 were used. Different regression models were used to address stationarity. After controlling for economic conditions and external events, the primary findings are as follows: first, pedestrian and personal vehicle crossings are sensitive to exchange rate fluctuations. Second, the economic cycle has a slightly higher impact on pedestrians than personal vehicle crossings. Third, an increase in the hostile environment toward immigration in the U.S. may negatively impact pedestrian crossings, especially in Texas. Moreover, a rolling regression was used to examine the impact of people's sentiments on crossings over time.

Keywords: border crossings; sentiments; personal vehicles; pedestrians; US–Mexico; Google Trends

Citation: Cabral, R.; García-Flores, F.; Saucedo, E. The Influence of Sentiments of Economic Agents on Pedestrians and Vehicle Crossings along the US–Mexico Border. *Appl. Sci.* **2022**, *12*, 2512. <https://doi.org/10.3390/app12052512>

Academic Editors: Anselme Muzirafuti, Dimitrios S. Paraforos, Giovanni Randazzo and Stefania Lanza

Received: 26 October 2021

Accepted: 10 February 2022

Published: 28 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The NAFTA agreement (now known as USMCA) has boosted economic integration between the United States (the U.S.) and Mexico over the last two and a half decades. From 1994 to 2019, trade in goods between the 2 countries increased 5-fold, from \$100.34 billion to \$614.54 billion Data from www.census.gov (accessed on 26 September 2020). Trade growth in Mexico was even more dramatic, with exports of goods increasing from \$49.49 billion in 1994 to \$357.97 billion in 2019. Labastida-Tovar [1] documented a significant growth in export levels in both countries' border cities. Indeed, more impoverished border cities, such as Port Arthur, California, and Reynosa, Mexico, have gained more from trade, with higher growth rates than San Diego or Monterrey. From 1996 to 2019, pedestrian crossings increased by 44.2%, from 34.10 million to 49.18 million, but the growth rate of personal vehicle crossings was 17.1% Data from Bureau of Transportation Statistics.

Labastida-Tovar [1] argued that because of NAFTA, economic integration and trade liberalization have intensified, benefiting both countries. According to Nicita [2], trade liberalization favored northern Mexican states more than southern states. In addition, Hanson [3] demonstrated that a high level of trade liberalization increases the demand for local products in foreign markets, boosting salaries due to the relocation of manufacturing facilities in the border regions.

Figure 1 depicts the total annual northbound border crossings of pedestrians and personal vehicles from 2004 to 2019. Data from 2020 is not considered due to border restrictions as part of the public health measures employed to tackle the COVID-19 pandemic. Pedestrian crossings totaled 48 million in 2004. From 2005 to 2007, an increasing trend was noted, followed by a considerable downturn, which was most likely influenced by the global financial crisis, with a low point in 2010 (39.9 million crossings). In 2019, there were 47.5 million pedestrian crossings, which is roughly the same level as that of 15 years earlier.

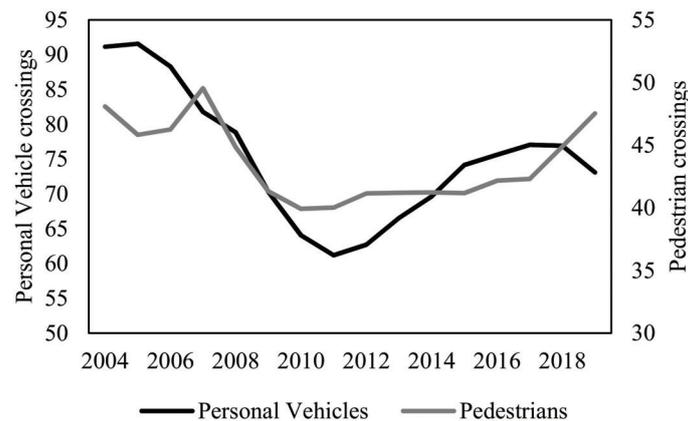


Figure 1. Annual pedestrian and personal vehicle crossings from 2004 to 2019 (millions). Note: Annual data for the 2004–2019 period. Data in millions. Pedestrian crossings are on the right vertical axis. Source: Own estimations using data from the Bureau of Transportation Statistics.

A substantial decline is registered for personal vehicle crossings after 2005, reaching a minimum in 2010 (from 91.5 million crossings in 2005 to 39.9 million in 2010). However, this was followed by a recovery of close to 77 million crossings, which is similar to the 2008 level. Figure 1 depicts periods where pedestrian crossings improved while personal vehicles decreased, suggesting that various factors may influence each type of crossing differently.

Fullerton and Walke [4] claimed that it is not cost effective for pedestrians to shop for specific retail goods categories or travel beyond the immediate border zone. Cross-border shopping requires the price differential of the local and foreign countries to exceed the transaction costs of purchasing across the border. Therefore, as Chandra et al. [5] pointed out, border shoppers live near the border. Baruca and Zolfagharian [6] demonstrated that hedonic shopping motivation may influence economic agents (e.g., consumers) to cross the border to seek fun and pleasure experienced in cross-border shopping trips.

The primary goal of this study was to investigate the impact that people’s sentiments toward crossing the border have on personal vehicle and pedestrian crossings throughout the entire US–Mexico border. This study stems from behavioral economics studies that have examined the relationship between economic agents’ sentiments and economic variables. In those studies, various proxies for assessing investor sentiment were developed, and they can be categorized based on the source of data employed [7]. In this analysis, the use of proxies with data extracted from the volume of internet searches is proposed. The hypothesis to be tested is that policies towards illegal immigration influence the sentiment of economic agents at the aggregate level and thus affect the northbound border crossing of pedestrians and personal vehicles along the US–Mexico border.

This study also builds on prior research examining the determinants of border crossings between Mexico and the U.S. This study makes three significant contributions to the literature. First, data on border crossings were retrieved from all the U.S. Ports of Entry (PoEs) along the US–Mexico border and grouped into three regions. Second, it used sentiment variables built from the volume of internet search queries as a proxy to assess decision-makers’ sentiments. Lastly, a rolling regression analysis was performed to examine the relationship between the sentiment variables and border crossings over time.

The rest of this paper proceeds as follows. The literature review is examined in the next section. Then, in the third section, the variables and data are described. Section 4 presents the method used in the study, and the results are presented in Section 5. The discussion and trends for future research are proposed in Section 6, and the conclusion is presented in the last section.

2. Literature Review

The US–Mexico border region is home to approximately 14 million people concentrated in 14 border-city pairs. Quintana et al. [8] documented that these binational urban areas range from larger metropolitan areas, such as San Diego-Tijuana (2.9 million people), to minor city pairs, such as Nogales in Arizona and Nogales in Sonora (0.2 million). The border is shared by four U.S. states and six Mexican states. The border is 3200 km long and includes 39 Mexican municipalities, 25 U.S. counties, and 25 land ports. In 2019, close to 47.5 million pedestrians crossed northbound, and 73 million personal vehicle crossings were registered.

2.1. Border-Crossing Literature: General Overview

Heyman [9] defined a port of entry (PoE) as nodes in the world trade system where people and goods enter a nation. According to Quintana et al. [8], personal vehicle and pedestrian crossings have increased substantially since NAFTA was signed. Population growth, Mexican border industrialization, and commerce expansion are important factors in explaining the increase in crossings during this period. In addition, Lee and Wilson [10] pointed out that more than 70% of the bilateral trade between the U.S. and Mexico flows through the border's land ports. Tourism is another important driver. According to Lee and Wilson [10], approximately 85% of Mexican tourists arrive in the U.S. through land ports, impacting the economy of border cities in the U.S.

The border-crossing literature has analyzed the relationship between border crossings and the real exchange rate [5,11–14], GDP [13,15], gasoline prices [16], unemployment [4,17], personal income [18,19], trade agreements [17,20], and geopolitical and security issues, such as the 9/11 event [20–23], and the economic impact of foreign visitors in border cities [24,25].

2.2. The US–Mexico Border

Patrick and Renforth [26] found that the devaluation of the Mexican peso affects Texas retailers, showing their reliance on Mexican customers (economic agents). Moreover, the influence of currency rate fluctuations differs by city, distance from the border, retail sector, and domestic market size. Mexican consumption is determined by purchasing power after considering currency conversion, inflation, interest rates, and available disposable income.

Gerber [27] investigated the influence of Mexican tourists on retail sales in eight U.S. border counties by employing the simple compound growth model of retail sales. According to the research, currency fluctuations impact retail sales, with nondurable goods, fashion merchants, and general merchandize outlets being more vulnerable. Fullerton [14] examined the relationship between the exchange rate and the cross-border flows in three international bridges in El Paso. He used monthly data for same-day personal vehicle, passenger, and pedestrian trips; the exchange rate; and the consumer price index. The data revealed each bridge's border-crossing behavior. The findings demonstrate that crossing flows are not random, and peso devaluation has a varying effect on traffic on each bridge. The Mexican peso's depreciation negatively impacts the bridge in that more personal vehicles cross, although it has a positive effect on the bridge where pedestrians prevail.

Fullerton et al. [28] studied how toll tariffs and currency exchange rates influence border crossings using 1990–2006 monthly pedestrian, personal vehicle, and cargo crossing data. The authors found a negative relationship between the toll and flow of border crossings using ARIMA transfer functions. They also found a negative relationship in land ports where pedestrians and personal vehicles predominate and a positive relationship in ports where cargo crossings dominate.

Cabral et al. [13] analyzed the relationship between exchange rate and personal vehicle passenger crossings in the 9 most active PoEs using monthly data from 1997 to 2018. Using panel data fixed effects and augmented mean group models and controlling for the difference in economic growth rates between Mexico and the U.S., they found a negative

relationship between the exchange rate and border crossings and an increase in crossings when the Mexican economy improves faster than the U.S. economy.

2.3. The U.S.–Canada Border

Di Matteo and Di Matteo [19] used quarterly data about same-day vehicle crossings to examine the determinants of cross-border shopping for 7 Canadian provinces that share a border with the U.S. from 1979 to 1992. The authors employed the Gerber [27] demand model and estimated a log–log model using ordinary least squares regression techniques. They found that the main factor for border crossings in several Canadian provinces is the exchange rate. Moreover, the per capita income variable is the most important determinant; however, British Columbia’s gas price is the most relevant variable that explains border crossings.

2.4. Decision-Maker Sentiment

Traditional economic theories leave no room for irrational behavior; they assume that decisions are made through rational decision-making processes [29], and economic agents consider all publicly available information when making decisions [30]. However, empirical evidence reveals that agents are constrained in the amount of information they can process [31], questioning the rationality assumptions in classical theory. Furthermore, behavioral economics has highlighted the relevance of economic agents’ sentiments in financial markets [7], and their behavioral biases, which lead to overly optimistic/pessimistic beliefs and drive an irrational behavior [32].

The literature provides several definitions for the term “sentiment”. In the financial literature, investor sentiment is commonly characterized as either the proclivity to take risks and speculate or the overall sense of optimism or pessimism toward risky assets [33]. For example, De Long et al. [34] related sentiment to noise trading. In contrast, Baker and Wurgler [35] attribute it to feelings of optimism or pessimism about risky assets.

Scholars have devised various proxies for measuring the sentiments of economic agents, especially investors. Based on the data source from which the proxy is extracted, Zhang et al. [7] grouped these proxies into three categories. Indirect proxies are based on stock market data and survey results. Baker and Wurgler [35] used six proxies for sentiment retrieved from the stock market, including the closed-end fund discount and NYSE share turnover. Based on noise-trader sentiment models, Lemmon and Portniaguina [36] employed the University of Michigan survey of consumer sentiment and the Conference Board survey of consumer confidence as proxies for investor sentiment.

Da et al. [37] suggested that earlier indirect sentiment measures do not capture all decision-makers’ sentiments. They claimed that market-based indicators can capture more than investors’ sentiments. Moreover, survey-based proxies are infrequent, and respondents cannot be incentivized to submit truthful answers. Because of this, they stated that the volume of internet search inquiries should be used to create direct sentiment indicators.

Emerging literature has applied internet search volume data as a source to construct proxies of economic and noneconomic variables [38]. For example, Ettredge et al. [39] analyzed the relationship between the search volume of employment-related terms with monthly U.S. unemployment data. Ginsberg et al. [40] used Google search volume to predict the incidence of influenza diseases. Choi and Varian [41] used search volume from Google to estimate macroeconomic variables, including automobile sales, initial claims for unemployment benefits, travel destination planning, and consumer confidence. Guzman [42] employed search query data to examine inflation expectations. Vosen and Schmidt [43] constructed a private consumption measure based on search queries. Finally, Da et al. [37] hold that sentiment at the market level can be measured using search volume data. Thus, using online searches relevant to households in the U.S., the authors developed a Financial and Economic Attitudes Revealed by Search (FEARS) index.

In this study, it is important to emphasize that the primary goal was to explore the relationship between people’s sentiment to cross the border at the aggregate level and

its effect on border crossings. This study proposes that online search query volume data should be used to develop a direct measure, based on the claim of Ettredge et al. [39] that people’s search behavior reveals information about their needs, desires, preferences, and concerns. As Guzman [42] proposed, search behavior can be interpreted as a measure of disclosed expectations. According to Da et al. [37], aggregating search volume data reveals market-level sentiment connected to specific topics. Moreover, Google search volume data is used as to gauge people’s intentions and concerns over crossing the border.

3. Data Analysis

The Bureau of Transportation Statistics provides monthly statistics for inbound crossings between the U.S. and Mexico at the PoE level. The data are classified by port, state, and means of transportation. In this study, the dependent variables were the monthly number of pedestrians and the personal vehicle crossings from Mexico to the U.S.

All PoEs were grouped into three regions: California, Texas, and other (Arizona and New Mexico). Table 1 presents the descriptive statistics of pedestrians and personal vehicles aggregated by region. In 2019, California accounted for 42.6% of the pedestrian crossings (42.9% for personal vehicle); Texas, 41.8% (44.1%); and Arizona and New Mexico, 15.6% (13%).

Table 1. Descriptive statistics of pedestrians and personal vehicles crossing the border from Mexico to the U.S. (2004-2M2020).

Regions	Statistics	Variables	
		Pedestrians	Personal Vehicles
California	Mean	1,416,099	2,478,202
	Std. Dev.	173,819	303,700
	Min.	896,650	1,882,214
	Max.	1,825,929	3,137,415
Texas	Mean	1,504,445	3,016,686
	Std. Dev.	174,942	485,218
	Min.	1,168,377	2,195,680
	Max.	2,105,010	4,069,654
Other	Mean	713,134	767,981
	Std. Dev.	153,993	90,761
	Min.	473,677	541,297
	Max.	1,224,899	937,418
Total	Mean	3,633,678	6,262,869
	Std. Dev.	342,316	800,376
	Min.	3,007,144	4,619,191
	Max.	4,788,991	8,074,001

Note: The data are up to February 2020. Source: Own estimations using data from the Bureau of Transportation Statistics.

Google is the largest and most popular search engine on the internet, and since 2004, it has provided the Google Trends services, in which the historical Search Volume Index (SVI) of search terms can be downloaded [37]. This tool provides the search volume of each term in hourly, daily, weekly, or monthly frequencies. In addition, the SVI data for each keyword within a particular geographical region is scaled from 0 to 100 by the period’s maximum.

To test the effect that an “anti-immigrant environment” has on economic agents’ sentiments (e.g., shopping-trip travelers) and, hence, on their attitudes and decisions toward northward pedestrian and personal vehicle crossings along the US–Mexico border, it is proposed that two variables should be constructed as proxies to capture the phenomenon from both sides of the border. According to Ettredge et al. [39] and Guzman [42], people’s search behavior can be expected to reveal matters relevant to them. These variables are derived from the Google Trends database.

If hedonic shopping drives economic agents to cross the border to pursue fun and enjoyment [6], it is reasonable to expect that an “anti-immigrant environment” in the U.S. border cities will affect economic agents’ attitudes and sentiments toward crossing the

border for shopping trips, thereby influencing border crossings. Therefore, it is proposed that a variable that captures the sentiments of these economic agents at the aggregate level should be developed using the online search volume data as a proxy for their concerns about and interests in the Mexican side of the border. Based on the online activity of economic agents in the U.S., a second variable was introduced to reflect the “anti-immigrant environment” in the U.S.

For the first variable, different monthly SVI data of border-crossing-related terms, including terms associated with border crossings and migration (e.g., “tiempo de cruce,” “puente internacional,” and “cruce fronterizo”), were individually retrieved using the Google Trends tool; the search was restricted to Mexico and the period analyzed. It was followed by a “snowball technique”, which was implemented by including the related terms that Google Trends suggests in the analysis. After analyzing the stationarity and correlation of the time series of the SVI of each term with the border-crossings data, the more significant search terms were identified, which are “migrantes”, “deportaciones”, “muro fronterizo”, “border patrol”, and “patrulla fronteriza”. Lastly, the SVI of these terms is grouped by adding each month’s values to construct a single index named Border Economic Migrant Sentiment (BEMS). Because the BEMS variable comprises search terms related to security and “anti-immigration” issues, it is expected that it will capture people’s concerns about crossing the border [39,42].

The SVI of terms related to border crossings is also explored, with the geographic scope limited to the U.S. A similar research strategy is employed for the other variables, with the “immigration” search term being the most relevant. It is assumed that when an anti-immigration sentiment occurs in the U.S., the SVI of the “immigration” search term will increase. Therefore, adding the “immigration” variable (Imm) may help determine whether a rise in the hostile environment in the U.S. toward immigration may influence northbound border crossings.

Table 2 displays statistics for the main variables. The monthly average crossings for pedestrian and personal vehicles are 3.6 and 6.2 million, respectively. Figure 2 depicts the northbound border crossings for pedestrians and personal vehicles grouped by region. The annual pedestrian crossings are depicted in Figure 2a. The other region experienced a rising trend from 2004 to 2007, followed by a significant decline, which was most likely caused by the global financial crisis, reaching a minimum in 2014. Subsequently, the border crossings experienced a sluggish recovery, with nearly 7 million crossings flattening in recent years. An annual decrease in pedestrian crossing is recorded from 2004 to 2009 in the Californian PoEs, with a temporary recovery in 2007. Texan ports reported similar behavior, with border crossings declining from 2004 to 2011 and with a temporary increase in 2007.

Table 2. Descriptive statistics of the main variables (2004-2M2020).

Statistics	Variables						
	Pedestrians	Personal Vehicles	BEMS	Imm	Real Exch. Rate	IGAE	Real Gas Price
Mean	3,633,678	6,262,869	35.61	36.05	81.79	98.19	1.26
Std. Dev	342,316	800,376	27.9	13.2	7.00	10.05	0.25
Min.	3,007,144	4,619,191	9	18	68.19	79.19	0.79
Max.	4,788,991	8,074,001	167	100	101.05	117.83	1.88

Note: The monthly data are up to February 2020. Source: Own estimations from sources described in the data section.

From 2009 to 2011, although pedestrian crossings in Californian ports increased, those in Texas decreased. However, from 2012 to 2015, there was a decrease in pedestrian crossings in California ports. The reverse is observed in Texas, where crossings in 2019 were lower than those registered 15 years earlier. This could indicate that regional factors may influence pedestrian border crossings.

Figure 2b exhibits the personal vehicle border crossings for each region. From 2004 to 2011, all 3 regions experienced a similar declining trend. However, a recovery was

registered from 2011 to 2017 in all regions, followed by a downward pattern in the last years of the sample.

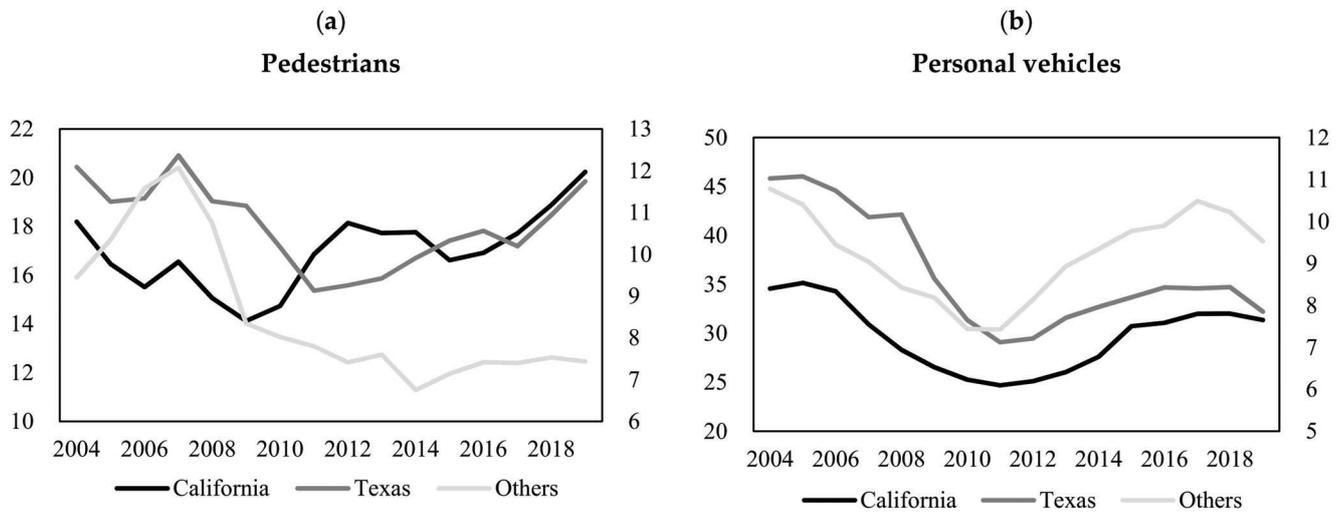


Figure 2. Northbound PoE's annual pedestrian and personal vehicle crossings by regions, from 2004 to 2019 (million). Note: The annual data are from 2004 to 2019 and are in millions. In Figures (a) and (b), the crossings for other are on the right vertical axis. Source: Own estimations using data from the Bureau of Transportation Statistics.

Following the literature [13,14,26–28], the real exchange rate movements are expected to explain the evolution of border crossings. For example, a depreciation of the Mexican peso makes shopping and leisure activities more costly, reducing the number of pedestrian and personal vehicles crossings. Likewise, two variables are introduced to capture economic factors. First, the Global Indicator of Economic Activity (IGAE) variable is used to track the Mexican economic cycle. Second, the real U.S. gas price is used to measure its influence in border-crossing fluctuations. The gas price difference between the U.S. and Mexico is a relevant variable for passenger vehicles crossings, especially across the Texas border. Moreover, the difference tends to be positive, influencing the crossing of Mexican citizens to the U.S. to look for cheaper gas, when such price differences are significant.

The data on the exchange rate between the Mexican peso and the U.S. dollar are retrieved from Banco de Mexico (Mexican Central Bank), which publishes a monthly real exchange rate index based on a weighted basket of numerous currencies. The gas price is retrieved from the U.S. Energy Information Administration. In this study, the monthly retail gasoline prices of all grades of formulations in the U.S. are used. The data are adjusted in real terms.

The monthly IGAE indicator and the real exchange rate from 2004 to February 2020 are displayed in Figure 3. The Mexican economy experienced an upper trend during this period, which was caused by trade liberalization and other economic reforms. However, the figure reveals that during the 2008 global financial crisis, the Mexican economy suffered a slowdown pattern, with the Mexican peso depreciating significantly. As a result, from 2016 onwards, the economy and the real exchange rate experienced a sideways trend, reducing the economic growth rate as compared with that of the previous years.

A dummy variable is included in the econometric model to capture the exogenous events that impacted the U.S. and may influence border crossings. The information about the U.S. economic recessions is retrieved from the National Bureau of Economic Research (NBER). Regarding the D_US_CRISIS variable, from December 2007 to June 2009 (including the 2008 global financial crisis) is assigned a value of 1.

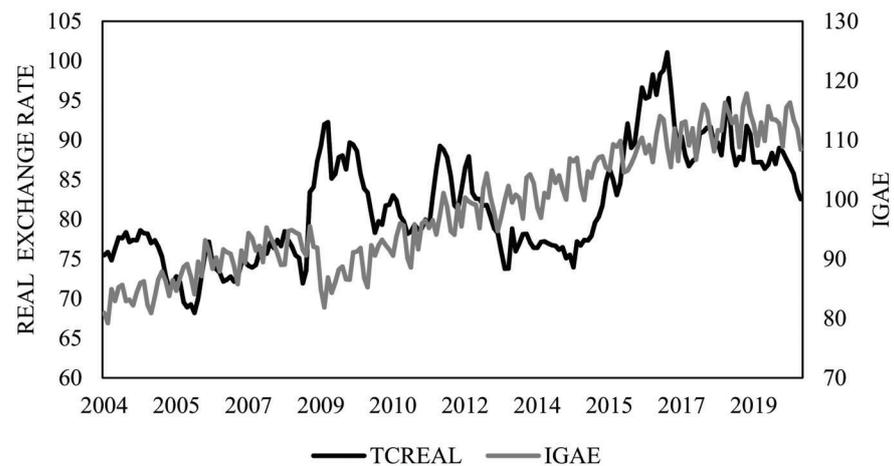


Figure 3. IGAE and Real Exchange Rate from 2004 to 2020. Note: The monthly data are from 2004 to February 2020. IGAE stands for the Global Indicator of Economic Activity and is displayed on the right vertical axis. Source: Own estimations using data from INEGI and Banco de Mexico.

Table 3 presents the correlation matrix of the dependent and the main independent variables. The real exchange rate’s negative sign regarding pedestrians and personal vehicles is as expected. Hence, there is some initial evidence that northbound border crossings decline when the Mexican peso depreciates. In the case of pedestrians and personal vehicles, the IGAE coefficient is negative, although the correlations are small. Furthermore, as projected, the real gas price is negative for personal vehicles. Lastly, the correlation between the two Google Trend variables is 0.408.

Table 3. Correlation matrix.

	Pedestrians	Personal Vehicles	BEMS	Imm	Real Exch. Rate	IGAE	Real Gas Price
Pedestrians	1						
Personal Vehicles	0.65	1					
BEMS	0.10	0.24	1				
Imm	0.26	0.50	0.41	1			
Real Exch. Rate	−0.21	−0.32	−0.06	−0.47	1		
IGAE	−0.06	−0.28	−0.06	−0.55	0.60	1	
Real Gas Price	−0.22	−0.40	−0.16	0.02	−0.44	−0.21	1

Source: Own estimations.

4. The Empirical Model

The empirical model proposed in this study, which uses a log–log specification, closely follows the study of Di Matteo and Di Matteo [19]:

$$\Delta \ln(C_t) = \alpha_i + \beta_1 \Delta \ln(Exch\ rate)_t + \beta_2 \Delta \ln(GR)_t + \beta_3 \Delta \ln(GAS)_t + \sum_{i=1}^n \theta_i \Delta \ln(GT_i)_t + \gamma_1 (D_US_CRISIS)_t + e_i \tag{1}$$

where the dependent variable C_t denotes the t month total number of pedestrian or personal vehicle crossings. In Equation (1), $Exch\ rate$ represents the real exchange rate variable, GR denotes the Mexican economy growth rate, GAS is the real gas price in the U.S., D_US_CRISIS is the dummy variable that captures economic turmoil, and GT_i represents the variables constructed from Google Trends that capture decision-maker’ sentiments (BEMS and Imm variables). The time series was computed in logarithms for both the dependent and independent variables, except the dummy variable.

Two sets of models were estimated: the first was for the entire sample of PoEs, and the second set was divided into three regions (California, Texas, and other). Endogeneity is

not expected to be a concern because exchange rates are defined in international financial markets. Therefore, border crossings from Mexico are not expected to directly affect the value of the Mexican peso. Furthermore, these border crossings can have little effect on national gas prices in the U.S. or the evolution of Mexico's national economy.

The log–log specification allowed us to identify the elasticities between the pedestrian and personal vehicle crossings and the independent variables. Due to the wealth effect resulting from the currency movements, it was expected that a depreciation of the local currency would negatively affect Mexicans' propensity to cross northbound. Thus, it was anticipated that the sign of the coefficient estimated would be $\beta_1 < 0$.

When the Mexican economy expands, its residents have more resources to cross northbound to take advantage of price differences in goods and services on both sides of the border [26]. Indeed, border crossings are explained by Mexican residents' shopping trips [26,44]. Mexicans have greater resources for cross-border shopping trips near the top of the economic cycle. Thus, the coefficient of this variable was predicted to be $\beta_2 > 0$.

For decades, the government set gas prices in Mexico, with price differentials expected along the border [4,16,19]. Thus, it can be assumed that the U.S. gas price may play a role in the decision to cross the northbound border. If the U.S. gas price increases, Mexicans are less motivated to cross the border. The sign for this coefficient was expected to be $\beta_3 < 0$.

Based on previous studies that used people's online search activity as a proxy for their sentiment [33,37,45,46], it can be argued that people's sentiment may influence border-crossing fluctuations. An increase in the negative sentiment about Mexicans and immigration in the U.S. will result in fewer border crossings. If the variables based on Google Trends data captured the negative sentiments about border crossings, a negative relationship was anticipated ($\theta_i < 0$).

5. Results

We begin by revisiting the series' stationarity before estimating the empirical model presented in Section 4 of this study. The primary estimations are then reported by border-crossing type and region, followed by robustness checks.

5.1. Testing for Stationarity

Table 4 reports the stationarity tests of the main dependent and independent variables in logarithms for both levels and the first differences. The Augmented Dicky Fuller (ADF) unit root tests were conducted, and the results are recorded in the first column of Table 4. The ADF tests the null hypothesis that a time series has a unit root against the alternative hypothesis that it is $I(0)$. The level variables are presented in the first panel, revealing that the data are not stationary. The Phillips–Perron (P.P.) unit root test was also performed, and the results are reported in the second column of Table 4. For this unit root test, the null and alternative hypotheses were the same as those of the ADF test. The results indicate that the data in levels are not stationary. Therefore, the data were computed in its first difference to minimize the nonstationary problems. The data in the first differences are stationary and reported in the second column of Table 4. Stationarity tests for the individual time-series components of the border-crossing sentiment index were performed, which are stationary at the first difference (the results are not presented in Table 4 but are available on request).

5.2. Main Estimates

The results of the regressions of Equation (1) for personal vehicle and pedestrian crossings are presented in Table 5. The Newey–West method was used to compute the robust standard errors to account for autocorrelation problems. The results support the negative relationship between border crossings and real exchange rates. The coefficients range from -0.19 to -0.48 (at least with a significance level of 10%) for pedestrian and personal vehicle crossings. Therefore, the depreciation of the Mexican currency negatively influences the propensity to cross northbound. An increase of 1% in the exchange rate leads

to approximately a -0.20% decrease in personal vehicle crossings, and that of pedestrians is close to -0.40% .

Table 4. Unit root tests.

Variables	ADF	PP
Ln levels		
Pedestrians	0.102 (0.714)	-0.459 (0.515)
Personal Vehicles	-0.678 (0.422)	-0.767 (0.383)
BEMS	-0.407 (0.536)	-0.435 (0.523)
Imm	-0.798 (0.369)	-1.150 (0.228)
Real Exch. Rate	0.224 (0.750)	0.221 (0.749)
IGAE	1.392 (0.959)	2.259 (0.994)
Gas Price (Real terms)	-1.910 (0.054)	-1.921 (0.053)
Ln 1st Differences		
Pedestrians	-4.367 (0.000)	-29.083 (0.000)
Personal Vehicles	-2.590 (0.001)	-35.761 (0.000)
BEMS	-13.665 (0.000)	-45.096 (0.000)
Imm	-13.069 (0.000)	-36.152 (0.000)
Real Exch. Rate	-11.945 (0.000)	-11.813 (0.000)
IGAE	-1.978 (0.046)	-24.749 (0.000)
Gas Price (Real terms)	-9.122 (0.000)	-7.361 (0.000)

Note: *p*-values are reported in parentheses. ADF refers to the Augmented Dicky Fuller unit root test, and PP refers to the Phillips–Perron unit root test. Source: Own estimations.

Table 5. Personal vehicles and pedestrians OLS estimations (period: 2004–2020M02).

Variables	Personal Vehicles		Pedestrians			
	1	2	2004M01	2020M02	5	6
d(ln Real Exch. Rate)	-0.278 *** (0.097)	-0.200 * (0.104)	-0.193 * (0.099)	-0.484 *** (0.172)	-0.383 *** (0.147)	-0.368 ** (0.150)
d(ln IGAE)	1.096 *** (0.103)	1.070 *** (0.097)	1.074 *** (0.095)	1.391 *** (0.145)	1.362 *** (0.141)	1.364 *** (0.141)
d(ln Real Gas Price)	-0.040 (0.031)	-0.001 (0.035)	-0.016 (0.032)			
Dummy_us_crisis	-0.001 (0.006)	-0.001 (0.007)	-0.002 (0.006)	-0.006 (0.017)	-0.007 (0.016)	-0.008 (0.017)
d(lnBEMS)	-0.019 *** -0.006		-0.014 *** (0.005)	-0.026 *** (0.010)		-0.020 ** (0.010)
d(lnImm)		-0.071 *** (0.014)	-0.061 *** (0.015)		-0.102 *** (0.026)	-0.088 *** (0.028)
C	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.003)	-0.002 (0.003)	-0.002 (0.003)
R2adj	0.413	0.434	0.455	0.311	0.332	0.351
AIC	-3.365	-3.401	-3.435	-2.401	-2.433	-2.456
DW	2.918	2.903	2.876	2.644	2.626	2.590

Note: The standard errors reported are robust to heteroscedasticity. *, **, and *** denote significance levels of 10%, 5%, and 1%, respectively.

The coefficient of the IGAE variable is positive (1% significance level), as expected. Thus, Mexicans tend to cross northbound more frequently when they have more resources. With an increase of 1% in the IGAE indicator, increases closes to 1% in personal vehicle crossings and 1.4% in pedestrian crossings are expected. Thus, the elasticities for IGAE are somewhat higher for pedestrian crossings than personal vehicles. These differences are statistically significant (1% significance level). To conduct this analysis, the Stata’s *suest*-based test of equality of coefficients was utilized, by comparing the *d(lnIGAE)* coefficients of model 1 to model 4, model 2 to model 5, and model 3 to model 6 (Table 5). Moreover,

there is no evidence that the price of gasoline in the U.S. is a motive for Mexican residents to cross the border.

In Table 5, the BEMS index variable is added to models 1 and 4, the immigration variable is added to models 2 and 5, and both variables are included in models 3 and 6. The coefficient of both variables is negative, as predicted. An increase of 1% in the BEMS index leads to a 0.02% decrease in border crossings. Moreover, a 1% increase in the Imm leads to an approximately 0.07% decrease in northbound border crossings. By comparing the $d(\ln \text{Imm})$ coefficients calculated using models 2 and 5 in Table 5, which is significant at the 10% level, it is found that Imm is slightly larger for pedestrian crossings than for personal vehicles. These differences are statistically significant (5% significance level), using the Stata’s *suest*-based test of equality of coefficients. After examining models 3 and 6, we find that the coefficients of Imm are larger than the coefficients of the BEMS index for both pedestrian and personal vehicle crossings. These differences are statistically significant (5% significance level), using the Stata’s *suest*-based test of equality of coefficients.

Table 6 (personal vehicle crossings) and Table 7 (pedestrian crossings) report the results for the 3 regions in which the PoEs are grouped. The BEMS index variable is included in models 1, 4, and 7 in Tables 6 and 7; the Imm variable in models 2, 5, and 8; and both variables in models 3, 6, and 9. By analyzing both tables, some interesting insights are obtained. First, in Texas, when the Imm is added, the relationship between personal vehicle crossings and the real exchange rate is not statistically significant, and for pedestrian models, the significance reduces. Moreover, for personal vehicle models in the other two regions, the significance reduces. However, the statistical significance level of the real exchange rate does not change when Imm is included in California and other region’s pedestrian crossings models.

If Imm captures some of the possible hostile immigration-related environment, an anti-immigrant environment weakens the economic motivation for crossing, especially in Texas PoEs. However, additional research may be conducted to examine this argument, as previous findings suggest a possible relationship between the real exchange rate, the anti-immigrant environment, and border crossings flow. However, such a relationship should be viewed with caution, as we provide no strong evidence to support such findings.

If tourism and shopping are two of the main reasons for Mexicans to cross the border [4,26], an increase in an anti-immigrant environment may diminish the economy of border cities.

Table 6. OLS regressions results of personal vehicle crossings clustered by regions (2004–2020M02).

Variables	California			Texas		Other			
	1	2	3	4	2004M01 2020M02 5	6	7	8	9
$d(\ln \text{ Real Exch. Rate})$	−0.271 ** (0.118)	−0.213 * (0.117)	−0.207 * (0.114)	−0.286 ** (0.111)	−0.185 (0.120)	−0.176 (0.118)	−0.239 ** (0.094)	−0.189 * (0.109)	−0.178 * (0.097)
$d(\ln \text{ IGAE})$	1.259 *** (0.121)	1.240 *** (0.115)	1.242 *** (0.112)	0.965 *** (0.122)	0.932 *** (0.119)	0.936 *** (0.118)	1.078 *** (0.106)	1.057 *** (0.105)	1.062 *** (0.102)
$d(\ln \text{ Real Gas Price})$	−0.018 (0.036)	0.011 (0.037)	0.000 (0.035)	−0.059 (0.045)	−0.012 (0.050)	−0.028 (0.047)	−0.035 (0.049)	0.003 (0.055)	−0.018 (0.051)
<i>dummy_uscrisis</i>	0.005 (0.007)	0.004 (0.007)	0.004 (0.007)	−0.005 (0.008)	−0.005 (0.009)	−0.006 (0.008)	0.000 (0.009)	0.001 (0.010)	−0.001 (0.009)
$d(\ln \text{ BEMS})$	−0.014 ** (0.006)		−0.010 ** (0.005)	−0.021 *** (0.007)		−0.015 *** (0.006)	−0.024 *** (0.007)		−0.020 *** (0.007)
$d(\ln \text{ Imm})$		−0.054 *** (0.013)	−0.046 *** (0.013)		−0.090 *** (0.018)	−0.078 *** (0.019)		−0.058 *** (0.017)	−0.043 ** (0.018)
C	−0.003 (0.002)	−0.003 (0.002)	−0.003 (0.002)	−0.003 (0.002)	−0.003 (0.002)	−0.003 (0.002)	−0.002 (0.002)	−0.003 (0.002)	−0.002 (0.002)
R2adj	0.435	0.445	0.454	0.291	0.331	0.353	0.395	0.368	0.413
AIC	−3.242	−3.261	−3.271	−3.028	−3.087	−3.115	−3.262	−3.218	−3.287
DW	2.828	2.833	2.813	2.839	2.791	2.772	2.940	2.971	2.915

Note: The standard errors reported are robust to heteroscedasticity. *, **, and *** denote 10%, 5%, and 1% significance levels, respectively.

Table 7. Pedestrian crossings OLS regression results clustered by regions (period: 2004–2020M02).

Variables	California			Texas		Other			
	1	2	3	4	5	6	7	8	9
d(ln Real Exch. Rate)	−0.470 ** (0.186)	−0.380 ** (0.172)	−0.375 ** (0.174)	−0.466 ** (0.214)	−0.344 * (0.190)	−0.323 * (0.194)	−0.536 ** (0.244)	−0.454 ** (0.218)	−0.431 ** (0.216)
d(ln IGAE)	1.329 *** (0.149)	1.306 *** (0.148)	1.307 *** (0.148)	1.386 *** (0.185)	1.351 *** (0.185)	1.353 *** (0.184)	1.466 *** (0.211)	1.439 *** (0.203)	1.441 *** (0.205)
dummy_uscrisis	−0.004 (0.024)	−0.005 (0.023)	−0.006 (0.024)	0.001 (0.022)	0.0003 (0.022)	−0.002 (0.022)	−0.023 (0.014)	−0.022 * (0.013)	−0.025 * (0.014)
d(lnBEMS)	−0.012 (0.011)		−0.007 (0.011)	−0.034 *** (0.010)		−0.026 *** (0.009)	−0.034 *** (0.013)		−0.028 ** (0.014)
d(lnImm)		−0.076 *** (0.020)	−0.072 *** (0.024)		−0.127 *** (0.034)	−0.108 *** (0.036)		−0.100 ** (0.040)	−0.079 * (0.044)
C	−0.001 (0.004)	−0.002 (0.004)	−0.002 (0.004)	−0.003 (0.003)	−0.003 (0.003)	−0.003 (0.003)	−0.001 (0.005)	−0.002 (0.005)	−0.001 (0.005)
R2adj	0.224	0.247	0.245	0.235	0.254	0.278	0.224	0.217	0.241
AIC	−2.174	−2.205	−2.197	−1.966	−1.990	−2.018	−1.779	−1.771	−1.796
DW	2.484	2.448	2.442	2.693	2.693	2.648	2.563	2.556	2.539

Note: Standard errors reported are robust to heteroscedasticity. The symbols *, **, and *** refer to 10%, 5%, and 1% significance levels, respectively.

However, in terms of elasticities, a 1% increase in Imm in Texas results in a −0.127% decline in pedestrian crossings, which is slightly greater than the −0.090% elasticity of personal vehicles. These differences are statistically significant (5% significance level), using the Stata’s suest-based test of equality of coefficients. This indicates that pedestrian crossings are more sensitive to changes in Imm.

Second, in all regions, the Mexican economic cycle represented by IGAE is statistically significant at the 1% level. As expected, when Mexicans have more resources, they cross the border more frequently. However, gas price is not statistically significant, suggesting that most people who cross the border may be motivated by shopping, tourism, or leisure activities rather than purchasing gasoline.

Lastly, except for the pedestrian models in California (Table 7), the BEMS index is statistically significant in all models. This might imply that economic motivators may be more important in influencing pedestrians’ border-crossing decisions in California than in the other regions. However, future research may further analyze this line of reasoning. By comparing models 4 in Tables 6 and 7, it is found in Texas that the BEMS index elasticities are slightly higher for pedestrian crossings (−0.03%) than personal vehicles (−0.02%). When both sentiment variables are added, the Imm coefficient has a higher value than the BEMS index coefficient in California and Texas for both types of crossings. Models 3 and 6 in Tables 6 and 7 were examined. These differences are statistically significant (10% significance level in California, 5% significance level in Texas), using the Stata’s suest-based test of equality of coefficients.

These results might suggest that people’s sentiments play a role in their decision to cross the border. Imm is significant in all models and regions, especially in California and Texas. If this proxy captures some of the variations in the anti-immigrant sentiment, the findings might imply that the hostile anti-immigration atmosphere may have a negative impact on border crossings and the economies of U.S. border cities.

5.3. Robustness Checks

Rolling regression analysis was conducted as a robustness check. This statistical method seeks to analyze the relationship between the dependent and independent variables. However, unlike other methods, a specific window size is defined and moved progressively along the sample period. Therefore, a window size of 24 observations was selected in this study generating 174 subsamples, which were used to perform the rolling regression with a step of 1 (Windows of 12 and 48 lengths were tested, yielding similar results that can be provided upon request). Henceforth, the window was two years long and moved progressively from one month to the next. The results can be delivered upon request.

This analysis intended to evaluate the behavior of the coefficients of the BEMS and Imm variables using the following specifications:

$$\Delta \ln(PV_t) = \alpha_i + \beta_1 \Delta \ln(Exch\ rate)_t + \beta_2 \Delta \ln(GR)_t + \beta_3 \Delta \ln(GAS)_t + \sum_{i=1}^n \theta_i \Delta \ln(GT_i)_t + e_i \quad (2)$$

$$\Delta(PD_t) = \alpha_i + \beta_1 \Delta \ln(Exch\ rate)_t + \beta_2 \Delta \ln(GR)_t + \sum_{i=1}^n \theta_i \Delta \ln(GT_i)_t + e_i \quad (3)$$

where the dependent variables PV_t and PD_t denote the t month total number of personal vehicle and pedestrian crossings. In Equations (2) and (3), $Exch\ rate$ represents the real exchange rate and GR denotes the Mexican economy growth rate; in Equation (2) GAS is the U.S. real gas price, and GT_i denotes the BEMS and Imm variables.

Figure 4a depicts the coefficients for the first difference of the BEMS logarithm for personal vehicle crossings. Throughout the period, the BEMS had a negative relationship with personal vehicles. However, short-run episodes with positive coefficients arose, such as the summer of 2012 and the 2013 and 2014 winter seasons. Pedestrian crossings followed a similar pattern (Figure 4c). Although Figure 4a depicts 2 main downward trends—one from 2006 to 2008 and the other from 2014 to 2019—with the latter coinciding with the Donald Trump Administration, a similar pattern was found for pedestrian crossings. The mean of the $d(\ln\ BEMS)$ coefficients for personal vehicle (pedestrian) crossings was -0.031 (-0.041), with a standard deviation of 0.020 (0.027), suggesting that pedestrian crossings are slightly more sensitive to fluctuations in the BEMS index than personal vehicles.

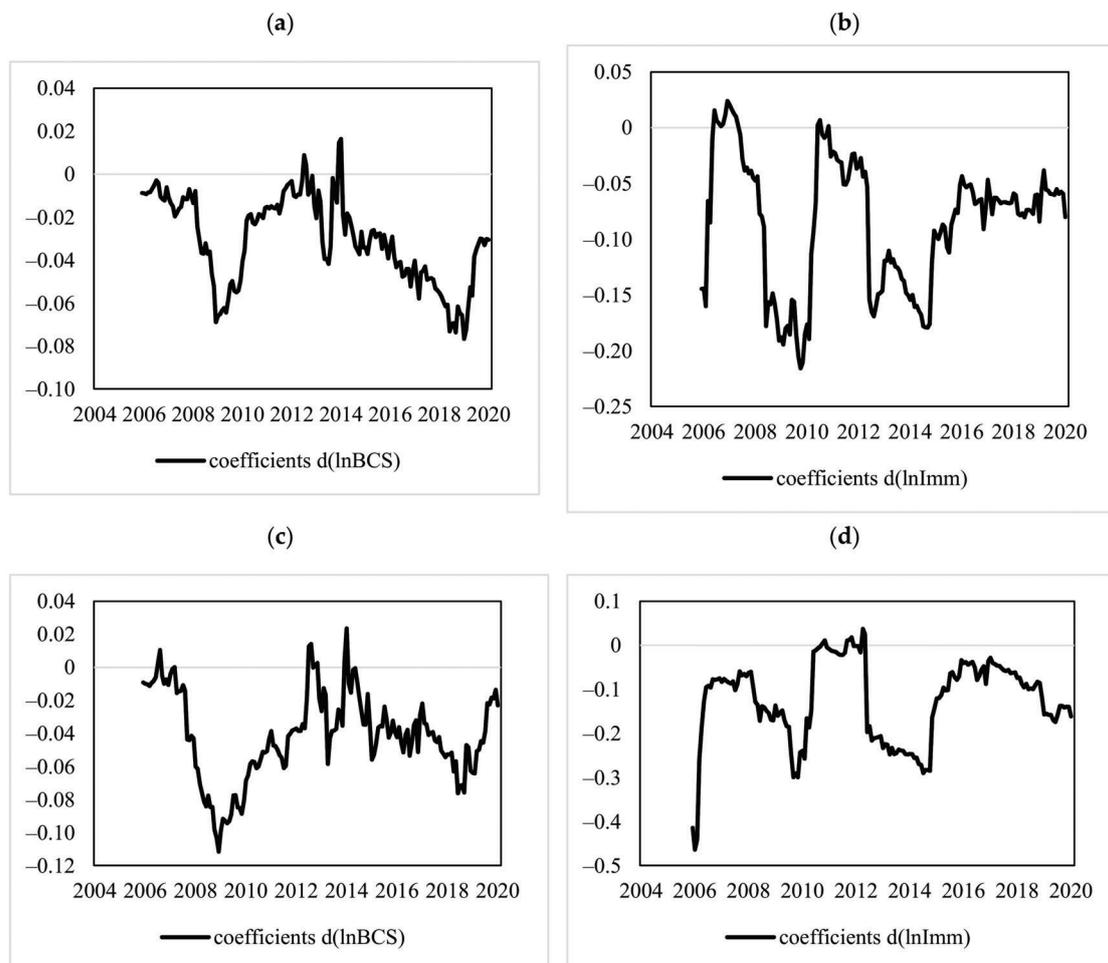


Figure 4. Rolling regression coefficients for the BEMS and Imm variables (24-month window). (a) Dependent variable: LN personal vehicles; (b) Dependent variable: LN personal vehicles; (c) Dependent variable: LN pedestrians; (d) Dependent variable: LN pedestrians.

Figure 4b reveals that the coefficients of the first difference of the Imm logarithm for personal vehicle crossings are negative, except from January 2006 to May 2007 and 2 subperiods in 2010. Two rising trends were followed by major decreases in 2006 and 2010. In addition, a low point is observed in 2014, coinciding with the 2014 immigration crisis. Due to violence and persecution, an increasing number of Central American families seek asylum in the U.S. [47]. Similar behavior is observed in Figure 4d for pedestrian crossings. The coefficients are more stable in recent years, as seen in Figure 4b for personal vehicle crossings, whereas pedestrian crossings (Figure 4d) experienced a downward trend. In the case of personal vehicle (pedestrian) crossings, the mean of the coefficients of $d(\ln \text{Imm})$ was -0.085 (-0.125), with a standard deviation of 0.059 (0.094). These findings might imply that pedestrian crossings are more responsive to Imm.

The 4 panels in Figure 4 depict a rise in the elasticity between the BCI and Imm and border crossings during the 2008 and 2012 economic slumps, as demonstrated by the downtrend of the IGAE indicator in Figure 3. This suggests that sentiments may have a larger influence on the choice to cross the border during economic turmoil.

6. Discussion

The topics examined in this study may be relevant to those overseeing the economic development of the US–Mexico border regions. This manuscript reveals that the U.S. authorities can have an impact on Mexicans' intention to cross the border based on their perception of the new developments of U.S. immigration policies. It is observed that controlling for economic factors, such as the real exchange rate and economic activity, border crossings are also affected by the sentiments of people toward new immigration policies. Even when those policies have, in principle, no impact on legal immigration, how people perceive policy changes affects their willingness to cross the northbound border. The findings are relevant for communities in the U.S. that receive a significant number of visitors from Mexico.

This study also offers some potential lines of further research. COVID-19 has changed the dynamics of border crossings. Nowadays, Mexican pedestrians and personal vehicles have new reasons to visit the U.S. border regions. This is because the U.S. government is ahead of the Mexican government in the COVID-19 vaccination. Moreover, the situation has not precluded Mexican citizens from obtaining a vaccine in the U.S. before taking it in their own country. The U.S.–Mexican border was closed in March 2020 but opened in November 2021. This offers new motives for Mexicans for crossing, which are to look for booster shots, to be immunized with the vaccine of their choice, or to obtain vaccines for the underage population, whom the Mexican government has been reluctant to include in its vaccination policies. These new motives can also complement the previous motives for crossings discussed in this study. Indeed, the impact of closing and opening the border can also offer an interesting setting for testing hypotheses about the nature of border crossings along the US–Mexico border.

7. Conclusions

This study tested the hypothesis that the anti-immigrant environment in the U.S. may influence the sentiment of economic agents and thus affect the crossings along the US–Mexico border. Unlike previous studies, variables constructed from online search volume data were used. In addition, a rolling regression analysis was performed to examine the relationship between the sentiment variables and border crossings over time. Finally, all the PoEs along the border were included and grouped into three geographic regions.

The elasticities obtained suggest a negative relationship between the real exchange rate and both types of border crossings, which is consistent with previous literature that indicates that the depreciation of the Mexican peso has an adverse effect on border crossings. The influence of the Mexican economic cycle was slightly more significant for pedestrians (1.39%) than personal vehicle crossings (1.09%). Therefore, the border is crossed more frequently when the Mexican economy grows.

When Imm was included, the relationship between the real exchange rate and personal vehicle crossings was not statistically significant in Texas; this might imply that an anti-immigrant environment reduces the economic incentive for crossing. In contrast, for the pedestrian crossings in California, the inclusion of Imm did not reduce the statistical significance.

Following the studies of Ettredge et al. [39] and Guzman [42], it can be argued that the BEMS index variable is expected to capture people's sentiment to cross the border at the aggregate level. The BEMS was statistically significant in all models for both crossings, except pedestrians in California, which might indicate that the other factors analyzed have a greater influence in this region than the sentiment captured by the BEMS.

As suggested by Baruca and Zolfagharian [6], if a hedonic shopping motive drives customers to cross the border to pursue fun and enjoyment, the preceding findings may have some practical implications. Border crossings in Texas are more sensitive to changes in the anti-immigration environment and people's sentiment, as captured by the BEMS index; hence, it is vital to implement public policies that enhance a friendlier environment for those interested in crossing the border because of the economic impact of Mexicans' shopping trips on the U.S. border-city economies [25].

The rolling regression results demonstrate that the relationship between the sentiment proxy variables and border crossings is negative, except for short-run subperiods, which are positive, although with small coefficients. Thus, pedestrian crossings are more sensitive to changes in the sentiment proxy variables than personal vehicles. In addition, the elasticities of the sentiment proxy variables and border crossings increase during economic turmoil.

Author Contributions: Conceptualization and methodology, R.C., F.G.-F. and E.S.; software, F.G.-F.; validation and formal analysis, R.C., F.G.-F. and E.S.; data curation, F.G.-F.; writing—original draft preparation, F.G.-F.; writing—review and editing, R.C., F.G.-F. and E.S.; supervision, R.C. and E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Labastida-Tovar, E. The impact of NAFTA on the Mexican-U.S. border region. In *The Politics, Economics, and Culture of Mexican-U.S.*, 1st ed.; Ashbee, E., Clausen, H.B., Pedersen, C., Eds.; Palgrave Macmillan: New York, NY, USA, 2007; pp. 107–132.
2. Nicita, A. Who benefited from trade liberalization in Mexico? Measuring the effects on household welfare. In *Policy Research Working Paper, No.3265*; The World Bank: Washington, DC, USA, 2004.
3. Hanson, G. *What Has Happened to Wages in Mexico since NAFTA?* National Bureau of Economic Research: Cambridge, MA, USA, 2003. [\[CrossRef\]](#)
4. Fullerton, T.M.; Walke, A.G. Cross-Border Shopping and Employment Patterns in the Southwestern United States. *J. Int. Commer. Econ. Policy* **2019**, *10*, 1950015. [\[CrossRef\]](#)
5. Chandra, A.; Head, K.; Tappata, M. The Economics of Cross-Border Travel. *Rev. Econ. Stat.* **2014**, *96*, 648–661. [\[CrossRef\]](#)
6. Baruca, A.; Zolfagharian, M. Cross-Border Shopping: Mexican Shoppers in the US and American Shoppers in Mexico: Cross-Border Shopping. *Int. J. Consum. Stud.* **2013**, *37*, 360–366. [\[CrossRef\]](#)
7. Zhang, W.; Li, X.; Shen, D.; Teglio, A. Daily Happiness and Stock Returns: Some International Evidence. *Physica A* **2016**, *460*, 201–209. [\[CrossRef\]](#)
8. Quintana, P.J.E.; Ganster, P.; Stigler Granados, P.E.; Muñoz-Meléndez, G.; Quintero-Núñez, M.; Rodríguez-Ventura, J.G. Risky Borders: Traffic Pollution and Health Effects at US–Mexican Ports of Entry. *J. Borderl. Stud.* **2015**, *30*, 287–307. [\[CrossRef\]](#)
9. Heyman, J. Ports of Entry as Nodes in the World System. *Identities* **2004**, *11*, 303–327. [\[CrossRef\]](#)
10. Lee, E.; Wilson, C.E. The state of trade, competitiveness and economic well-being in the US-Mexico border region. In *Border Research Partnership Working Paper*; Wilson Center Mexican Institute: Washington, DC, USA, 2012.

11. Baggs, J.; Fung, L.; Lapham, B. Exchange Rates, Cross-Border Travel, and Retailers: Theory and Empirics. *J. Int. Econ.* **2018**, *115*, 59–79. [[CrossRef](#)]
12. Bello, P. Exchange rate fluctuations and border crossings: Evidence from the Swiss-Italian border. In *IdEP Economic Papers*; USI Università Della Svizzera Italiana: Lugano, Switzerland, 2017.
13. Cabral, R.; Mollick, A.V.; Saucedo, E. Northbound Border Crossings from Mexico to the U.S. and the Peso/Dollar Exchange Rate. *Res. Transp. Bus. Manag.* **2019**, *33*, 100423. [[CrossRef](#)]
14. Fullerton, T.M., Jr. Currency Movements and International Border Crossings. *Int. J. Public Adm.* **2000**, *23*, 1113–1123. [[CrossRef](#)]
15. Anderson, W.P.; Maoh, H.F.; Burke, C.M. Passenger Car Flows across the Canada–US Border: The Effect of 9/11. *Transp. Policy (Oxf.)* **2014**, *35*, 50–56. [[CrossRef](#)]
16. Fullerton, T.M., Jr.; Jiménez, A.A.; Walke, A.G. An Econometric Analysis of Retail Gasoline Prices in a Border Metropolitan Economy. *N. Am. J. Econ. Financ.* **2015**, *34*, 450–461. [[CrossRef](#)]
17. Adkisson, R.V.; Zimmerman, L. Retail Trade on the U.S.-Mexico Border during the NAFTA Implementation Era. *Growth Chang.* **2004**, *35*, 77–89. [[CrossRef](#)]
18. Burke, C.M. A Comparative Analysis of Cross Border Travel Influences at the Port Level: Pacific Highway/Douglas, B.C.—Blaine, WA and Windsor, ON—Detroit, MI. *Res. Transp. Bus. Manag.* **2015**, *16*, 95–101. [[CrossRef](#)]
19. Di Matteo, L.; Di Matteo, R. An Analysis of Canadian Cross-Border Travel. *Ann. Tour. Res.* **1996**, *23*, 103–122. [[CrossRef](#)]
20. Andreas, P. *A Tale of Two Borders: The U.S.-Mexico and U.S.-Canada Lines after 9/11*; Center for Comparative Immigration Studies, University of California at San Diego: La Jolla, CA, USA, 2003.
21. Bradbury, S.L. The Impact of Security on Travelers across the Canada–US Border. *J. Transp. Geogr.* **2013**, *26*, 139–146. [[CrossRef](#)]
22. Ferris, J.S. Quantifying Non-Tariff Trade Barriers: What Difference Did 9/11 Make to Canadian Cross-Border Shopping? *Can. Public Policy* **2010**, *36*, 487–501. [[CrossRef](#)]
23. Fullerton, T.M., Jr.; Walke, A.G. Homicides, Exchange Rates, and Northern Border Retail Activity in Mexico. *Ann. Reg. Sci.* **2014**, *53*, 631–647. [[CrossRef](#)]
24. Charney, A.H.; Pavlakovich-Kochi, V.K. The Economic Impacts of Mexican Visitors to Arizona. In *Economic and Business Research Program*; Eller College of Business and Public Administration, University of Arizona: Tucson, AZ, USA, 2002.
25. Ghaddar, S.; Brown, C.J. *The Economic Impact of Mexican Visitors along the US-Mexico Border: A Research Synthesis*; University of Texas-Pan American: Edinburg, TX, USA, 2006; pp. 1–20.
26. Patrick, J.M.; Renforth, W. The Effects of the Peso Devaluation on Cross-border Retailing. *J. Borderl. Stud.* **1996**, *11*, 25–41. [[CrossRef](#)]
27. Gerber, J. *The Effects of a Depreciation of the Peso on Cross-Border Retail Sales in San Diego and Imperial Counties*; University of California San Diego: San Diego, CA, USA, 1999.
28. Fullerton, T.M., Jr.; Molina, A.L., Jr.; Walke, A.G. Tolls, Exchange Rates, and Northbound International Bridge Traffic from Mexico. *Reg. Sci. Policy Pract.* **2013**, *5*, 305–321. [[CrossRef](#)]
29. Sharpe, W.F. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *J. Financ.* **1964**, *19*, 425–442. [[CrossRef](#)]
30. Fama, E.F. The Behavior of Stock-Market Prices. *J. Bus.* **1965**, *38*, 34–105. [[CrossRef](#)]
31. Simon, H.A. Bounded Rationality. In *Utility and Probability*; Palgrave Macmillan UK: London, UK, 1990; pp. 15–18. [[CrossRef](#)]
32. Ofek, E.; Richardson, M. DotCom Mania: The Rise and Fall of Internet Stock Prices. *J. Financ.* **2003**, *58*, 1113–1137. [[CrossRef](#)]
33. Qadan, M.; Nama, H. Investor Sentiment and the Price of Oil. *Energy Econ.* **2018**, *69*, 42–58. [[CrossRef](#)]
34. De Long, J.B.; Shleifer, A.; Summers, L.H.; Waldmann, R.J. Noise Trader Risk in Financial Markets. *J. Polit. Econ.* **1990**, *98*, 703–738. [[CrossRef](#)]
35. Baker, M.; Wurgler, J. Investor Sentiment and the Cross-Section of Stock Returns. *J. Financ.* **2006**, *61*, 1645–1680. [[CrossRef](#)]
36. Lemmon, M.; Portniaguina, E. Consumer Confidence and Asset Prices: Some Empirical Evidence. *Rev. Financ. Stud.* **2006**, *19*, 1499–1529. [[CrossRef](#)]
37. Da, Z.; Engelberg, J.; Gao, P. The Sum of All FEARS Investor Sentiment and Asset Prices. *Rev. Financ. Stud.* **2015**, *28*, 1–32. [[CrossRef](#)]
38. Irresberger, F.; Mühlnickel, J.; Weiß, G.N.F. Explaining Bank Stock Performance with Crisis Sentiment. *J. Bank. Financ.* **2015**, *59*, 311–329. [[CrossRef](#)]
39. Ettredge, M.; Gerdes, J.; Karuga, G. Using Web-Based Search Data to Predict Macroeconomic Statistics. *Commun. ACM* **2005**, *48*, 87–92. [[CrossRef](#)]
40. Ginsberg, J.; Mohebbi, M.H.; Patel, R.S.; Brammer, L.; Smolinski, M.S.; Brilliant, L. Detecting Influenza Epidemics Using Search Engine Query Data. *Nature* **2009**, *457*, 1012–1014. [[CrossRef](#)]
41. Choi, H.; Varian, H. Predicting the Present with Google Trends: Predicting the Present with Google Trends. *Econ. Rec.* **2012**, *88*, 2–9. [[CrossRef](#)]
42. Guzmán, G. Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations. *J. Econ. Soc. Meas.* **2011**, *36*, 119–167. [[CrossRef](#)]
43. Vosen, S.; Schmidt, T. Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends. *J. Forecast.* **2011**, *30*, 565–578. [[CrossRef](#)]
44. Diehl, P.N. The Effects of the Peso Devaluation on Texas Border Cities. *Tex. Bus. Rev.* **1983**, *57*, 120–125.

45. Da, Z.; Engelberg, J.; Gao, P. In Search of Attention. *J. Financ.* **2011**, *66*, 1461–1499. [[CrossRef](#)]
46. Preis, T.; Moat, H.S.; Stanley, H.E. Quantifying Trading Behavior in Financial Markets Using Google Trends. *Sci. Rep.* **2013**, *3*, 1684. [[CrossRef](#)]
47. Musalo, K.; Lee, E. Seeking a Rational Approach to a Regional Refugee Crisis: Lessons from the Summer 2014 “Surge” of Central American Women and Children at the US-Mexico Border. *J. Migr. Hum. Secur.* **2017**, *5*, 137–179. [[CrossRef](#)]

Article

Isolating the Role of the Transport System in Individual Accessibility Differences: A Space-Time Transport Performance Measure

Alberto Dianin ^{1,2,*}, Michael Gidam and Georg Hauger ^{1,*}

¹ Faculty of Architecture and Spatial Planning, Vienna University of Technology, Karlsplatz 11, A-1040 Vienna, Austria; michael.gidam@tuwien.ac.at

² Eurac Research, Institute for Regional Development, Viale Druso 1, I-39100 Bolzano, Italy

* Correspondence: alberto.dianin@eurac.edu (A.D.); georg.hauger@tuwien.ac.at (G.H.)

Abstract: Accessibility differences across individuals are a core topic in the transport equity debate. Space-Time Accessibility measures (*STAs*) have often been used to show such differences, given their sensitiveness to individual spatial and temporal constraints. However, given their complexity, *STAs* cannot properly isolate the specific role of the transport system in individual accessibility differences, since it is mixed with several other spatial, individual and temporal factors. To isolate the role of the transport system, this study introduces a Space-Time Transport Performance measure (*STTP*) that (a) grounds on the individual daily schedule of fixed activities, (b) calculates the generalised transport costs each individual has to bear to perform such schedule, and (c) weights it against the Euclidean distance between the activities of such a schedule. *STTP* is tested together with *STA* for a small sample of individuals living and performing their daily activities within the 22nd district of Vienna. This test provides two main findings: first, individual differences registered by *STTP* tend to be smaller than those highlighted by *STA*, according to the former's more narrowed and transport-specific approach. Second, individuals with the highest *STA* do not necessarily register the highest *STTP* (and vice versa). Indeed, some may experience limited transport performances when running their mandatory daily schedule, while registering a high degree of access to discretionary activities according to their constraints and opportunities at disposal (and vice versa). Considering these results, *STTP* may be seen as a complementary indicator to be used together with *STA* to analyse both general and transport-specific individual accessibility differences. Its role is particularly important for transport policy makers, who should understand which accessibility differences are directly linked to the performances of the transport system and could be remediated through transport policies.

Keywords: transport equity; distributional analysis; accessibility; space-time model; transport policy

Citation: Dianin, A.; Gidam, M.; Hauger, G. Isolating the Role of the Transport System in Individual Accessibility Differences: A Space-Time Transport Performance Measure. *Appl. Sci.* **2022**, *12*, 3309. <https://doi.org/10.3390/app12073309>

Academic Editors: Giovanni Randazzo, Anselme Muzirafuti, Dimitrios S. Paraforos and Stefania Lanza

Received: 18 February 2022

Accepted: 21 March 2022

Published: 24 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As highlighted by van Wee and Mouter [1], “in the transport policy literature, there is consensus that ‘sound’ policies have to meet three criteria: they should be effective, efficient and fair” [2]. Effectiveness and efficiency have received significant attention in the last decades [1], while the same does not apply to fairness except for contributions addressing social exclusion (e.g., ref. [3]). In recent years, the attention on transport fairness has increased thanks to the growing importance of inequality reduction at the international level, e.g., among the Sustainable Development Goals of the United Nations [4]. The lack of studies in this field is linked to the normative nature of fairness, making it difficult to measure and apply it to a cost–benefit analysis (one of the most diffused policy assessment tools [1,5]). Due to this normative issue, most studies in transport fairness focus on distributional analyses, i.e., how transport effects (such as air pollution variations or safety variations) are distributed over people [1,6]. One of the most addressed effects is the variation of *individual accessibility differences* [5]. Indeed, accessibility is one of the critical

pillars for a sustainable mobility paradigm [7]. Its improvement is one of the core concerns for transportation ministries throughout the world [8].

However, most accessibility measures developed in the literature are not suitable to point out such individual differences since they focus on the physical separation among places and overlook the accessibility differences that could exist among people living in the same area (e.g., because of their different modal restrictions or daily schedules [9]). To fill this gap, so-called *person-based* accessibility measures have been developed [10,11]. Among them, the *Space-Time Accessibility* measure (*STA*) is one of the most diffused given its sensitiveness to individual space-time constraints [12–14]. Although *STA* is very suitable for investigating individual accessibility differences in general, its high complexity does not allow a clear understanding of the specific role the transport system plays in such differences, differences which are fed by a multitude of spatial, transport, individual and temporal factors. This is a relevant limit of *STA* especially from the perspective of transport policy makers, who should introduce transport policies aimed at reducing such differences. Indeed, accessibility differences are often unavoidable and the transport system cannot remediate them [1]. For instance, inevitably, people living close to a large facility (e.g., a hospital) have better access to it than people living far away. However, this is unavoidable due to land-use constraints (hospitals cannot be built in any municipality), and transport policies cannot eliminate distance. Given this issue, it is necessary to complement *STA* with transport performance measures (With the term “transport performance” we mean a set of transport indicators that describe the efficiency of the transport system, such as commercial speed, network capacity, service period, average waiting or transfer time, or monetary cost of travel) able to isolate the specific role of the transport system in individual accessibility differences, which could be avoided or remediated through transport policy interventions. For this purpose, this paper introduces a so-called *Space-Time Transport Performance* measure (hereinafter *STTP*).

The rest of the article is organised as follows. Section 2 reviews main place- and person-based accessibility measures used in literature and points out the elements that prevent *STA* from isolating the role of the transport system in individual accessibility differences. On this basis, Section 3 introduces *STTP* by describing its key features and its computation process. *STTP* is then tested in Section 4 together with *STA* to determine the complementary results that are achievable with the proposed measure. Moreover, it discusses the limits of *STTP*. Section 5 concludes the contribution by highlighting potential contexts of application.

2. Place- and Person-Based Accessibility Measures

In general terms, place-based measures calculate accessibility for a location by assuming that all people in that location register the same accessibility. Conversely, person-based calculations analyse accessibility for individuals living in the same area to understand how accessibility varies across them [15]. Sections 2.1 and 2.2 summarise the main place- and person-based measures developed and used in literature, while Table 1 displays their definition, mathematical formulation and main conceptual and operational pros and cons. Afterwards, Section 2.3 discusses the factors that prevent *STA* from isolating the role of the transport system in individual accessibility differences.

2.1. Place-Based Measures

Place-based measures calculate how easy it is for people departing from a place to reach opportunities located in another [9,16]. Following this definition, three main types of place-based measures have been developed: *cumulative-opportunity*, *gravity-based* and *adapted gravity-based* measures ([9,17]; Table 1). All of these are a function of two core elements: (a) an attraction factor given by the amount, spatial distribution and quality of opportunities to access; and (b) an impedance factor given by the effort needed to reach these opportunities [18]. These are combined in different ways. The *cumulative-opportunity measure* (e.g., refs. [19–21]) counts the number of opportunities reachable from

an origin location within a predetermined threshold usually defined in terms of travel time or cost. It is a very straightforward approach, but it shows some limits since counted opportunities have the same importance regardless of the effort needed to reach them. The *gravity-based measure* (e.g., refs. [22–24]) addresses these limits since it calculates the accessibility of an origin as a function of the number and importance of opportunities at the destination, weighed against the travel effort needed to reach them. However, it neglects the competition between the demand for and supply of opportunities. The *adapted gravity-based measure* (e.g., refs. [25–27]) incorporates them through a double constrained spatial interaction model with two mutually-dependent balancing factors [17].

These three measures have two main limitations that person-based measures aim to overcome [13,28]. First, they assume that all individuals who depart from the same origin location experience the same level of accessibility regardless of their different spatial, temporal and modal constraints. Second, they calculate accessibility for a single reference location (typically home place) overlooking the fact that people generally perform a sequence of daily activities that are differently located in space and time, and this sequence affects their accessibility.

2.2. Person-Based Measures

Three main types of person-based measures address the gaps of the place-based measures: the *utility-based*, *individual integral* and *space-time* measures ([11,17]; Table 1). The *utility-based measure* (e.g., refs. [9,29,30]) grounds on economic theories and calculates accessibility as the maximal economic utility individuals can get from the access to spatially distributed opportunities based on their perception of the utility of the options at their disposal. The *individual integral measure* (e.g., refs. [31–33]) is a gravity-based measure adjusted to be person-specific. The adjustment is performed either by disaggregating data and analysis by, e.g., trip purposes, transport modes, age or income groups; or by using a non-zonal method as the point-based approach, which allows a focus on specific point locations and measurement of point-to-point travel costs at the individual level. Although person-based, this last measure focuses on a single reference location, and it overlooks the spatio-temporal constraints affecting people on a daily basis [34]. The *space-time measure* addresses these limitations in the most comprehensive manner (e.g., refs. [14,35,36]).

For this reason, this is considered an effective approach to measure accessibility at the individual level and to discuss individual accessibility differences [10]. It derives from the time geography framework elaborated by Hägerstrand [37] and focuses on the set of discretionary opportunities (i.e., non-mandatory daily activities) that individuals could reach on a daily basis given the spatio-temporal constraints posed by their fixed daily activity chain (i.e., mandatory daily activities) [38]. This set is called a Feasibility Opportunity Set (*FOS*), and is obtained in three steps. First, the daily sequence of fixed activities constrained in space and time for a person is schematised. This sequence generates a so-called Space-Time Path (*STPA*). Based on the *STPA*, the Potential Path Areas (*PPAs*) are calculated for each couple of the following fixed activities in the *STPA*. Each *PPA* includes all the locations that an individual could visit between two subsequent fixed activities, given the mandatory departure time from the former, the mandatory arrival time at the latter, the time needed to travel between them, and the time required to visit such locations. By extending the calculation of the *PPA* to all couples of sequential fixed activities, the Daily Potential Path Area is obtained (*DPPA*). All the opportunities that belong to the *DPPA* constitute the *FOS* and define the space-time accessibility measure (*STA*).

Table 1. Main place- and person-based accessibility measures and their key features.

Accessibility Measure	General Definition	Mathematical Formulation *	Conceptual/Operational Pros	Conceptual/Operational Cons	Sample References
Place-based measures	Cumulative opportunity	Set of opportunities reachable from a location within a pre-defined travel time or cost $A_i = \sum_{j=1}^{j=n} O_j f(TC_{ij})$ $f(TC_{ij}) = \begin{cases} 1 & \text{for } TC_{ij} \leq T \\ 0 & \text{otherwise} \end{cases}$	Easily understandable and applicable	Opportunities have the same importance regardless of their features	[19–21]
	Gravity-based	Set of opportunities reachable at destination weighed by the travel effort needed to reach them $A_i = \sum_{j=1}^{j=n} O_j f(TC_{ij})$	It considers the role of the travel effort in decreasing accessibility	It neglects temporal constraints and competition effects	[22–24].
	Adapted gravity-based	Gravity-based measure including competition effects between demand for and supply of opportunities $a_i = \sum_{j=1}^{j=n} \frac{1}{b_j} O_j f(TC_{ij})$ $b_j = \sum_{i=1}^{i=n} \frac{1}{a_i} D_i f(TC_{ij})$	It includes competition effects between demand for and supply of opportunities	Particularly difficult to operationalize and adopt in concrete analyses	[25–27].
Person-based measures	Utility-based	Maximal economic utility individuals can get from the access to spatially distributed opportunities $A_u = \frac{1}{\lambda} \ln \sum_{z \in C_u} \exp(\lambda V_{zu})$	Utility is measured at individual level and results can feed economic evaluations	Temporal constraints are neglected while results are not easily understandable	[9,29,30]
	Individual integral	Gravity-based measure adapted to analyse individual accessibility through disaggregation or point-based approach $A_{ui} = \sum_{j=1}^{j=n} O_j^u f(TC_{ij}^{u,k})$	It allows focusing on e.g., specific trip purposes, transport modes, or age groups	It measures accessibility for a single location and neglects spatio-temporal constraints	[31–33]
	Space-time	Set of opportunities reachable by an individual according to his/her daily activity programme and constraints $A_u = \sum_{w=1}^{w=n} O_w I(w)$ $I(w) = \begin{cases} 1 & \text{if } w \in DPPA \\ 0 & \text{otherwise} \end{cases}$ $DPPA = \{(w, t) \mid t_a + \frac{d_{a,w}}{v} \leq t \leq t_{a+1} + \frac{d_{w,a+1}}{v}\}$	Sensitive to individual transport specificities and spatio-temporal constraints	It needs peculiar input data and skip transport factors beyond travel time	[14,35,36]

* Where: **Cumulative opportunity:** i is an origin location; j is a destination location; O_j are the opportunities available at destination; TC_{ij} is the cost of travelling from i to j ; $f(TC_{ij})$ is the travel cost function, which may assume different forms such as linear, Gaussian, logistic or negative exponential; T is the travel cost threshold set in the analysis. **Gravity-based:** i, j, O_j, TC_{ij} and $f(TC_{ij})$ are defined above. **Adapted gravity-based:** i, j, O_j, TC_{ij} and $f(TC_{ij})$ are defined above; a_i is the balancing factor for demand in location i ; b_j is the balancing factor for supply in location j ; D_i is the demand for opportunities in i . **Utility-based:** u is a user for whom accessibility is calculated; λ is the travel cost coefficient; z is one of the choices that u can make; C_u is the set of choices z that u can make; V_{zu} is the systematic utility of the choice z for u . **Individual integral:** u, i, j, O_j and $f(TC_{ij})$ are defined above; $TC_{ij}^{u,k}$ is the travel cost for u from i to j by transport mode k . **Space-time:** u is defined above; w_{1-n} are the locations of discretionary opportunities; O_w are the discretionary opportunities O_w available in w_{1-n} ; $DPPA$ is the Daily Potential Path Area; t is the time needed to participate in a discretionary opportunity O_w ; t_a is the ending time of a fixed activity a ; t_{a+1} is the starting time of the following fixed activity $a+1$; $d_{a,w}$ is the physical distance between a and w ; $d_{w,a+1}$ is the physical distance between w and $a+1$; v is the average speed on the transport network.

2.3. Limits of STA in Describing the Role of the Transport System

Thanks to its individual sensitiveness and capacity to comprise all the four accessibility components (land-use, transport, individual and temporal; [17]), STA is often adopted in the analysis of accessibility differences (e.g., refs. [10,12,39]). Nevertheless, STA presents some limits when it comes to isolating the role of the transport system in individual accessibility differences. In particular:

- **Limited relevance of the transport system performances in the FOS:** STA is represented by the FOS, which mostly depends on the amount and spatial distribution of the discretionary opportunities and the spatio-temporal constraints of the STPA [11]. Therefore, STA focuses highly on spatial and temporal accessibility components and less on transport performances [17]. This is a gap when the aim is to isolate the specific role of a transport system in individual accessibility differences. For instance, let us

assume two people who both have one fixed activity during their day, departing and headed to the same locations simultaneously, and using the same transport system with the same performances. The former works full-time while the latter part-time. According to the *STA* concept, the different time constraints of the two individuals would lead to an accessibility difference since the part-time worker has more occasions to engage in discretionary activities than the full-time worker. However, the performance of the transport system does not play a role in such accessibility differences.

- **Unsuitability of the *FOS* to represent accessibility differences:** As stressed by Pritchard et al. [40,41], the choice of the accessibility measure may significantly influence the outcomes of the analysis. Therefore, it is crucial to deploy a measure that is as suitable as possible to discuss accessibility distribution. Specifically, the estimate should represent an optimisation factor for the observed individuals, i.e., a good they generally aim to increase [42]. This is the case, e.g., with income, which is one of the critical indicators for distributional analyses in socio-economic sciences [43]. *STA* cannot be easily labelled as an optimisation factor since it is not straightforward to state that individuals aim to maximise the number of discretionary opportunities they could reach on a daily basis. For instance, a person could have a small *FOS* because (s)he has a tight schedule of fixed activities and no room to engage in discretionary ones. Nevertheless, (s)he could be not much interested in further activities. At the same time, the transport system could be efficient in allowing them to reach all the fixed activities with a reasonable effort [5].

Based on these limits, we introduce the so-called Space-Time Transport Performance measure (*STTP*) in order to complement *STA* by isolating the role of the transport system in individual accessibility differences.

3. Space-Time Transport Performance Measure (*STTP*)

3.1. Key Features of *STTP*

STTP aims to measure the performances of the transport system based on the spatio-temporal and individual constraints characterising the daily life of each individual. To meet this purpose, *STTP* grounds on three key features, described below in detail.

- **Focus on the Individual Daily Travel Cost (*IDTC*) incurred for the daily fixed activities:** *STTP* is not focused on the sum of the discretionary opportunities potentially reachable given the schedule of fixed activities (i.e., the *FOS*). Instead, it focuses on the individual travel cost incurred to perform the daily schedule of fixed activities (from now on named Individual Daily Travel Cost; *IDTC*). *IDTC* is calculated as the generalised cost of transport incorporating both monetary and non-monetary cost (see Section 3.2 for further details). This shift of perspective allows *STTP* to focus on a factor that is transport-specific rather than land-use-specific and thus isolate the performance of the transport system. Moreover, it allows *STTP* to focus on an indicator that is suitable to represent transport-related accessibility differences. Indeed, individuals tend to minimise the transport cost needed to reach their fixed daily destinations [42,44]. However, by focusing on *IDTC*, *STTP* also excludes the discretionary part of accessibility typically included in *STA* and representing the potential for activities offered by surrounding amenities. This choice requires *STTP* to be complemented with *STA* to capture the potential component of accessibility.
- **Weighting of *IDTC* against the Individual Daily Distance (*IDD*):** *IDTC* is usually influenced by the distance daily travelled: the higher the distance, the higher *IDTC*. This may be misleading for evaluating individual transport-performance differences, since even in the case of identical transport performances, people who travel longer distances would be more disadvantaged than those travelling short ones. To isolate the role of the transport system performance in individual accessibility differences, this variable needs to be controlled to exclude its influence from the analysis. For this purpose, *STTP* weights *IDTC* against the Individual Daily Distance (*IDD*). This is the sum of the Euclidean distances between each couple of subsequent activities

belonging to an individual’s daily schedule (see Section 3.2 for further details). This choice makes *STTP* a measure of transport performance rather than a measure of the daily transport effort of individuals (which would be influenced also by the distance daily covered).

- **Estimation of *IDTC* based on temporal and individual constraints:** To incorporate the temporal constraints, *STTP* calculates *IDTC* by considering the actual location and timing of the fixed activities daily performed by an individual. Also, the individual constraints are incorporated in the *IDTC* computation in two ways. First, the actual modal choices of individuals for each daily travel are considered according to individual constraints such as the ability to drive or car ownership. Second, the non-monetary cost part of *IDTC* (i.e., travel-time costs) are estimated at the individual level based on income (as described in detail in Section 3.2).

These three features shape *STTP* as a transport performance indicator that stems from spatio-temporal and individual constraints. On the one hand, this allows *STTP* to isolate the role of the transport system in individual accessibility. On the other hand, it suggests how *STA* and *STTP* should be deployed together to get both a comprehensive picture of space-time accessibility and more narrowed information on the transport component. Figure 1 summarises the relation between *STA* and *STTP*, while Section 3.2 describes how *STTP* is calculated.

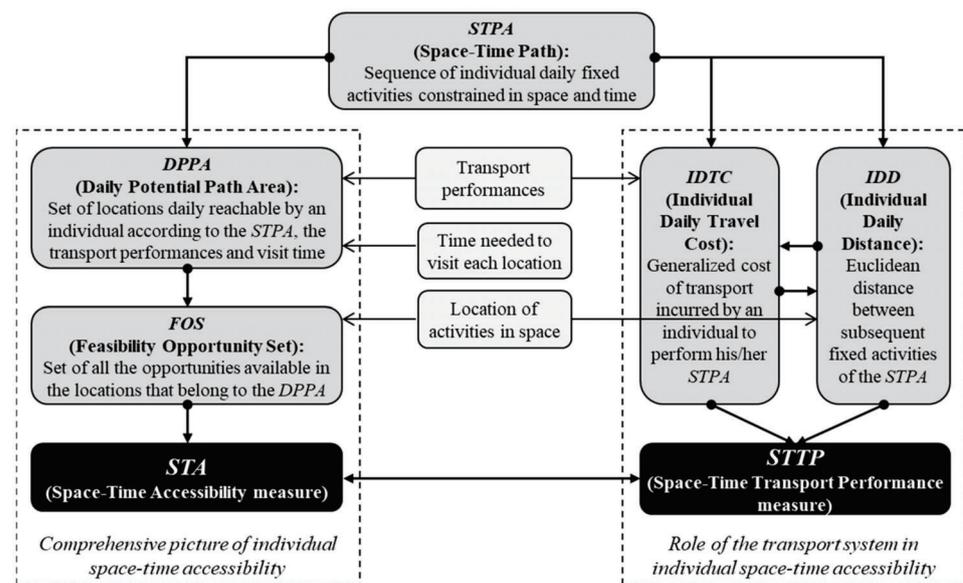


Figure 1. Relation between *STA* (left) and *STTP* (right) and their key features.

3.2. Calculation of *STTP*

STTP (Formula (1)) is calculated by following three steps: (A) the setup of the *STPA*; (B) the calculation of the *IDTC* figures; (C) the calculation of the *IDD* figures.

(A) The setup of the *STPA* is made in the same way as for *STA*. The daily sequence of fixed activities (a_{1-n}) constrained in space and time for the observed individual is schematised. This includes the location where each a takes place (address); its category (home-stay, work, education or other); the duration of each a given by its mandatory starting and ending time; the transport mode(s) usually used to travel between each couple of subsequent activities (a); and the degree of fixity for each a_n according to the flexibility of its location and timing. This is made through a 1–5 Likert scale, where 1 indicates maximum flexibility of the location and/or timing of a , while 5 shows a minimum one. The data needed to reconstruct the *STPA* are collected from observed individuals using travel diaries. Interviewed individuals are asked to fill them out by considering a typical weekday of their daily life. Table 2 shows an exemplificative (and fictional) *STPA*.

Table 2. Exemplificative STPA to be set up for the calculation of STTP.

Main STPAs Information	Fixed Activity a_1	Fixed Activity a_2	Fixed Activity a_3	Fixed Activity a_4	Fixed Activity a_5
Activity category:	Home-stay	Other	Work	Other	Home-stay
Activity location:	Street 1	Street 2	Street 3	Street 2	Street 1
Activity timing:	00:00–07:30	07:50–08:00	09:00–17:00	17:30–18:00	19:00–24:00
Activity fixity degree:	4	5	5	3	4
* Transport mode(s):	-	Car driver	Car driver	Car driver	Car passenger

* Notes: the transport mode used to reach the activity.

(B) Once the STPA is set, IDTC is calculated (Formula (2)). IDTC is the sum of the Individual Transport Costs incurred by an individual for each daily travel performed by the transport mode k between each couple of subsequent fixed activities $a, a+1$ ($ITC_{a,a+1}^k$). Each $ITC_{a,a+1}^k$ value is calculated through a generalised cost function, including a series of monetary and non-monetary (but monetizable) costs [42]. This consists of the monetary cost of travel (Cm), the cost of the in-vehicle travel time ($Civtt$), and the cost of out-of-vehicle travel time ($Covtt$). Cm encompasses the costs for the usage of infrastructures (e.g., tolls and parking fares), the operating costs of vehicles (e.g., fuel, usage-related depreciation and insurance), and the costs for access to services (e.g., public transport; from now on PT). $Civtt$ includes the time spent within private, shared or pooled vehicles. $Covtt$ covers the cost of the time to access the first transport system (first mile), the waiting time for transport services, the transfer time among transport services, and the time to reach the final destination (last mile [42]). $Civtt$ and $Covtt$ are monetised based on unitary Values of Travel Time (VTT). As demonstrated in the literature, VTT may vary by income, country, travel purpose, mode of transport and distance. It depends on the approach with which it is estimated (e.g., stated vs. revealed preference surveys [44]). The wage rate method is used for STTP: different wage rate shares are assumed depending on the country of investigation, travel purpose, and transport mode. Moreover, the actual wage rates of observed individuals are used to make the estimation individual.

(C) Once IDTC is calculated, this has to be weighed against IDD (Formula (3)). This is the sum of the distances measured between each couple of subsequent fixed activities $a, a+1$ ($D_{a,a+1}$). Each $D_{a,a+1}$ value is measured as Euclidean and not travelled distance along with the transport network. Indeed, the travelled distance may be influenced by the design of the transport system and not only by the land-use system. For instance, this is the case with fast transport systems such as motorways and high-speed railways, which tend to generate much more detours than slower systems (a phenomenon called “spatial inversion” by Bunge [45]). This detour is an aspect that transport planners can potentially address e.g., by modifying the shape of the PT lines and distribution of stops. Therefore, it is a factor to be included in the accessibility computation. Conversely, the Euclidean distance solely depends on the land-use system (i.e., the amount and location of opportunities in space) therefore it is used to weight IDTC and point out the role of the transport system in accessibility differences. Figure 2 summarises the calculation process for STA (in red) and STTP (in blue) based on STPA (in black), which is the common element between them.

$$STTP = \frac{IDTC}{IDD} \tag{1}$$

$$IDTC = \sum_{a_1}^{a_n} ITC_{a,a+1}^k \quad \text{with} \quad ITC_{a,a+1}^k = Cm_{a,a+1}^k + Civtt_{a,a+1}^k + Covtt_{a,a+1}^k \tag{2}$$

$$IDD = \sum_{a_1}^{a_n} D_{a,a+1} \tag{3}$$

where: a_{1-n} are the fixed activities performed by an individual on a daily basis, k is the mode(s) of transport used by an individual between each couple of subsequent as , IDTC

is the Individual Daily Travel Cost incurred by an individual on a daily basis, IDD is the Individual Daily Distance between the as performed by an individual on a daily basis, $ITC^k_{a,a+1}$ is the individual transport cost by mode k between each couple of subsequent as , $Cm^k_{a,a+1}$ is the monetary cost of transport by mode k between each couple of subsequent as , $Civtt^k_{a,a+1}$ is the cost of in-vehicle travel time by mode k between each couple of subsequent as , $Covtt^k_{a,a+1}$ is the cost of out-of-vehicle travel time by mode k between each couple of subsequent as , and $D_{a,a+1}$ is the Euclidean distance between each couple of subsequent as .

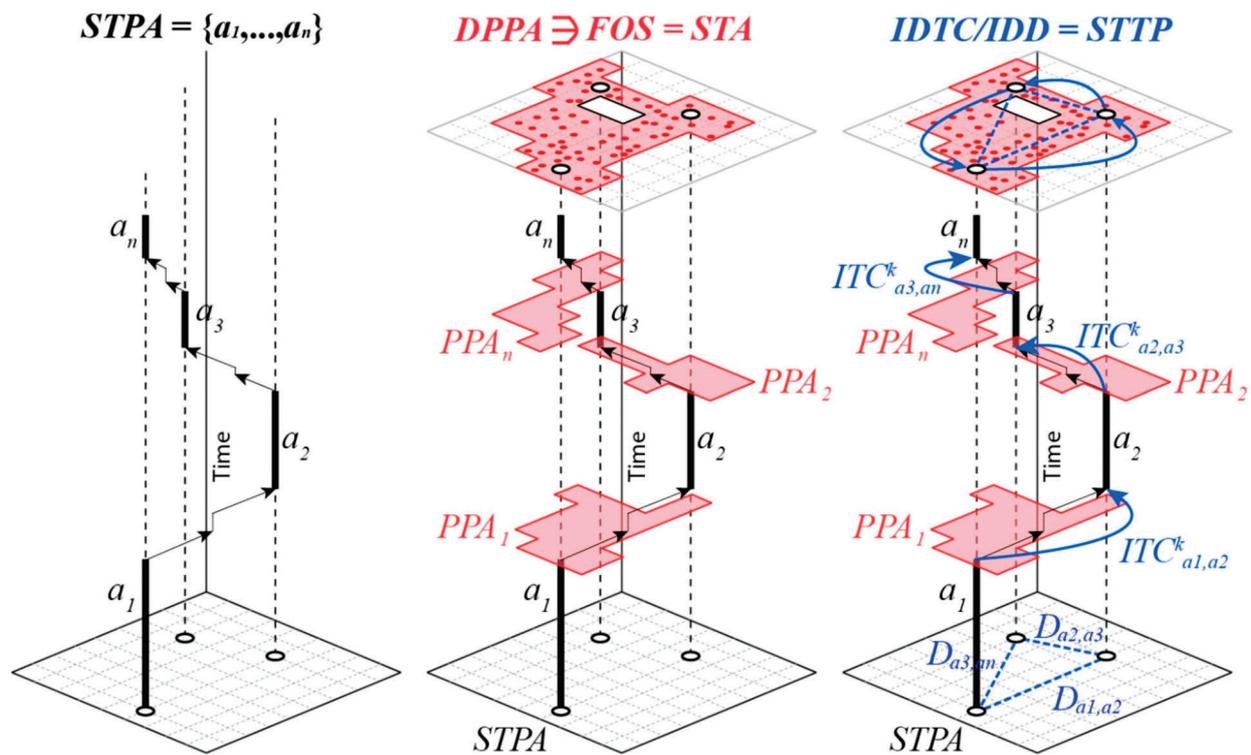


Figure 2. Process for the computation of $STPA$ (black), STA (red) and $STTP$ (blue).

4. Joint Test of STA and $STTP$ in the City of Vienna

The test aims to show how $STTP$ may lead to complementary results for STA , providing insights on the role of the transport system in individual accessibility differences. Similar to other studies focused on the methodological integration of the space-time approach (e.g., refs. [36,46]), we run the test for a small sample of five individuals, for whom STA and $STTP$ are calculated. We perform such a small test for two reasons: first, because the purpose is to provide a methodological test and not to get statistically relevant results about a specific phenomenon (e.g., gender or income-related accessibility differences). Second, focusing on a few individuals allows a more detailed reflection on results, e.g., pointing out the main differences for each individual (Section 4.3). This would not be feasible with a test involving many individuals, which would be more suitable for statistical comparison. Nevertheless, focusing on such a small test also has some limits, which are discussed in Section 4.4.

4.1. Study Area and $STPAs$

The test is run in the City of Vienna, Austria. The analysed individuals (A–E) live and perform their fixed activities within the 22nd district (Donaustadt). This is the second northernmost district of Vienna, with the greatest surface (ca 102 km²), the second-highest population (ca 198,800 inhabitants), and the second-lowest population density out of the 23 city districts (1943 inhabitants/km²). The district is served by 26 bus lines, two light rail lines and two subway lines. Moreover, it is served by a road network that is denser

in the core part of the district characterised by a higher urban density and much lower at the fringes, where green areas are predominant. Given the heterogeneous availability of transport means, the area represents a suitable case study to explore the individual accessibility differences related to the transport system. Figure 3 displays the road and PT network of the study area, the location of the available opportunities, the location of the home places, and the fixed activities of the individuals.

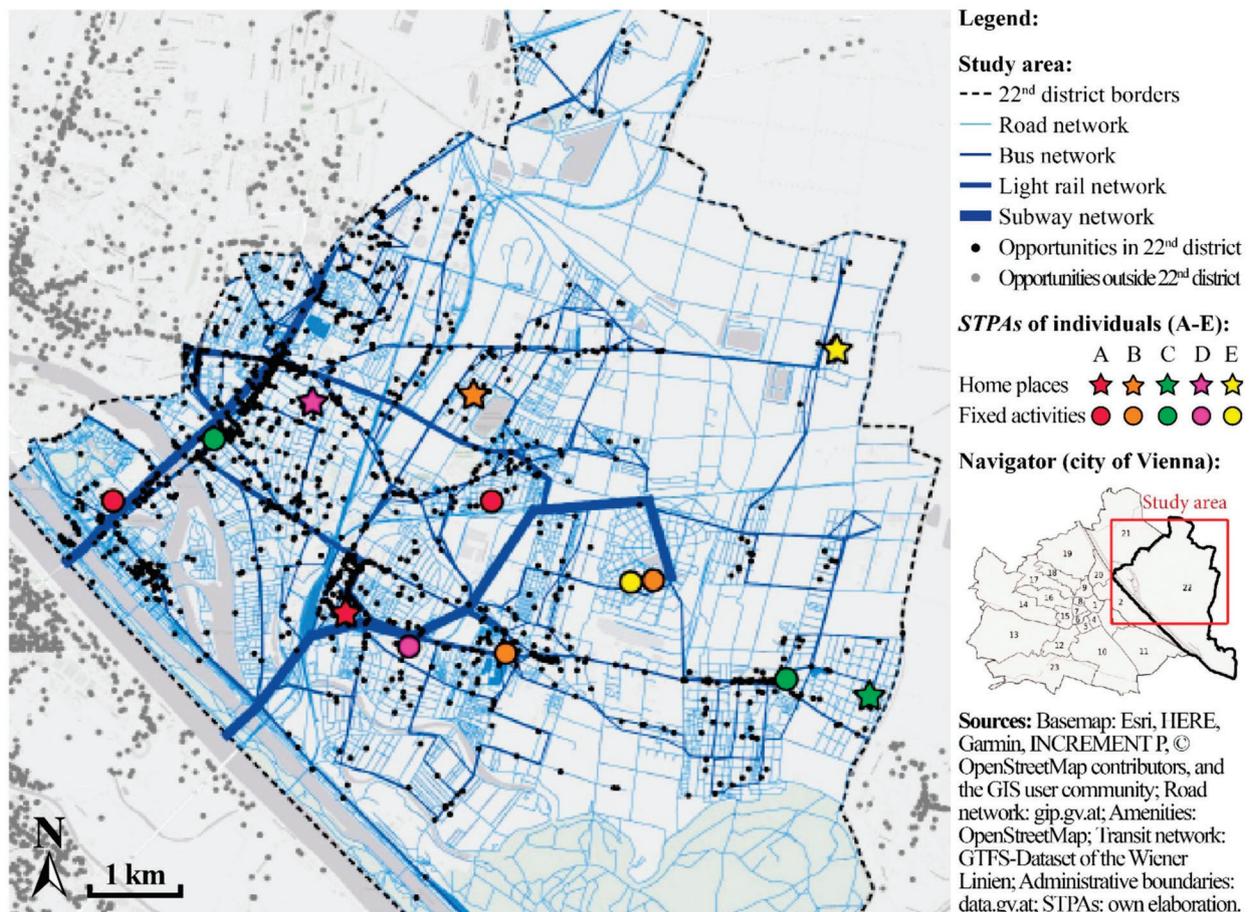


Figure 3. Study area and key locations of the STPAs of the individuals A–E.

After defining the study area, the first step for calculating both *STA* and *STTP* is the setup of the *STPAs* (summarised in Table 3 for individuals A–E). Each *STPA* describes the categories of fixed activities daily performed, their location (not listed in Table 3), their starting and ending time, their fixity degree (expressed with a 1–5 Likert scale), and the transport mode(s) used to reach them. Individual A is a full-time worker who takes their child to school before reaching the work place, stays at the work place until late afternoon, and then comes back home in the evening. (S)He always travels by car. Individual B is a part-time worker who works in the morning and picks up their child after school on the way home at lunchtime. In the afternoon, (s)he has to stay at home for some household duties in the timespans 13:00–15:00 and after 17:30, while (s)he has free time to engage in discretionary activities between 15:00 and 17:30. (S)He always travels by car for all the mandatory travels. Individual C is a part-time worker too. In the early morning and at lunchtime (s)he takes their child to school and back home by walking. In between, (s)he has some household duties to perform at home (from 08:20 to 09:30 and from 12:00 to 14:00) and a free time to engage in discretionary activities in between (from 09:30 to 12:00). From the afternoon until the evening, (s)he works part-time. (S)He travels by PT to and from the workplace. Individual D is a pensioner who visits the hospital in the morning on a

daily basis and travels to and from the hospital by PT. In the afternoon, (s)he has free time to engage in discretionary activities (between 15:00 and 18:00), while (s)he has to stay at home for the other hours of the afternoon. Finally, Individual E is a teenager who goes to school. In the morning, a parent takes them to school by car. In the early afternoon, (s)he has to come back home by PT and stays at home for their mandatory activities till 18:00. Afterwards, (s)he has free time in the late afternoon to engage in discretionary activities, before having to be again at home at 20:00.

Table 3. The STPAs of individuals A–E.

Individuals	Main STPAs Information	STPAs of Individuals (A–E)							
		Fixed Activity a_1	Fixed Activity a_2	Fixed Activity a_3	Fixed Activity a_4	Fixed Activity a_5	Fixed Activity a_6	Fixed Activity a_7	Fixed Activity a_8
A	Activity category:	Home-stay	Other	Work	Home-stay				
	Activity timing:	00:00–07:30	07:40–07:45	08:00–17:30	18:00–24:00				
	Activity fixity degree:	5	5	5	4				
	Transport mode(s):	-	Car driver	Car driver	Car driver				
B	Activity category:	Home-stay	Work	Other	Home-stay	Home-stay			
	Activity timing:	00:00–06:30	07:00–12:00	12:30–12:35	13:00–15:00	17:30–24:00			
	Activity fixity degree:	5	5	5	4	3			
	Transport mode(s):	-	Car driver	Car driver	Car driver	PT, Walking			
C	Activity category:	Home-stay	Other	Home-stay	Home-stay	Other	Home-stay	Work	Home-stay
	Activity timing:	00:00–07:25	07:45–07:50	08:20–09:30	12:00–14:00	14:15–14:20	14:35–17:00	18:00–23:00	24:00–24:00
	Activity fixity degree:	4	5	3	4	5	4	5	5
	Transport mode(s):	-	Walkig	Walking	PT, Walking	Walking	Walking	PT	PT
D	Activity category:	Home-stay	Other	Home-stay	Home-stay				
	Activity timing:	00:00–09:00	10:00–11:30	13:00–15:00	18:00–24:00				
	Activity fixity degree:	4	5	4	3				
	Transport mode(s):	-	PT	PT	PT, Walking				
E	Activity category:	Home-stay	Education	Home-stay	Home-stay				
	Activity timing:	00:00–07:30	08:00–14:00	15:00–18:00	20:00–24:00				
	Activity fixity degree:	5	5	4	4				
	Transport mode(s):	-	Car passenger	PT	PT, Walking				

Notes: **Activity category:** We define four activity categories: Home-stay, Work, Education, and Other. For individuals B–E, two consecutive home-stays occur in their STPAs. This is because they have two consecutive mandatory activities to perform at home and a free-time span in between to potentially engage in discretionary activities. **Activity timing:** This indicates the starting and ending time of each activity. The first and last activities always start at 00:00 and end at 24:00. **Activity fixity degree:** 1–5 Likert scale, with 1 indicating maximum flexibility and 5 maximum fixity. **Transport mode:** This indicates the transport mode(s) used to reach the related activity. We define four transport modes: Car driver, Car passenger, PT, and Walking. Since the first activity is always the early-morning home-stay, there is no transport mode assigned to reach it.

4.2. STA and STTP Calculation

To calculate STA, (A) the travel-time performances of the transport mode(s) used by the individuals A–E have to be estimated, and (B) the discretionary opportunities in the study area have to be mapped. These steps are implemented in ArcGIS by calculating different route analyses and service-area analyses through the Network Analyst extension. The estimation of these two components is described below in detail.

- (A) **Travel-time performances ($tt_{a,a+1}$):** $tt_{a,a+1}$ by car, PT and walking is estimated via GIS by using the GTFS-Dataset of the Wiener Linien and the Austrian Graphenintegrations-Plattform GIP [47,48]. Road network performances include speed limits, one-way streets, turn prohibitions and actual traffic conditions. According to time schedules, PT performances include travel time between stops and waiting time at the stops.

The transfer time between lines or modes is not yet available for the city of Vienna. Therefore, we assume an average value of one minute for buses and light rail and three minutes for the subway, plus the related waiting time. Finally, $tt_{a,a+1}$ by walking is estimated based on the existing network of sidewalks and an assumed walking speed of 5 km/h.

- (B) **Discretionary opportunities (O_w):** The set of O_w available in the study area is geo-referenced using OpenStreetMap as a core data source. These comprise all the study areas' amenities apart from workplaces, schools, and other educational facilities. Therefore, they mainly include groceries, shopping facilities, healthcare facilities, leisure facilities and other services such as post offices and banks. We consider all O_w to have the same importance for all individuals for the STA computation. Additionally, we assume that all O_w need at least a 10-min stay to be considered in the $PPAs$.

To calculate $STTP$, it is necessary to estimate (A) the unitary value of travel time for both $Civtt^k_{a,a+1}$ and $Covtt^k_{a,a+1}$; (B) the unitary monetary cost for $Cm^k_{a,a+1}$; and (C) the Euclidean distances among fixed activities, i.e., $D_{a,a+1}$. Even these components are implemented in ArcGIS through the Network Analyst extension by calculating different route analyses and described below in detail.

- (A) **Unitary value of travel time (VTT):** VTT is estimated with the wage rate method [44]. According to this approach, the value of travel time outside working hours (called Off-The-Clock Travel Time) for the driver is empirically found to be approximately 60% of the wage rate, excluding benefits. This percentage tends to decrease to 45% for passengers (of cars and PT) and increases to 100% when considering any kinds of out-of-vehicle travel time (i.e., walk-access, waiting, and transfer times). These differences depend on the perceptions of disutility of travel time for different modes of transport. Generally, travel time by PT or as car passenger has a higher utility since it is possible to make a profitable use of that time (e.g., to read, work or relax). When focusing on travel time during working hours (called On-The-Clock Travel Time), a percentage of 100% is considered for any kind of in-vehicle and out-of-vehicle travel time. Since we do not have individual income data at our disposal, we rely on the average hourly wage rates registered in Austria in 2018 for four different categories of people relevant for our case study, i.e., full-time workers (€16.22/h), part-time workers (€13.78/h), pupils (€9.88/h) and pensioners (€8.89/h) [49]. Accordingly, VTT is calculated for each individual and transport mode as summarised in Table 4. Combining these values with the travel-time performances, we obtain the $Civtt^k_{a,a+1}$ and $Covtt^k_{a,a+1}$ figures for each individual.

Table 4. VTT values applied to the different modes of transport and categories of individuals.

Categories	Full-Time Workers	Part-Time Workers	Pensioners	Pupils
Related individuals of the test	A	B and C	D	E
Hourly wage rate	€16.22/h	€13.78/h	€8.89/h	€9.88/h
VTT for car drivers	€9.73/h	€8.27/h	€5.33/h	€5.93/h
VTT for car and PT passenger	€7.30/h	€6.20/h	€4.00/h	€4.45/h
VTT for out-of-vehicle travel time	€16.22/h	€13.78/h	€8.89/h	€9.88/h

- (B) **Unitary monetary cost (UMC):** UMC is estimated for private cars and PT in two different ways. For private vehicles, we rely on the average kilometric Vehicle Operating Cost (VOC) for passenger cars in Austria. This includes the average cost of fuel and oil, maintenance and repair, tyres, and kilometric-dependent depreciation. According to the EU report by Infrast [50] and the yearly values provided by ACEA for all EU countries, [51], a VOC of €0.42/km is assumed for Austria. This is multiplied by the distance travelled to obtain $Cm^k_{a,a+1}$ figures for each individual travelling by car. As

for PT, the transport operator of the city of Vienna offers different yearly subscriptions covering the whole urban transport system [52]. Given the age and mobility habits of individuals, three subscriptions are considered: the annual ticket for adults (€365/year), for seniors +65 (€235/year), and for students till 24 years old (€79/year). According to these fares, a UMC of €1/day, €0.64/day and €0.21/day is taken as $Cm^k_{a,a+1}$ for individual C, D and E, respectively.

- (C) **Euclidean distances ($D_{a,a+1}$):** The Euclidean distances are first measured for each couple of subsequent fixed activities and then summed up to obtain the total daily Euclidean distance (IDD). Each $D_{a,a+1}$ value is obtained via GIS and then merged for each individual.

Table 5 presents the values of the components discussed above for the individuals A–E. Based on these components, Figure 4 displays the STA and STTP results. Figure 4 (left side) focuses on STA by showing the extension of the DPPA and the related FOS for each individual. For individuals B–E (who have a wide free-time span available for discretionary activities either in the morning or afternoon), results are divided into two clusters. The first includes the DPPA and FOS resulting from the time available between fixed activities occurring in different locations ($DPPA_{fa}$ and FOS_{fa}). The second includes the additional DPPA and FOS deriving from the free-time span available between consecutive mandatory home stays ($DPPA_{hs}$ and FOS_{hs}). Figure 4 (right side) shows the results of STTP. Each individual shows the $ITC^k_{a,a+1}$ and $D_{a,a+1}$ segments on the map, plus the overall IDTC and IDD figures.

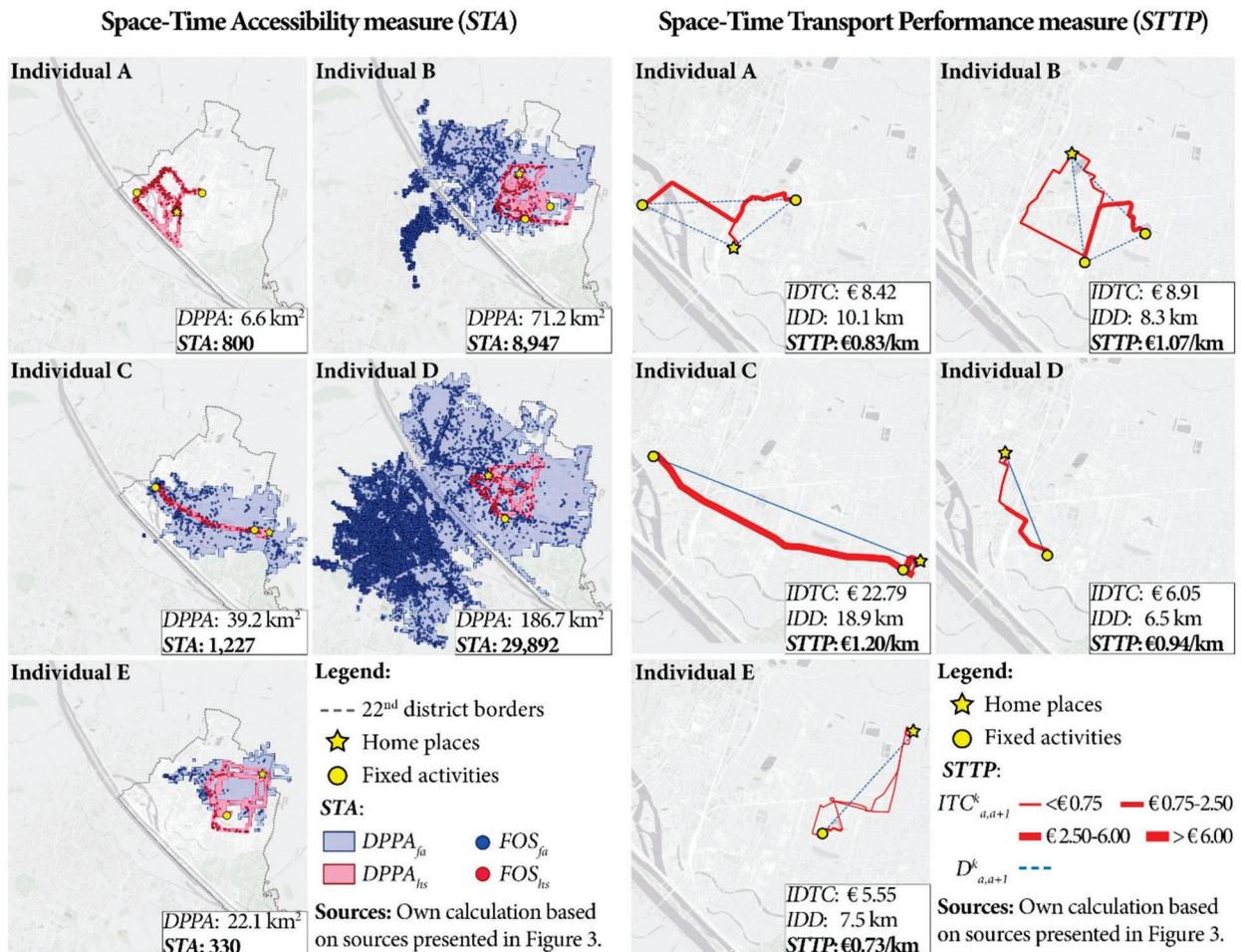


Figure 4. Left, results of the STA computation; right, results of the STTP computation.

Table 5. Components for the calculation of *STA* and *STTP* for individuals A–E.

Individuals	STPA			STA					STTP			
	Subsequent Fixed Activities	Time Span *	Mode(s) of Transport	$tt_{a,a+1}$	PPA	$O_w \in PPA$	VTT	UMC	$ICT^k_{a,a+1}$			$D_{a,a+1}$
	-	Min.	-	Min.	km ²	Num.	€/h	€/km	$Cm^k_{a,a+1}$	$Civtt^k_{a,a+1}$	$Covtt^k_{a,a+1}$	km
A	$a_1 \rightarrow a_2$	15	Car driver	5	0.88	119	9.73	0.42	1.42	0.74	0.00	2.30
	$a_2 \rightarrow a_3$	15	Car driver	7	0 [†]	0 [†]	9.73	0.42	2.41	1.18	0.00	4.64
	$a_3 \rightarrow a_4$	30	Car driver	5	6.32	779	9.73	0.42	1.81	0.86	0.00	3.16
B	$a_1 \rightarrow a_2$	30	Car driver	9	4.08	82	8.27	0.42	2.12	1.21	0.00	3.16
	$a_2 \rightarrow a_3$	30	Car driver	6	3.64	104	8.27	0.42	1.68	0.94	0.00	1.98
	$a_3 \rightarrow a_4$	30	Car driver	7	6.24	357	8.27	0.42	2.12	0.84	0.00	3.17
	$a_4 \rightarrow a_5^{**}$	150	PT, Walking	45	70.56	8924	-	-	-	-	-	-
C	$a_1 \rightarrow a_2$	20	Walking	9	0.32	1	13.78	0.00	0.00	0.00	2.09	0.52
	$a_2 \rightarrow a_3$	30	Walking	9	0.32	1	13.78	0.00	0.00	0.00	2.09	0.52
	$a_3 \rightarrow a_4^{**}$	150	PT, Walking	45	39.16	1227	-	-	-	-	-	-
	$a_4 \rightarrow a_5$	15	Walking	9	0.32	1	13.78	0.00	0.00	0.00	2.09	0.52
	$a_5 \rightarrow a_6$	15	Walking	9	0.32	1	13.78	0.00	0.00	0.00	2.09	0.52
	$a_6 \rightarrow a_7$	60	PT	43	2.32	373	6.20	1.00 ^{***}	0.50	3.20	2.85	8.45
	$a_7 \rightarrow a_8$	60	PT	39	2.32	373	6.20	1.00 ^{***}	0.50	3.00	2.36	8.45
D	$a_1 \rightarrow a_2$	60	PT	27	3.52	259	4.00	0.64 ^{***}	0.32	1.07	1.60	3.23
	$a_2 \rightarrow a_3$	90	PT	29	6.92	616	4.00	0.64 ^{***}	0.32	1.27	1.47	3.23
	$a_3 \rightarrow a_4^{**}$	180	PT, Walking	60	186.68	29,892	-	-	-	-	-	-
E	$a_1 \rightarrow a_2$	30	Car passenger	9	7.44	101	4.45	0.42	0.00	0.70	0.00	3.78
	$a_2 \rightarrow a_3$	60	PT	34	1.84	30	4.45	0.21 ^{***}	0.21	0.74	3.90	3.78
	$a_3 \rightarrow a_4^{**}$	120	PT, Walking	30	20.16	279	-	-	-	-	-	-

Notes: * The time span between the ending time of a fixed activity and the starting time of the following one. ** Subsequent fixed activities occurring at the same location (i.e., home place) with a free-time span in between. Travels that could be potentially performed between them are considered in the *STA* computation, but not in the *STTP* one. *** For the values regarding PT, the unit of measure is €/day and not €/km. † PPA and O_w are equal to 0 because there is less than 10 min at disposal to engage in discretionary activities.

4.3. Discussion of Results

The results of *STA* and *STTP* are summarised in Table 6. Since they are expressed in different measurement units, they are also converted in percentages (*rSTA* and *rSTTP*). The highest *STA* and *STTP* figures across the five individuals are 100.00%, while the other values are rescaled accordingly. The coefficient of variation (CV) is measured for both *STA* and *STTP* to point out their differences in terms of distribution across individuals. To better understand these differences in distribution, we take into account a set of time, space and transport variables, which have a relevant influence on *STA* and *STTP* (reported in Table 6). These include: the amount of constrained time in a day, i.e., the number of hours/day spent in fixed activities (*CT*); the density of discretionary opportunities available within the *DPPA* of each individual (*OD*); the average speed of the travels linking the subsequent fixed activities of each individual (*AS*); the average kilometric monetary cost (considering the travelled distance) incurred by individuals to travel among their fixed activities (*AKC*); the average travel-time cost to perform daily travels by all modes (*ATC*); the detour effect experienced by each individual expressed as the ratio between the travelled distance and the Euclidean distance (*DE*); and finally the preferred mode(s) of transport used by the individuals for their fixed activities (*PMT*). The first two variables are particularly relevant to explain *STA* results, while the others have a higher influence on *STTP*. The following two

paragraphs first discuss the overall accessibility differences registered by *STA* and *STTP*, and then explore the main reasons for these differences for each individual.

Table 6. Results of *STA* and *STTP* for individuals A–E.

Individuals	Time, Space and Transport Variables							STA			STTP		
	CT	OD	AS	AKC	ATC	DE	PMT	STA	rSTA	CV	STTP	rSTTP	CV
	h/day	n./km ²	km/h	€/km	€/min	%	Preferred mode(s) of transport	-	%	%	€/km	%	%
A	23	120.4	47	0.42	0.16	133	Car driver	800	2.68		0.83	88.09	
B	20	125.7	42	0.42	0.14	182	Car driver	8947	29.93		1.07	68.48	
C	18	31.3	11	0.05	0.18	114	PT, Walking	1227	4.10	153	1.20	61.21	19
D	18.5	160.1	9	0.08	0.10	127	PT	29,892	100.00		0.94	78.46	
E	20	14.9	17	0.02	0.12	161	Car passenger, PT	330	1.10		0.73	100.00	

Overall accessibility differences: Accessibility differences measured by *STA* are sensibly higher than those registered by *STTP* (CV equal to 153% and 19%, respectively). This is consistent with the different approaches of the two measures. The primary purpose of *STTP* is to isolate the individual accessibility differences that are directly related to the performance of the transport system by excluding other variables not affected by changes in the transport system. Accordingly, *STTP* varies depending on the travel-time and monetary costs incurred to travel between fixed activities (*ATC* and *AKC*). Moreover, it depends on the level of the detour (*DE*), since travel costs are weighed according to the Euclidean (and not travelled) distance among fixed activities. Accordingly, the individuals with the lowest *rSTTP* (B and C) register the highest average time cost (*ATC* = €0.18/min for individual C), monetary cost (*AKC* = €0.42/km for Individual B), detour effect (*DE* = 182% for individual B), and one of the lowest average speeds between fixed activities (*AS* = km11/h for individual C). In contrast, *STA* considers a wider range of factors, including the availability and distribution of discretionary opportunities across space and the individual amount of constrained and free time on a daily basis. These two factors play a crucial role in determining the higher differences registered by *STA*. Indeed, the three individuals scoring the lowest *STA* (E, A and C) also have the highest amount of daily constrained time (*CT* equal to h23/day and h20/day for individuals A and E); and the lowest density of discretionary opportunities within their *DPPAs* (*OD* equal to 14.9/km² and 31.3/km² for individuals E and C).

Accessibility differences at the individual level:

- Individual A has the second highest *rSTTP* value (88.09) and the second lowest *rSTA* value (2.68). This is related to the lack of time to engage in discretionary activities (h23/day are constrained) that decreases *STA*; it also depends on the high average speed of private car (*AS* = km47/h, the highest value registered) and the moderate detour effect (*DE* = 133%) that increase *STTP*.
- In contrast, Individual B has the second highest *rSTA* (29.93) and the second lowest *rSTTP* (68.48). The high density of discretionary opportunities at disposal (*OD* = 125.7/km²) affects *STA* positively, as well as their possibility to reach the city centre of Vienna during their free-time span in the afternoon (see Figure 4). Conversely, the high detour effect experienced during mandatory travels (*DE* = 182%) and the monetary cost (*AKC* = €0.42/km) affects *STTP* negatively.
- Compared with the other individuals, Individual C registers low values for both *rSTA* and *rSTTP* (4.10 and 61.62). On the one hand, this depends on the lack of discretionary opportunities in their *DPPA* (*OD* = 31.3/km²), which decreases *STA*. On the other hand, this condition is influenced by a low average speed of mandatory travel and the high average travel-time cost associated with them (*AS* = km11/h; *ATC* = €0.18/min), which affect *STTP*. In particular, it is worth mentioning that Individual C spends almost 1 h each day in out-of-vehicle travel time (by walking, waiting and transferring), which

corresponds to about 50% of their daily travel time. This has great impacts on their *AS* and *ATC*.

- Contrary to Individual C, Individual D registers high or average values for both *STA* and *STTP*. This is linked to various factors. As for *STA*, (s)he has the second lowest amount of constrained time ($CT = h18.5/\text{day}$) and the highest density of opportunity at disposal ($OD = 160.1$). These two conditions are interlinked: thanks to the amount of free time and proximity of the home location to a stop of the subway line U1, Individual D could reach the central district of Vienna (see Figure 4). As for *STTP*, Individual D has the lowest average speed among individuals ($AS = km9/h$). However, this negative factor is offset by the low travel-time and monetary cost paid for their daily travels ($AKC = €0.08/\text{km}$; $ATC = €0.10/\text{min}$). Indeed, Individual D is a pensioner and their time has the lowest value among the observed individuals (see Table 4), meaning that the low speed has a smaller impact on his *STTP*. At the same time, (s)he benefits from the convenient PT subscriptions offered from the City of Vienna to people aged +65.
- Finally, Individual E registers the lowest *STA* and highest *STTP*. In this case, *STA* is most negatively influenced by their home location, which is in a mainly non-urbanised area. As a consequence, the density of discretionary opportunities is the lowest registered ($OD = 14.9/\text{km}^2$). Additionally, Individual E has free time only during the evening (from 18:00 to 20:00; see Table 3), when the PT provision is least competitive. In contrast, when travelling to their fixed activities, Individual E may take advantage of a ride from their parents in the morning and make use of PT in the afternoon. This makes their average speed higher than those registered by the other individuals travelling by PT ($AS = km17/h$), while keeping their kilometric cost low ($AKC = €0.02/\text{km}$). *AKC* is also influenced by the attractive annual subscription offered by the Wiener Linien to students up to 24 years old, which is 78% lower than the standard subscription for adults (€79/year vs. €365/year). Finally, even the low cost of travel time ($ATC = €0.12/\text{min}$) plays a positive role in *STTP*.

4.4. Added Value of *STTP* for *STA* and Its Limits

According to the developed methodology and test, the main added value of *STTP* for *STA* consists in its capacity to focus explicitly (and exclusively) on the performance of the transport system. Although this allows *STTP* to point out the accessibility differences that are strictly linked to the transport system, it also makes individual differences less evident (as demonstrated by the test results). Indeed, other aspects such as the daily amount of free time, the distance daily travelled, and the availability of discretionary opportunities are excluded from the computation. This confirms that *STTP* should be seen as a complementary (and not alternative) measure to be used together with *STA* to deal with the broad topic of accessibility equality. On the one hand, *STA* measures individual accessibility from a broad perspective by taking into account spatial, transport, individual and temporal factors and by holding the potential dimension of accessibility. On the other hand, *STTP* narrows the focus down, by isolating the performances of the transport system in allowing individuals to carry out their schedule of daily fixed activities.

Beside this key added value of *STTP* for *STA*, it is important to mention also some limits of *STTP* that need to be addressed in future applications. First, *STTP* results highly depend on the estimation of *VTT*, which has to be as accurate as possible and performed at the individual level. However, to estimate *VTT* at the individual level with the wage rate method, personal income data has to be collected. This is not always feasible since people tend to be not willing to answer income-related questions [53]. An alternative may be to estimate *VTT* through stated and revealed preference methods [44]. However, this approach is very data demanding and time-consuming and it requires the availability of individuals to answer a lengthy questionnaire that should include (a) preliminary socio-economic and demographic questions; (b) the travel diary for *STPA* computation; and (c) stated-preference questions to estimate *VTT*. Second, regarding the test, there are some

computational limits to be refined. First, the opening/closure time of the discretionary opportunities should be integrated to properly select those to be counted in *STA*. Second, discretionary opportunities could be weighed according to the importance assigned by individuals to different categories of opportunities such as groceries, shopping facilities, or post/bank offices. However, the inclusion of these elements in the computation highly depends on the data at disposal in the study area and the availability of individuals to answer longer questionnaires. Third, the walking speed assigned to individuals (km5/h in our test) could be differentiated in specific cases, e.g., for elderly people with physical hindrances that make them walk slower. These limits need to be addressed to highlight precisely accessibility difference issues at individual level (e.g., refs. [54–56]).

5. Conclusions

Considering that many accessibility differences are unavoidable and irremediable by transport policies [1], and that *STA* measures tend to incorporate a broad variety of factors, this study has introduced a transport performance measure isolating the role of the transport system in individual accessibility differences. As suggested by the test, the results obtained by *STTP* may be sensibly different from those of *STA*. On the one hand, individual differences tend to be smaller, according to the more narrowed and transport-specific approach of *STTP*. On the other hand, the individuals registering the highest *STA* do not necessarily correspond to those with the highest *STTP* and vice versa. Indeed, an individual might register some transport difficulties in running their daily schedule but at the same time have enough time to engage in several discretionary activities (like Individual D in Section 4). Even the opposite may apply: a person could easily reach their usual daily destinations but have little free time and few surrounding amenities so as to register a low space-time accessibility (like Individual E in Section 4).

These results highlight the complementary value of *STTP* for *STA*. Indeed, researchers and policymakers might gain relevant benefits from the combined analysis of these two measures since they may evaluate both the overall and transport-specific impacts of various transport policies on individual accessibility differences (e.g., refs. [57–59]). This would be particularly relevant to assess transport policies that are expected to trigger controversial impacts on individual accessibility differences. For instance, in the case of high-speed railways, autonomous vehicle applications and transport sharing services, which are often found to increase accessibility in general, but also the accessibility differences across (groups of) people (e.g., refs. [57,59,60]). At the same time, *STTP* might be used to discuss the individual accessibility impacts of growing mobility trends, such as the usage of individual slow mobility solutions (such as e-scooters), which are progressively replacing walking travels, especially for the first- and last-mile.

To prove the suitability of the combination of *STA* and *STTP*, it is necessary to refine the estimation of user transport costs on the one hand and extend the application of this methodology to a broader case study on the other. The former challenge is mainly related to reliable estimation of the value of travel time (*VTT*). This study relied on the wage rate method, given its popularity and ease of application. However, more articulated estimation approaches could be used (such as stated and revealed preferences) to differentiate better *VTT* values depending on, e.g., the travel purposes, length of the travel and transport modes [61,62]. The latter challenge implies collecting *STPA* information from a substantial sample of individuals and applying sound statistical distribution analyses to the results. The first aspect is problematic for many studies using space-time measures, and it is one of the main practical reasons for the low usage of this kind of measure compared with, e.g., gravity-based measurements [15]. To overcome this limit, recent studies have deployed multi-stage stratified random sampling approaches, reducing the survey sampling rates while maintaining high accuracy [12]. The usage of statistical distribution analyses is problematic for a conceptual reason, since they imply the focus on one or more equity typologies [63]. For instance, calculating the coefficient of variation (*CV*) over the whole sample is closely linked to an egalitarian point of view, since the *CV* measures how spread

the distribution of a good over a population is. Conversely, calculating the percentile ratio (PR) by comparing the accessibility values scored at the median (50%) with those scored at e.g., 10% is linked to the vertical-equity point of view, since it focuses on the individuals with the lowest accessibility values. To guarantee a scientifically solid distributional analysis, future applications of *STTP* and *STA* to a broader case study would require a methodological effort in selecting a broad set of statistical distribution analyses to apply to the results.

Despite these challenges, *STTP* provides an added value for *STA*, and it may be deployed to complement the analysis of individual accessibility and of the distributional implications of transport policies, which represent an increasingly relevant priority of policymakers and transport planners.

Author Contributions: Conceptualization, A.D., G.H.; methodology, A.D., G.H.; software, M.G.; validation, A.D., M.G.; formal analysis, A.D., M.G.; investigation, A.D.; resources, A.D.; data curation, M.G.; writing—original Draft Preparation, A.D.; writing—review and editing, A.D., M.G., G.H.; visualization, A.D., M.G.; supervision, G.H.; project administration, G.H.; funding acquisition, A.D., M.G., G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in whole or in part by the Austrian Science Fund (FWF) [I 5224-G Internationale Projekte]. For Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. The research leading to these results has received funding from the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano, within the framework of the FWF Joint Projects between Austria and South Tyrol, under Project number I 5224-G. The APC was funded by the Austrian Science Fund (FWF) [I 5224-G Internationale Projekte].

Data Availability Statement: Data used and developed in this study will be uploaded to an institutional repository in case of approval. In that case, details for data access will be provided in this section.

Acknowledgments: Open Access Funding by the Austrian Science Fund (FWF). The authors warmly thank Elisa Ravazzoli (Eurac Research, Institute for Regional Development) for her support to the conceptualisation and revision of the article and her contribution within the framework of the RAAV research project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van Wee, B.; Mouter, N. Evaluating Transport Equity. In *Advances in Transport Policy and Planning*; Academic Press: Cambridge, MA, USA, 2020.
2. Young, W.; Tilley, F. Can Businesses Move beyond Efficiency? The Shift toward Effectiveness and Equity in the Corporate Sustainability Debate. *Bus. Strategy Environ.* **2006**, *15*, 402–415. [[CrossRef](#)]
3. Lucas, K. Transport and Social Exclusion: Where Are We Now? *Transp. Pol.* **2012**, *20*, 105–113. [[CrossRef](#)]
4. UN THE 17 GOALS | Sustainable Development. Available online: <https://sdgs.un.org/goals> (accessed on 10 December 2021).
5. Van Wee, B.; Geurs, K. Discussing Equity and Social Exclusion in Accessibility Evaluations. *Eur. J. Transp. Infrastruct. Res.* **2011**, *11*. [[CrossRef](#)]
6. Camporeale, R.; Caggiani, L.; Ottomanelli, M. Modeling Horizontal and Vertical Equity in the Public Transport Design Problem: A Case Study. *Transp. Res. Part A Policy Pract.* **2019**, *125*, 184–206. [[CrossRef](#)]
7. Banister, D. The Sustainable Mobility Paradigm. *Transp. Pol.* **2008**, *15*, 73–80. [[CrossRef](#)]
8. Van Wee, B. Accessible Accessibility Research Challenges. *J. Transp. Geogr.* **2016**, *51*, 9–16. [[CrossRef](#)]
9. Handy, S.L.; Niemeier, D.A. Measuring Accessibility: An Exploration of Issues and Alternatives. *Environ. Plan. Econ. Space* **1997**, *29*, 1175–1194. [[CrossRef](#)]
10. Franssen, K.; Farber, S. 4-Using Person-Based Accessibility Measures to Assess the Equity of Transport Systems. In *Measuring Transport Equity*; Lucas, K., Martens, K., Di Ciommo, F., Dupont-Kieffer, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 57–72; ISBN 978-0-12-814818-1.
11. Kwan, M.-P. Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-Based Framework. *Geogr. Anal.* **1998**, *30*, 191–216. [[CrossRef](#)]
12. Chen, Z.; Yeh, A.G.-O. Socioeconomic Variations and Disparity in Space–Time Accessibility in Suburban China: A Case Study of Guangzhou. *Urban Stud.* **2021**, *58*, 750–768. [[CrossRef](#)]

13. Kwan, M.-P. Gender and Individual Access to Urban Opportunities: A Study Using Space–Time Measures. *Prof. Geogr.* **1999**, *51*, 210–227. [\[CrossRef\]](#)
14. Miller, H.J. Measuring Space-Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures. *Geogr. Anal.* **1999**, *31*, 187–212. [\[CrossRef\]](#)
15. Kwan, M.-P. Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 1078–1086. [\[CrossRef\]](#)
16. Song, S. Some Tests of Alternative Accessibility Measures: A Population Density Approach. *Land Econ.* **1996**, *72*, 474–482. [\[CrossRef\]](#)
17. Geurs, K.T.; van Wee, B. Accessibility Evaluation of Land-Use and Transport Strategies: Review and Research Directions. *J. Transp. Geogr.* **2004**, *12*, 127–140. [\[CrossRef\]](#)
18. Koenig, J.G. Indicators of Urban Accessibility: Theory and Application. *Transportation* **1980**, *9*, 145–172. [\[CrossRef\]](#)
19. Kolcsár, R.A.; Szilassi, P. Assessing Accessibility of Urban Green Spaces Based on Isochrone Maps and Street Resolution Population Data through the Example of Zalaegerszeg, Hungary. *Carpat. J. Earth Environ. Sci.* **2018**, *13*, 31–36. [\[CrossRef\]](#)
20. O’Sullivan, D.; Morrison, A.; Shearer, J. Using Desktop GIS for the Investigation of Accessibility by Public Transport: An Isochrone Approach. *Int. J. Geogr. Inf. Sci.* **2000**, *14*, 85–104. [\[CrossRef\]](#)
21. Wu, H.; Levinson, D.; Owen, A. Commute Mode Share and Access to Jobs across US Metropolitan Areas. *Environ. Plan. Urban Anal. City Sci.* **2021**, *48*, 671–684. [\[CrossRef\]](#)
22. Bocarejo, J.P.; Escobar, D.; Hernandez, D.O.; Galarza, D. Accessibility Analysis of the Integrated Transit System of Bogotá. *Int. J. Sustain. Transp.* **2016**, *10*, 308–320. [\[CrossRef\]](#)
23. Cavallaro, F.; Dianin, A. An Innovative Model to Estimate the Accessibility of a Destination by Public Transport. *Transp. Res. Part Transp. Environ.* **2020**, *80*, 102256. [\[CrossRef\]](#)
24. Gutiérrez, J.; Monzón, A.; Piñero, J.M. Accessibility, Network Efficiency, and Transport Infrastructure Planning. *Environ. Plan. Econ. Space* **1998**, *30*, 1337–1350. [\[CrossRef\]](#)
25. Pan, Q.; Jin, Z.; Liu, X. Measuring the Effects of Job Competition and Matching on Employment Accessibility. *Transp. Res. Part Transp. Environ.* **2020**, *87*. [\[CrossRef\]](#)
26. Van Wee, B.; Hagoort, M.; Annema, J.A. Accessibility Measures with Competition. *J. Transp. Geogr.* **2001**, *9*, 199–208. [\[CrossRef\]](#)
27. Wilson, A.G. A Family of Spatial Interaction Models, and Associated Developments. *Environ. Plan. Econ. Space* **1971**, *3*, 1–32. [\[CrossRef\]](#)
28. Pirie, G.H. Measuring Accessibility: A Review and Proposal. *Environ. Plan. Econ. Space* **1979**, *11*, 299–312. [\[CrossRef\]](#)
29. Lu, R.; Chorus, C.G.; van Wee, B. The Effects of Different Forms of ICT on Accessibility – a Behavioural Model and Numerical Examples. *Transp. Transp. Sci.* **2014**, *10*, 233–254. [\[CrossRef\]](#)
30. Nassir, N.; Hickman, M.; Malekzadeh, A.; Irannezhad, E. A Utility-Based Travel Impedance Measure for Public Transit Network Accessibility. *Transp. Res. Part A Pol. Pract.* **2016**, *88*, 26–39. [\[CrossRef\]](#)
31. Guy, C.M. The Assessment of Access to Local Shopping Opportunities: A Comparison of Accessibility Measures. *Environ. Plan. B Plan. Des.* **1983**, *10*, 219–237. [\[CrossRef\]](#)
32. Cohn, J.; Ezike, R.; Martin, J.; Donkor, K.; Ridgway, M.; Balding, M. Examining the Equity Impacts of Autonomous Vehicles: A Travel Demand Model Approach. *Transp. Res. Rec.* **2019**, *2673*, 23–35. [\[CrossRef\]](#)
33. Wang, F. Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 1104–1112. [\[CrossRef\]](#)
34. O’Kelly, M.E.; Miller, E.J. Characteristics of Multistop Multipurpose Travel: An Empirical Study of Trip Length. *Transp. Res. Rec.* **1984**, *976*, 33–39.
35. Delafontaine, M.; Neutens, T.; Schwanen, T.; de Weghe, N.V. The Impact of Opening Hours on the Equity of Individual Space–Time Accessibility. *Comput. Environ. Urban Syst.* **2011**, *35*, 276–288. [\[CrossRef\]](#)
36. Lee, J.; Miller, H.J. Analyzing Collective Accessibility Using Average Space-Time Prisms. *Transp. Res. Part Transp. Environ.* **2019**, *69*, 250–264. [\[CrossRef\]](#)
37. Hägerstrand, T. What about People in Regional Science? *Pap. Reg. Sci. Assoc.* **1970**, *24*, 6–21. [\[CrossRef\]](#)
38. Landau, U.; Prashker, J.N.; Alpern, B. Evaluation of Activity Constrained Choice Sets to Shopping Destination Choice Modelling. *Transp. Res. Part Gen.* **1982**, *16*, 199–207. [\[CrossRef\]](#)
39. Chen, B.Y.; Wang, Y.; Wang, D.; Lam, W.H.K. Understanding Travel Time Uncertainty Impacts on the Equity of Individual Accessibility. *Transp. Res. Part Transp. Environ.* **2019**, *75*, 156–169. [\[CrossRef\]](#)
40. Pritchard, J.P.; Tomasiello, D.; Giannotti, M.; Geurs, K. An International Comparison of Equity in Accessibility to Jobs: London, São Paulo and the Randstad. *Findings* **2019**, 7412. [\[CrossRef\]](#)
41. Pritchard, J.P.; Tomasiello, D.B.; Giannotti, M.; Geurs, K. Potential Impacts of Bike-and-Ride on Job Accessibility and Spatial Equity in São Paulo, Brazil. *Transp. Res. A Part Pol. Pract.* **2019**, *121*, 386–400. [\[CrossRef\]](#)
42. Ricci, S. *Tecnica ed Economia dei Trasporti*; Hoepli: Milano, Italy, 2011; ISBN 978-88-203-4594-5.
43. Jeekel, J.F.; Martens, C.J.C.M. Equity in Transport: Learning from the Policy Domains of Housing, Health Care and Education. *Eur. Transp. Res. Rev.* **2017**, *9*, 53. [\[CrossRef\]](#)
44. Sinha, K.C.; Labi, S. *Transportation Decision Making*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007; ISBN 978-0-470-16807-3.

45. Bunge, W. *Theoretical Geography*; Royal University of Lund Sweden, Department of Geography: Lund, Sweden; Gleerup Publishers: Lund, Sweden, 1962.
46. Schwanen, T. Struggling with Time: Investigating Coupling Constraints. *Transp. Rev.* **2008**, *28*, 337–356. [CrossRef]
47. GIP.at Graphenintegrations-Plattform GIP: Der Multimodale, Digitale Verkehrsgraph Für Ganz Österreich. Available online: <http://gip.gv.at/> (accessed on 26 January 2022).
48. Open Data Österreich Wiener Linien–Fahrplandaten GTFS Wien. Available online: <https://www.data.gv.at/katalog/dataset/wiener-linien-fahrplandaten-gtfs-wien> (accessed on 26 January 2022).
49. Statistik Austria Verdienststruktur-Ergebnisse Im Überblick: Bruttostundenverdienste. Available online: https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/soziales/personen-einkommen/verdienststruktur/index.html (accessed on 20 January 2022).
50. Infrac. *COMPETE: Analysis of the Contribution of Transport Policies to the Competitiveness of the EU Economy and Comparison with the United States. Annex 1 to COMPETE Final Report: Analysis of Operating Cost in the EU and the US*; European Commission–DG TRE: Geneva, Switzerland, 2006.
51. ACEA. *ACEA 2021 Tax Guide*; ACEA European Automobile Manufacturers’ Association: Brussels, Belgium, 2021.
52. Wiener Linien-Tickets. Available online: <http://www.wienerlinien.at/tickets> (accessed on 21 January 2022).
53. Gideon, L. *Handbook of Survey Methodology for the Social Sciences*; Springer: New York, NY, USA, 2012; ISBN 978-1-4614-3876-2.
54. Noack, E. Are Rural Women Mobility Deprived?—A Case Study from Scotland. *Sociol. Rural.* **2011**, *51*, 79–97. [CrossRef]
55. Ranković Plazinić, B.; Jović, J. Mobility and Transport Potential of Elderly in Differently Accessible Rural Areas. *J. Transp. Geogr.* **2018**, *68*, 169–180. [CrossRef]
56. Rau, H.; Vega, A. Spatial (Im)Mobility and Accessibility in Ireland: Implications for Transport Policy. *Growth Chang.* **2012**, *43*, 667–696. [CrossRef]
57. Dianin, A.; Ravazzoli, E.; Hauger, G. Implications of Autonomous Vehicles for Accessibility and Transport Equity: A Framework Based on Literature. *Sustainability* **2021**, *13*, 4448. [CrossRef]
58. Velaga, N.R.; Beecroft, M.; Nelson, J.D.; Corsar, D.; Edwards, P. Transport Poverty Meets the Digital Divide: Accessibility and Connectivity in Rural Communities. *J. Transp. Geogr.* **2012**, *21*, 102–112. [CrossRef]
59. Kim, H.; Sultana, S. The Impacts of High-Speed Rail Extensions on Accessibility and Spatial Equity Changes in South Korea from 2004 to 2018. *J. Transp. Geogr.* **2015**, *45*, 48–61. [CrossRef]
60. Braun, L.M.; Rodriguez, D.A.; Gordon-Larsen, P. Social (in)Equity in Access to Cycling Infrastructure: Cross-Sectional Associations between Bike Lanes and Area-Level Sociodemographic Characteristics in 22 Large U.S. Cities. *J. Transp. Geogr.* **2019**, *80*, 102544. [CrossRef]
61. Chen, X.; Liu, Q.; Du, G. Estimation of Travel Time Values for Urban Public Transport Passengers Based on SP Survey. *J. Transp. Syst. Eng. Inf. Technol.* **2011**, *11*, 77–84. [CrossRef]
62. Hensher, D.A. Measurement of the Valuation of Travel Time Savings. *J. Transp. Econ. Pol.* **2001**, *35*, 71–98.
63. Thomopoulos, N.; Grant-Muller, S.; Tight, M.R. Incorporating Equity Considerations in Transport Infrastructure Evaluation: Current Practice and a Proposed Methodology. *Eval. Progr. Plan.* **2009**, *32*, 351–359. [CrossRef]

Article

Analytical Model for Enhancing the Adoptability of Continuous Descent Approach at Airports

Emad A. Alharbi ^{1,*}, Layek L. Abdel-Malek ², R. John Milne ³ and Arwa M. Wali ⁴¹ Saab AB, Surveillance Business Area, Riyadh 11322, Saudi Arabia² Department of Mechanical and Industrial Engineering, Newark College of Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA; malek@njit.edu³ David D. Reh School of Business, Clarkson University, Potsdam, NY 13699, USA; jmilne@clarkson.edu⁴ Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; amwali@kau.edu.sa

* Correspondence: emad.alharbi@saabgroup.com; Tel.: +966-11-205-7976

Featured Application: We present an analytical model, using a queueing theory framework, that identifies periods of time for air traffic controllers when they can permit the vast majority of approaching aircrafts to land using Continuous Descent Approach, thereby reducing noise, fuel consumption, and pollution, while enhancing air transportation sustainability.

Abstract: Continuous Descent Approach (CDA) is the flight technique for aircraft to continuously descend from cruise altitude with an idle thrust setting and without level-offs, contrary to the staircase-like Step-down Descent Approach (SDA). Important for air transportation sustainability, using CDA reduces noise, fuel consumption, and pollution. Nevertheless, CDA has been limited to low traffic levels at airports, often at night, because it requires more separation distance between aircraft arrivals and, thus, could decrease throughput. Insufficient attention has been given to helping air traffic controllers decide when CDA may be used. In this paper, we calculate the probability that an aircraft arriving during a particular brief period of time (e.g., 15 min) will need to revert to SDA when the controller tentatively plans to permit CDA for all aircrafts arriving during that time period. If this probability is low enough, the controller may plan to permit CDA during that time period. We utilize an analytical approach and queueing theory framework that considers factors such traffic and weather conditions to estimate the probability. We also provide the number of aircrafts that can be accommodated within the airport's stacking space using CDA. This number provides insight into whether a particular aircraft may use CDA.

Keywords: green transport; continuous descent approach; optimized profile descent; climate change; terminal maneuvering area; environmental impact; applied queueing theory; air traffic management; air transportation sustainability

Citation: Alharbi, E.A.; Abdel-Malek, L.L.; Milne, R.J.; Wali, A.M.

Analytical Model for Enhancing the Adoptability of Continuous Descent Approach at Airports. *Appl. Sci.* **2022**, *12*, 1506. <https://doi.org/10.3390/app12031506>

Academic Editor:
Giovanni Randazzo

Received: 31 December 2021

Accepted: 26 January 2022

Published: 30 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The air transportation and aviation industry face several challenges due to projected increases in demand for air travel and freight accompanying limited airspace congestion and airport capacity. The International Air Transport Association expects 7.2 billion passengers to travel in 2035, almost doubling the 3.8 billion air travelers in 2016, with the U.S. as the second-fastest-growing market, after China, with 484 million additional passengers per year forecasted for a total of 1.1 billion passengers [1]. With increased pressure on the infrastructure of terminals, runways, airspace around airports, and air traffic control operations, the industry is struggling to cope with this demand, yet it has to limit the harm that aircrafts cause to the environment through carbon emissions and noise levels.

With regard to aircraft emissions, the U.S. Environmental Protection Agency finalized its determination that greenhouse gas emissions from certain types of aircraft engines,

primarily engines used on large commercial jets, contribute to the pollution that causes climate change and endangers Americans' health and the environment [2]. Other countries are taking strict measures to limit emissions from aviation operations at airports by setting penalties for emissions levels above a specified limit. Under the European Union Emission and Trading System, all airlines operating in Europe, European and non-European alike, are required to monitor, report, and verify their emissions and to surrender allowances against those emissions that exceed certain levels from their flights per year [3]. Aircraft noise, on the other hand, is the biggest concern for airport officials at 29 of the 50 busiest U.S. airports [4]. Airports' support personnel who work in proximity to aircrafts idling on the ground or taking off and landing may suffer hearing loss. Residents of communities surrounding airports suffer sleep disorders and interference with speech, both of which may lead to reduced productivity in learning and work. Furthermore, recent studies have linked noise to non-auditory health effects, such as hypertension, heart disease, and stroke [5]. These issues represent critical challenges to air transportation and aviation industry sustainability, development, and prosperity.

The Continuous Descent Approach (CDA), also known as Optimized Profile Descent (OPD), is an advanced flight technique for aircrafts to descend continuously from cruising altitude to the Final Approach Fix (FAF) or touchdown without level-offs and with an idle, or near idle, thrust setting. Descending using the CDA procedure, an aircraft can stay as high as possible for a longer time than with a conventional descent, thereby expanding the vertical distance between the aircraft's sources of noise and the ground and thus, significantly reducing the noise levels for populated areas near airports. Furthermore, by descending with an idle, or near idle, engine setting, fuel burn is decreased, resulting in the reduction of fuel consumption and harmful emissions to the environment. A study that conducted flight trials of CDA at Kentucky's Louisville International Airport using an aircraft fleet of the United Parcel Service (UPS), an express package delivery company, quantified the benefits of CDA in terms of fuel savings by 400 lb. to 500 lb. per flight and noise level by 3.9 A-weighted decibels (dBA) [6]. Another study conducted at San Francisco International Airport estimated a reduction of CO₂ emissions of between 700 lb. and 10,000 lb. per flight with CDA flights [7]. To cut down on aircraft emissions, airplane manufacturer Airbus recently has been working on unique ways used by birds and emerging concepts like tandem flying that could reduce fuel burn by up to 10% [8]. When compared with the widely used Step-down Descent Approach (SDA), in which the arrival aircraft descends in a step-like fashion, CDA reduces flight time by around two minutes [9]. FedEx, another express transport and delivery company with one of the largest civil aircraft fleets in the world, has been using CDA at their World Hub, Memphis International Airport. Their use of CDA at Memphis reduced flight time by 2.5 min for each flight, and this translated into cost savings of \$105 million based on their field study from 2006 to 2009 [10].

These operational, economical, and environmental benefits from CDA procedures have made it a cornerstone in some aviation modernization programs at the national (e.g., FAA's Next Generation Air Transportation System, "NextGen"), continental (e.g., EU's Single European Sky Air Traffic Management Research, "SESAR"), and international (e.g., United Nations' International Civil Aviation Organization, "ICAO", Continuous Descent Operations, "CDO", initiative) levels. Although considered as an effective Noise Abatement Procedure, CDA is not widely implemented, especially during high density operations [11,12]. Due to safety considerations [13–15], CDA procedures may require more separation between aircraft arrivals, which may affect the airport arrival rate and runway throughput [16,17]. The larger separation spacing for a CDA aircraft is mainly due to two reasons: the difficulty for air traffic controllers to predict the future position of an aircraft with significantly variable speed [6] and the inability of the pilot to quickly decelerate during descent [18]. Although CDA has been proven to be feasible and without increasing the required spacing between aircraft under light traffic conditions, such as night-time operations [6], aircrafts flying CDA are most likely to be spaced further apart under heavy traffic conditions.

Thus, CDA implementation has been limited to low to moderate traffic levels. During these low traffic conditions, CDA has been used at more airports. To increase the use of CDA, several studies in the literature have used various approaches, such as simulation [19], mathematical modeling [15,20], and flight trials [6,14], to quantify CDA's benefits and/or suggest solutions to the problem of increasing CDA's usage at airports through the analysis of sequencing and merging [21], merging and spacing [18], scheduling and conflict detection and resolution [16], time and aircraft energy management during descent [22], fuel and flight-path management [23,24], and ground-to-air air traffic network vulnerability [25]. Other literature has applied quantitative methods to improve aircraft operation [26,27]. However, insufficient attention has been given to developing a quantitative measure to enable air traffic controllers to make informed decisions on safely accepting more CDA operations.

The contribution of this work is the development of a model that addresses this gap in CDA research and that helps air traffic controllers determine brief periods of time (e.g., 15-min periods) in which the vast majority of arriving flights may land using CDA. These time periods are based on time of aircraft arrival into the TMA. An aircraft entering the TMA during one of those time periods may begin its continuous descent upon entering the TMA while completing its continuous descent during a later time period, in which then-arriving aircrafts are no longer using CDA. In fact, although the time to descend using CDA depends on several factors (e.g., aircraft weight), it may be longer than 15 min. Data from previous authors [28,29] implies the time to descend using CDA may be 20–30 min.

Special attention is dedicated to factors that have a significant impact on CDA implementation, such as airspace structure around airports, airport arrival rate, and distance requirements for longitudinal separation between approaching aircrafts. Analyzing airspace structure around an airport offers a systematic way of developing an analytical model that adequately captures the elements associated with descent and approach procedures.

In particular, we calculate the probability that an aircraft arriving during a particular brief period of time will need to revert to SDA under the initial modeling assumption that the controller will permit CDA for all aircraft arrivals during that period. If this calculated probability is low enough, the controller may be comfortable in planning to permit CDA for all arriving aircrafts during that time period, and otherwise, they will not permit any of them to use CDA. Our model utilizes an analytical approach and queueing theory framework that considers factors such traffic and weather conditions to estimate the probability. The non-queueing portion of our modelling provides the number of aircrafts that can be accommodated within the airport's stacking space when CDA is used for all aircrafts arriving during the time period. This number provides insight to the controller on whether to permit a particular aircraft to use CDA during the period when all (or nearly all) arriving aircrafts will be permitted to descend with CDA (due to the low probability an aircraft will need to use SDA instead). Through the use of this modeling, it is our hope that CDA will be used more often and, thus, reduce noise, fuel consumption, harmful emissions and, thus, provide greener and more sustainable air transportation operations. This paper should draw attention to the opportunity to systematically increase the use of CDA for aviation green operations and air transportation sustainability.

The paper is organized as follows. Section 2 describes the airspace around airports with respect to the terminal maneuvering area and describes descent and approach operations. In particular, two approach operations (CDA and SDA) are discussed and compared. In Section 3, we present the adoptability of CDA and the factors that impact aircraft landing time; the Base of Aircraft Data (BADA) Aircraft Performance Calculation (APC) is estimated and validated against actual landing times of flights operated at Nashville International Airport (BNA). Section 4 presents a background on the queueing theory and the main assumptions and fundamental components used to develop our model, while Section 5 presents the concept of the probability of an aircraft being denied CDA entry as a key output of the queueing model. In Section 6, we illustrate the calculation of the model probabilities

using standard industry data (e.g., separation distance requirements) and actual flight data from the BNA airport. Finally, Section 7 presents our main findings and conclusions.

2. Preliminaries and Process Description

We begin this section by describing the airspace around airports, then generally describe aircraft descent and approach operations at airports, and we conclude with a comparison between the two most-commonly used descent approaches: CDA and SDA.

2.1. Structure of Airspace around Airports

Terminal Maneuvering Area (TMA). TMA refers to the designated area of airspace managed by air traffic control services around major airports that have high volumes of traffic. Normally, TMA airspace is designed in a cylindrical configuration, including all altitudes centered around the geographical coordinates of the airport. Geographical positions that define the boundaries of the TMA, known as entry fixes, are considered as entry points to the TMA (although each fix includes all altitudes and thus may be conceptualized as a line), and the arriving aircraft enters the TMA airspace via entry or arrival fixes. When crossing the TMA boundary over one of these entry fixes, the responsibility for the separating aircraft will be handed over usually from a controller at the air traffic control center responsible for separating the en route aircraft (e.g., Air Route Traffic Control Center “ARTCC”) to a controller at the air traffic control center responsible for separating the aircraft approaching an airport (e.g., Terminal Radar Approach Control “TRACON”). A simplified structure of a TMA is illustrated in Figure 1.

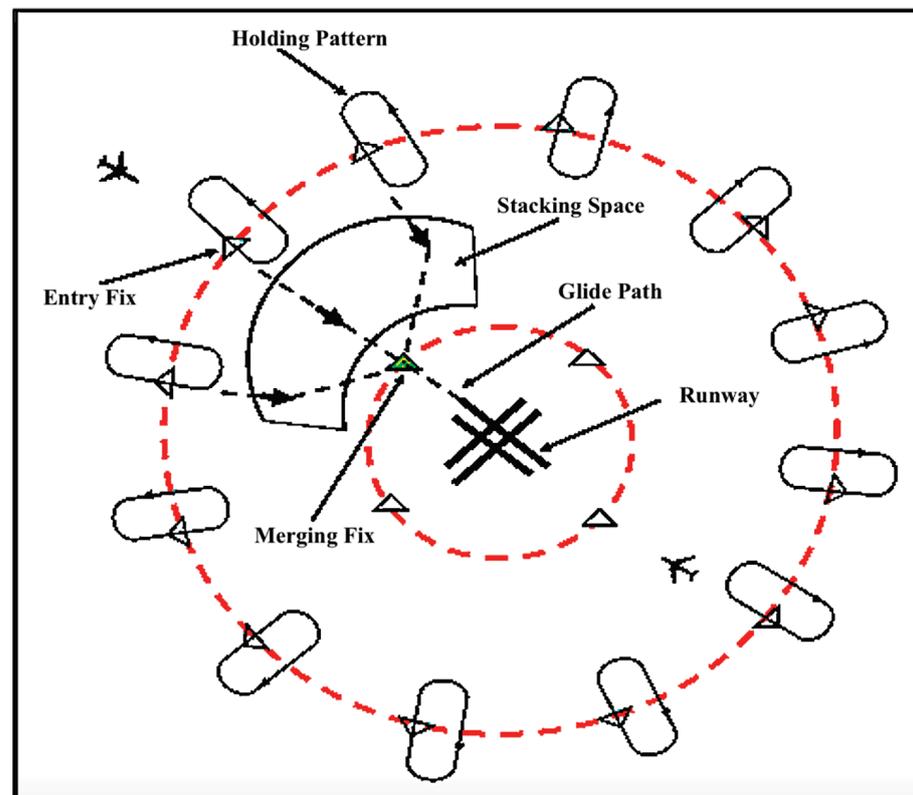


Figure 1. Typical structure of a TMA (top view).

As an arriving aircraft nears an entry fix, the air traffic controller may clear the pilot for the approach or, depending on traffic congestion and the separation and sequencing method used, may place the aircraft in a holding pattern. While aircraft separation aims to have the controller apply and maintain the separation distance requirements between

aircrafts for safety purposes, aircraft sequencing aims to have the controller organize a stream of aircrafts to provide an orderly sequence of continuous traffic flow towards the final approach path. In practice, there are a number of aircraft-sequencing methods for approach traffic management, but generally, all of them fall under two broad categories: procedural control (published procedures with altitude change and speed instructions) and radar vectoring (controller-generated instructions in terms of headings, altitude, and speeds to optimize traffic flow in order to maximize the number of aircrafts with the least average delay). Today, radar vectoring is one of the main methods to achieve efficient sequencing for aircrafts flying towards the final approach path.

Once an aircraft has been cleared by the controller to approach the airport or to leave a holding pattern, the aircraft approaches the merging fix. However, as the aircraft approaches the merging fix, it flies in the stacking space, the space that the controllers use from the available terminal airspace to stack arriving aircrafts, that is, orderly align aircraft arrivals for approach. In the stacking space, the controller manages air traffic and enhances airspace capacity by stacking arriving aircrafts using techniques such as minimal speed adjustments and path-stretching. This efficient management of air traffic flow enables the controller to bring together aircrafts that have crossed entry fixes from different directions to be stacked and merged at the merging fix. For instance, an aircraft may enter the TMA through one of about 12 entry fixes and then proceed to one of about four merger fixes as they get closer to the airport. The merging fix provides a transition for arriving aircrafts from the stacking space to approach, as it connects traffic from different directions into one merger fix to create one stream of aircraft arrivals to follow a standard published arrival procedure. This way, arrivals from several directions can be accommodated, and traffic flow is efficiently managed within a congested airspace. To safely merge arriving aircrafts, the controller synchronizes aircrafts based on joining window time on the air route leading to the merging fix considering sufficient spacing for other aircrafts to fit into the air traffic stream and while maintaining, at least, the minimum required separation between aircrafts.

2.2. Description of the Aircraft Descent and Approach Process at Airports

In this subsection, we first describe the aircraft descent operations. Then, we present SDA and finally introduce CDA and compare it with SDA.

2.2.1. Descent and Approach Operations

Aircraft descent could be initiated to attain an optimal profile from the cruise altitude all the way down to landing to minimize fuel burn, emissions, and noise exposure. However, due to Air Traffic Control (ATC) restrictions and aircraft performance limitations, an optimal descent profile may not be attained all the time. For an aircraft operating at typical cruise altitudes, descent will normally initiate at 100 to 130 nautical miles (nmi) from the destination airport. This distance varies primarily due to ATC service restrictions, aircrafts' equipment and performance capabilities, and weather conditions. The controller may issue crossing restrictions during the descent, as part of a Standard Terminal Arrival Route (STAR) or as a requirement for traffic sequencing. These crossing restrictions are generally issued to the cockpit crew in terms of altitude over a fix, and they may include a speed restriction as well [30].

A stabilized descent requires minimum control adjustments by the pilot in maintaining the planned descent path; more specifically, excessive corrections or control inputs indicate that the descent was improperly planned. Thus, planning the descent from cruise altitude is important because descending early results in more of the flight at a low altitude with increased fuel consumption and noise impacts, and starting the descent late results in problems with controlling both airspeed and descent rates later in the approach phase.

Prior to flight, pilots need to compute the fuel, time, and distance required to descend from the cruising altitude to the approach gate (an imaginary point used by the controller to provide headings (i.e., vectoring) for aircraft arrivals to the final approach course), with the objective of determining the most economical distance from the airport to begin descent.

This distance is referred to as the Top of Descent (TOD) point. The computations for the TOD point could be done manually prior to flight or automatically during flight using the Flight Management System (FMS). Conversely, in flight prior to the descent, pilots plan the descent from cruise by reviewing and verifying landing weather to include winds in their consideration, since tempestuous weather at the landing airport can cause slower descents. Furthermore, pilots need to know the cruise altitude and approach gate altitude (otherwise known as the initial approach fix (IAF) altitude), descent rate, and ground speed during descent.

Based on aircraft performance, approach constraints, aircraft weight, and weather data (such as winds, temperature, and icing conditions), the vertical component of the flight plan, which referred to as the Vertical Navigation (VNAV), is computed. Usually, the VNAV approach is computed from the TOD point down to the waypoint at which the descent ends, which is generally the runway or the Missed Approach Point. There are only two types of VNAV paths that the FMS uses: the performance path or geometric path. The performance VNAV path is computed using an idle or near-idle thrust from the TOD point to the first constrained waypoint, which is constrained by speed and/or altitude and represents a typical CDA. The geometric VNAV path is computed from point-to-point between two constrained waypoints or when a vertical angle is assigned, which may represent a typical SDA, as it is shallower than the performance VNAV path and is typically using a non-idle thrust. Detailed descriptions of SDA and CDA are presented in the following subsections.

2.2.2. Step-Down Descent Approach (SDA)

In air navigation, if the aircraft flies under Instrument Flight Rules (IFR), which represents a set of rules governing the navigation of aircraft using instruments, then the instrument approach procedures (IAP) must be followed. The IAP consists of four approach segments along the aircraft flight path, namely the initial, intermediate, and final approach and, as a backup plan to use if needed, a segment for a missed approach. Typically, the initial approach segment starts at the en route (i.e., cruise) altitude from an IAF and ends when the aircraft joins the intermediate approach segment, where the later ends at the final approach fix (FAF).

SDA is the conventional arrival procedure that pilots and air traffic controllers have been accustomed to for many years. In SDA, an aircraft begins its initial descent at the TOD point and continues descending gradually in a series of steps along the descent path. This step-down descent occurs because the aircraft descends over a stair-like path from the current altitude to a new altitude, due to the controller instructions and/or airspace constraints. During the SDA, the aircraft gradually levels-off by transitioning from the initial to the intermediate to the final approach segments through predefined fixes that indicate the start and end of each approach segment. To fly from the fix that marks the end of the previous approach segment to the fix that marks the subsequent one, the aircraft must increase speed by employing thrust to maintain altitude [31]. Figure 2 illustrates the SDA profile and the approach segments of the IAP.

SDA also requires communication between the pilot and controller to inform and authorize air movement, which means more workload on both the aircrew and controller during a critical phase of flight that requires situational awareness and additional concentration. Once an aircraft has reached the fix or waypoint that marks the end of the previous approach segment and marks the subsequent one at the new altitude assigned by the controller, the pilot needs to utilize engine thrust to maintain altitude and prepare for further instructions from the controller with respect to approach. Air traffic may be expedited during periods of high demand at airports when using SDA through radar vectoring; however, the utilization of engine power increases fuel burn, which, in turn, increases emissions and noise levels at lower altitudes [31].

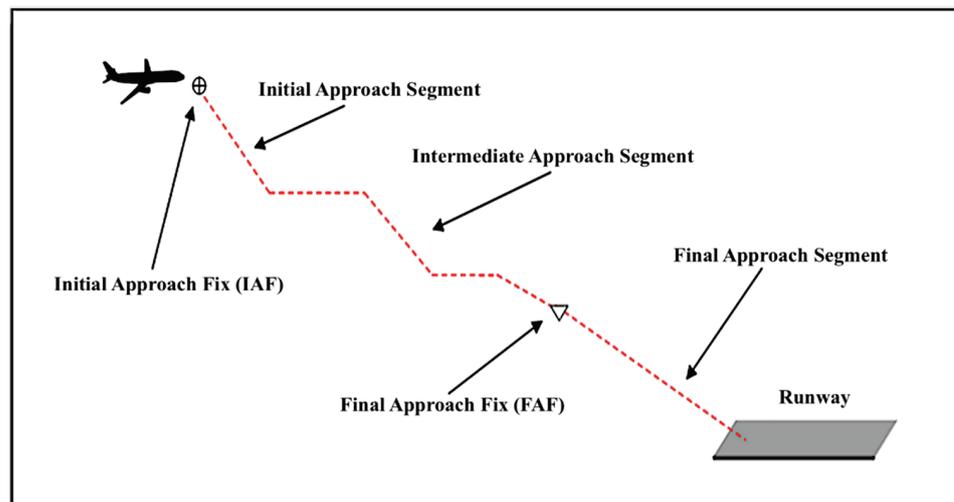


Figure 2. The vertical profile of SDA based on the IAP and approach segments.

2.2.3. Comparison between CDA and SDA

In this subsection, we provide a comparison between CDA and SDA from an operational perspective.

Considering aircraft approach speed, if a pair of aircrafts are approaching an airport for landing heading for the same runway, both aircraft approach speeds may not be the same when CDA is used, even with the same aircraft type. This is due to the fact that during descent, pilots make efforts to stabilize their approaches by controlling and balancing several parameters such as rate of descent, approach speed, thrust, and the aircraft’s attitude. With CDA, landing is conducted with idle thrust as the aircraft approach speed decreases just before touchdown [32]. With SDA, the pilot utilizes thrust and adjusts speed more frequently along the descent path, and the aircraft approach speed increases just before touchdown. This comparison is illustrated in Figure 3 below. Table 1 highlights some of the differences between CDA and SDA operations.

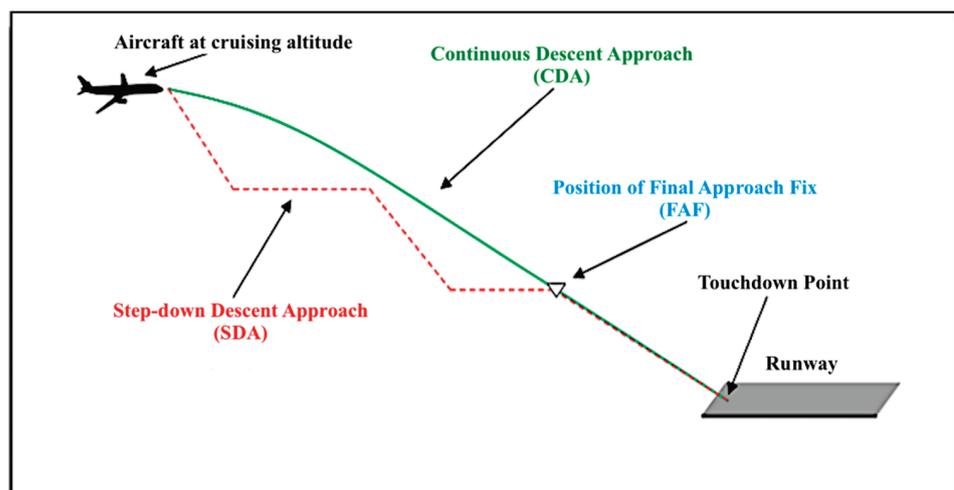


Figure 3. The vertical profile of CDA compared with SDA.

Table 1. Summary of the Differences between CDA and SDA.

Comparison Criteria	CDA	SDA
Operational Benefits	Reduces noise, emissions, and flight time and improves fuel efficiency.	May expedite air traffic during high volumes of arrivals.
Facilitation	Tactical radar vectoring, published arrival procedures (STAR), or a combination of these.	Subject to standard radar vectors with speed and altitude control.
Approach Type Based on Vertical Navigation	Performance path computed by the FMS using idle or near-idle thrust from the TOD point to the first waypoint.	Geometric path computed by the FMS between two constrained waypoints.
Sequencing and Separation of Air Traffic	Requires more spacing between aircrafts during radar vectoring and early sequencing that may require advanced sequencing tools.	Follows separation minima standards based on the sequencing method.
Impact on Airport Capacity	May reduce airport capacity.	No reduction in airport capacity
Descent Initiation	The pilot initiates descent from a TOD point that is as close to the airport as possible.	Normally, the pilot initiates descent from the TOD point at cruise altitude further from the airport than with CDA.
Aircraft Performance: Airspeed	Smooth speed profile, although the pilot may occasionally adjust speed at the controller's request to account or to balance the rate of descent.	Fluctuating speed profile as the pilot decelerates before level-off and accelerates to resume descent from a level.

3. CDA Adoptability and Aircraft Descent Times

This section introduces the factors that influence our model and then estimates and validates the time aircrafts take to land under CDA and SDA operations. The estimation and validation are essential for developing our model. In our model, we assume that during a given brief period of time, say 15 min or 30 min, for example, the air traffic controller will enable all arriving aircrafts to use CDA or not permit any aircrafts to use CDA. This assumption simplifies the controller's duties and avoids the complexity associated with a significant number of aircrafts within the TMA using CDA and a significant number using SDA. That being said, it remains possible for the controller to plan on permitting all aircrafts arriving in the brief period of time to use CDA, if suggested by the model, and yet on a case-by-case basis, decide whether to deny CDA for a particular aircraft. The case-by-case basis analysis is beyond the scope of our queueing model, but it is considered by the maximum number of aircrafts that may reside in the stacking space at any given time, as we determine much further below.

3.1. Factors Impacting CDA Adoptability and Aircraft Descent Time

Before we present the details of our model [33], we briefly discuss the concept of acceptance, and rejection, in the context of landing operations at an airport and, particularly, with CDA operations. In general, our queueing model assumes that at a given airport, the air traffic controller will either accept all CDA requests from aircraft arrivals to approach and land using CDA over a specified brief time period, say 15 min, or reject them all. Before the queueing model details are calculated, we first calculate the CDA Adoptability Factor (CDA_AF). This factor is a function of λ_{ss} , the average arrival rate of the aircraft that requests CDA at an airport (which our model assumes is all the aircrafts arriving at

the TMA's stacking space due to the advantages of continuous descent), and *AAR*, the Airport Arrival Rate, which is defined as the dynamic parameter that specifies the number of arriving aircrafts that an airport can accept during any consecutive 15-min period of time [34]. As shown from the equation below, *CDA_AF* represents the ratio of λ_{ss} to *AAR*:

$$CDA_AF = \frac{\lambda_{ss}}{AAR} \quad (1)$$

If the value of *CDA_AF* during the brief time period is high (e.g., over 100%), then there is no need to continue the analysis. It is obvious in that case that *CDA* will not be permitted during that time period. Conversely, if the value of *CDA_AF* is low (e.g., under 10%), then it is obvious that *CDA* will be permitted during the time period. The queueing calculations are performed only for those periods of time when the value of *CDA_AF* does not make it obvious whether or not *CDA* should be permitted during the time period.

Factors Impacting *CDA* Adoptability and Aircraft Descent Times

There are a number of factors that could impact the nature of aircraft arrivals at airports. Such factors could be operational, meteorological, planning, technological, or related to airspace structure and procedures design. We discussed some of these factors in Section 2 and discuss others briefly in the following subsections. Technology factors (e.g., the level of Air Traffic Management automation at an airport) are beyond the scope of this work. Other factors, however, such as traffic at neighboring airports and wind speed and direction, can be managed by reducing aircraft stacking space and increasing the minimum separation distance between aircrafts.

The Airport Arrival Rate (*AAR*) states the hourly capacity of airplane arrivals at an airport, and thus, it is critical to our model.

The aircraft fleet mix, or more generally fleet mix, refers to the ratio of various aircraft types that, based on wake turbulence categories, make up the total arrival traffic that operates at an airport. Fleet mix is essential in airport planning to determine the likely average landing speed and separation requirements on final approach, which are important factors that affect the *AAR* and, in turn, our model. Generally speaking, and from the perspective of runway capacity, which is defined as the expected number of landings that can be performed per hour on a runway, a relatively homogenous fleet mix, consisting of two dominant aircraft classes, is more favorable than a heterogeneous fleet mix.

The aircrafts' separation requirements determine the maximum number of aircrafts that can navigate each part of the airspace or can use a runway system per unit of time. The separation requirements for an aircraft landing on the same runway specify the minimum separation in longitudinal distance, or time, that must always be maintained between two aircrafts operating consecutively on the runway. These requirements are also specified for every possible pair of classes and every possible sequence of movements [35]. Table 2 exhibits the ICAO's minimum wake turbulence separation standards [36], and apparently, the larger the separation required, the lower the *AAR*. Furthermore, the more heterogeneous the fleet mix at an airport, the more influence there will be on *AAR* and our model. These separation distances are based upon *SDA* being used. If *CDA* is used, then the minimum separation distances are the same when the leading and trailing aircrafts are from the same weight turbulence category but longer than with *SDA* when the aircraft weight classes differ. Furthermore, the air traffic controller will be inclined to use longer minimum separation distances with *CDA* because of the greater challenge of controlling aircrafts using *CDA*.

Table 2. ICAO Minimum Wake Turbulence Separation Standards.

Leading Aircraft	Trailing Aircraft					
	Separation in Distance (nmi)			Separation in Time (s)		
	Heavy	Medium	Light	Heavy	Medium	Light
Heavy	4	5	6	105	131	158
Medium		3	4		79	105
Light			3			79

Among the usually considered weather conditions at airports, such as cloud ceiling and visibility, wind speed and direction are the most influential conditions on ATM operations in general and on approach operations, in particular. The two components of the wind, headwinds and tailwinds, have a significant impact on AAR. In fact, wind speed and direction dictate the availability and orientation of runways at any given time. Adverse wind conditions can reduce AAR due to the increased complexity of merging arrival traffic streams and separating aircrafts as they descend and change heading under intense or varying winds. Specifically, winds aloft may result in a phenomenon called compression, in which the separation between pairs of arriving aircrafts decreases rapidly as they descend to the final approach [37]. The results from applying our model indicate it captures the effect of wind speed.

In general, airport and airspace constraints refer to limitations that hinder airport capacity by creating difficulties for arrival aircrafts, largely due to airspace consideration. Often, such constraints are contingent on the original airspace design, which gradually became less efficient due to increasing demand and fluctuating traffic patterns, or airspace redesign, which necessitates consideration of nearby restricted airspace. On the other hand, a restricted airspace, which is an area of airspace typically used by military operations, could be close to an airport and would impose a specific airspace design that affects the pattern of the arriving aircraft. Other airspace constraints include the topographical nature and terrain (e.g., an airport close to a mountainous terrain). Airspace constraints, collectively as a single factor, are beyond the scope of this work. While estimating the effect of this factor is beyond the scope of this work, our model's behavior reflects its ability to capture such effect.

Growth in air traffic at airports within close geographical proximity likely will create congestion, especially if these airports are in a large, busy metropolitan area. The impact of air traffic at neighboring airports comes from systems of airports commonly referred to as *metroplexes*. Operationally, air traffic that flows into and out of airports within a metroplex airport system needs to be coordinated between airports in such systems to maintain efficient air traffic and individual airports' throughput, while contributing little (if any) impact to the AAR of an airport over another in such a system [38,39] and, therefore, limiting or even preventing the use of our model. The FAA is having ongoing efforts to accommodate CDAs within metroplexes, with plans to deploy during a later phase of NextGen.

3.2. Estimation of Aircraft Descent Time

In this section, we estimate the time an aircraft takes to descend, starting from the TOD point at cruise altitude down to the runway, under CDA and SDA operations, using version 3.11 of Base of Aircraft Data's (BADA) Aircraft Performance Model (APM) [40]. Estimating aircraft descent time under the two distinct approach operations is a fundamental step towards developing our model.

BADA is an APM developed and maintained by the European Organization for the Safety of Air Navigation, commonly known as EUROCONTROL, through active cooperation with aircraft manufacturers and operating airlines. To estimate aircraft landing time at airports using BADA, we used BADA's web-based Calculation Tool, the Aircraft Performance Calculation (APC), to calculate aircraft performance for the descent phase of flight.

Essentially, BADA's application software provides access to an online implementation of BADA APM, which consists of a database of aircraft operational performance files and formulas derived from the Total-Energy Model that EUROCONTROL relied on to model aircraft performance in categories such as aircraft, aerodynamics (e.g., drag), and engine thrust [41], as follows:

3.2.1. Aircraft Velocity and Lift Model

For a straight-and-level flight at cruise altitude, the aircraft speed (velocity) is given by

$$V_{TAS} = a_0 M_{cruise} \sqrt{\frac{T}{T_0}} \quad (2)$$

where V_{TAS} is aircraft's true airspeed (TAS) in nautical miles per hour (knots), a_0 is the speed of sound at sea level in knots, M_{cruise} is the aircraft's Mach number at cruise altitude, and T and T_0 are the temperatures at cruise altitude and at sea level, respectively. The lift coefficient, C_L , can be calculated using the classical formula for the lift force, L :

$$L = C_L \frac{1}{2} \rho V^2 S \quad (3)$$

where ρ is the density of air in kilograms per meter cubic, V is the aircraft speed in meters per second, and S is the aircraft's wing area in square meters.

In cruise flight, the lift force, L , in Newtons, may be assumed to be equal to the aircraft's weight in kilograms, m . Combining this relationship with Equation (3) and rearranging terms results in

$$C_L = \frac{2mg}{\rho V^2 S} \quad (4)$$

where g is the acceleration due to the earth's gravity. Assuming a no-wind scenario and that the flight path's angle in degrees is γ , then the relationship between ground speed and true airspeed is given by

$$V_{ground} = V_{TAS} \cdot \cos \gamma \quad (5)$$

3.2.2. Drag Model

Drag is the aerodynamic force acting on an aircraft body in terms of air resistance to aircraft motion through air. Similarly, to the lift force, the aerodynamic drag, D , is the product of the dynamic pressure and drag coefficient, as follows:

$$D = C_D \frac{1}{2} \rho V^2 S \quad (6)$$

The drag coefficient is given by the sum of zero-lift, C_{D_0} , and induced drag, C_{D_i} , coefficients, where the latter is a quadratic function of the lift coefficient, as follows:

$$C_D = C_{D_0} + C_{D_i} C_L^2 \quad (7)$$

Typically, C_{D_0} and C_{D_i} are functions of the aerodynamic configuration of the aircraft flight phase. Generally, drag coefficients are functions of the aircraft's Mach number and the Reynolds number ($Re = \rho VL / \mu$, where μ is the absolute viscosity coefficient of air). For each aerodynamic configuration, BADA models these coefficients as constants to provide computations for altitude and speed profile thresholds at pre-determined flight phases (i.e., takeoff, initial climb, clean, approach, and landing).

3.2.3. Thrust Model

BADA uses a general formula to calculate the maximum climb thrust, $Thr_{max,climb}$, at a standard atmosphere for three different types of engines: jet, turboprop, and piston engines. For jet engines, the general equation is given as

$$Thr_{max,climb} = C_{Tc,1} \times \left(1 - \frac{H_p}{C_{Tc,2}} + C_{Tc,3} \times H_p^2 \right) \quad (8)$$

Since BADA uses this maximum climb thrust for both take-off and climb phases, the descent thrust is then calculated from the maximum climb thrust using adjustment coefficients for cruise, approach, and landing configurations [41], respectively, as follows:

$$Thr_{des,low} = C_{Tdes,low} \times Thr_{max,climb} \quad (9)$$

$$Thr_{des,app} = C_{Tdes,app} \times Thr_{max,climb} \quad (10)$$

$$Thr_{des,ld} = C_{Tdes,ld} \times Thr_{max,climb} \quad (11)$$

where $C_{Tc,1}$, $C_{Tc,2}$, $C_{Tc,3}$, $C_{Tdes,low}$, $C_{Tdes,app}$, and $C_{Tdes,ld}$ are aircraft-specific coefficients, and H_p is the geo-potential pressure altitude, in feet. The rate, in feet per minute, at which an aircraft's altitude changes with respect to time when descending and approaching the runway for landing is the Rate of Descent (ROD). ROD is given by

$$ROD = \frac{dh}{dt} = \frac{(Thr_{des} - D)V_{TAS}}{mg} - \frac{V}{g} \frac{dV}{dt} \quad (12)$$

where dV/dt is the aircraft's vertical speed, in feet of descent per minute. Given that the typical target of a flight path is about 3 degrees, the flight path angle, γ , in degrees for a 3-degree flight over the descent path is

$$\gamma = \sin^{-1} \left(\frac{ROD}{V_{app}} \right) \quad (13)$$

where V_{app} is the aircraft approach speed, in knots. The distance, in nautical miles, that the aircraft covers over the descent path is given as follows:

$$\text{Distance} = \frac{(\Delta h \div 100)}{\gamma} \quad (14)$$

where Δh is the difference between the altitude that the aircraft is currently flying at and the altitude that the aircraft will descend to, in feet. Finally, the time, in minutes, that the aircraft takes to descend and land can be estimated by dividing the difference in altitude, in feet, by the rate of descent, in feet per minute, as follows:

$$\text{Aircraft Landing Time} = \frac{\Delta h}{ROD} \quad (15)$$

3.3. Evaluation of Aircraft Estimated Descent Time

To evaluate the calculations outputs of BADA APM, Figure 4 shows a comparison between the estimated landing times computed by BADA APC and actual landing times for an aircraft with CDA operated at Nashville International Airport (BNA) on 17 June 2015. There is a slight variation observed across the compared values between the estimated landing times and actual landing times. For example, with a CRJ9 aircraft that has estimated and actual landing times of 38 min and 27 min, respectively, there is an error of almost 29%, while for a CRJ7 aircraft with estimated and actual landing times of 20 min and 19 min, respectively, there is an error of 5.3%. On average, BADA APC have estimated the landing time for an aircraft with CDA operations to be 20 min. When compared with the actual

average landing time for an aircraft with CDA at BNA airport, which is 21 min, an error of 4.7% was generated from this estimation.

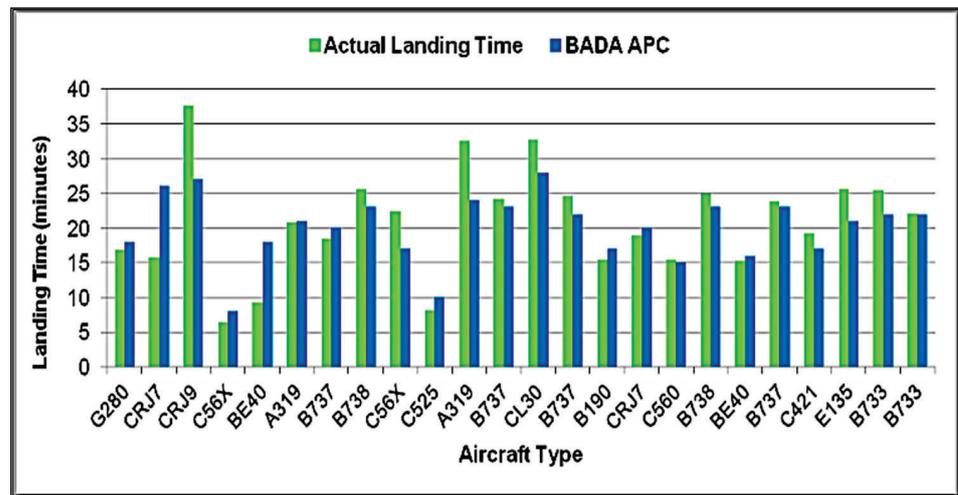


Figure 4. Evaluation of aircraft landing times with CDA at BNA airport.

Similarly, Figure 5 shows a comparison between the estimated landing times computed by BADA APM, using BADA APC, and actual landing times for an aircraft with SDA operated at BNA. It shows that for a B737 aircraft with estimated and actual landing times of 32 min and 36 min, respectively, BADA APC produced an error of about 11%, with an error of about 8% for an E135 aircraft with estimated and actual landing times of 39 min and 36 min, respectively. However, there are SDA instances where BADA APC was able to match the estimated landing time with the actual landing time, such as with the MD88 aircraft, or provide close to a match, such as with the FA50 aircraft. On average, BADA APC have estimated the landing time for an aircraft with SDA operations to be 21.7 min. When compared with the actual average landing time for an aircraft with SDA at BNA airport, which is 24 min, an error of 9.6% results from this estimation.

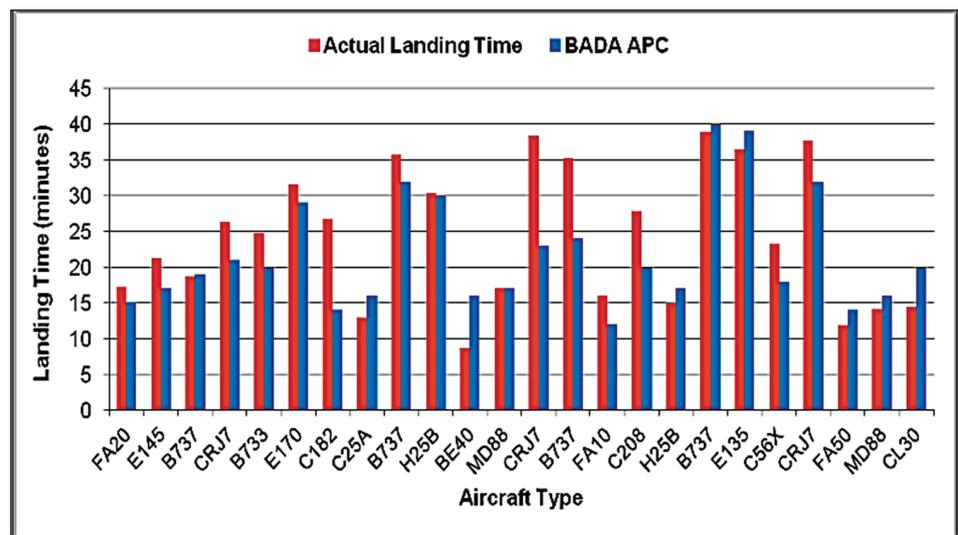


Figure 5. Evaluation of aircraft landing times with SDA at BNA.

4. Model Development

4.1. Background on Queueing Theory

In this subsection, we briefly discuss the fundamentals of queueing theory. Queues, or waiting lines, are common in people's daily lives. Queueing theory is the field of study within operations research (OR) that concerns the study of queueing models to represent the different types of queueing systems (systems that involve some sort of queue) that appear in real-world applications. Thus, these queueing models are helpful for determining how to operate a queueing system [42].

The basic process of most queueing models is that customers requiring service are generated over time by an input source (also known as a calling population). The arrival pattern by which customers are generated from the input source is statistically defined to accommodate the randomness of the customer arrival pattern. A common assumption is that customers arrive according to a Poisson process, that is, customers arrive at random but at a fixed mean rate, or equivalently, the time between consecutive customer arrivals, that is, interarrival time, follows an exponential distribution. These customers enter the queueing system and form a queue to wait for the required service.

Queues could be infinite or finite according to the maximum permissible number of customers that they can contain. At certain times, a customer is selected from the queue for service according to some defined rule referred to as the queue discipline (usually first-come-first-serve, shortly known as FCFS, or some priority-based rule). The service is then provided to the selected customer by a service mechanism that may consist of one or more service facilities, each of which contains one or more servers. The time elapsed from the commencement to the completion of service for a customer is referred to as the service time. Collectively, characteristics of queueing systems include arrival patterns of customers, service patterns of server(s), the number of servers, system capacity, queueing discipline, and the number of service stages, if more than one service stage exists [43].

A convenient notation for summarizing the basic characteristics of the queueing systems was developed by D. G. Kendall and is known in the literature as the Kendall notation. It follows the notation of $(a/b/c)$, where a = customer arrivals distribution, b = service time distribution, and c = number of servers [44]. For instance, the queueing model $(M/D/5)$ uses Markovian (or Poisson) arrivals (or equivalently, exponential interarrival time distribution), deterministic (constant) service time, and five parallel servers.

Generally, there are three basic measures of performance for queueing systems: the waiting time that a typical customer endures, the number of customers that may accumulate in the queue or system, and the idle time of the servers. Since most queueing systems follow random processes (i.e., stochastic processes), these measures are represented as random variables, and thus, their probability distributions need to be defined. Depending on if the main objective of modeling a queueing system is whether to determine some measure of effectiveness for a given process or to design the optimal system based on some defined criterion, the measures of performance could include the expected number of customers in the system, expected number of customers in the queue, expected waiting time in the system, expected waiting time in the queue, and expected number of busy servers.

Beyond the previously mentioned measures of performance, there is a measure of performance of particular interest that indicates the percentage of time the service facility within the queueing system is being utilized. This measure of performance represents the traffic intensity or utilization factor, which is the expected arrival rate of the customers to the queueing system, divided by the expected service rate, assuming one server in the service facility. If more than one server is available, then the number of servers must be multiplied by the expected service time. The utilization factor is an important performance measure of the queueing system.

4.2. Adopting Queueing Theory to Our Model

In this section, we introduce the fundamental parameters and essential conceptual elements for developing our model. In our model, aircrafts arriving at the TMA are viewed

as customers of a queuing system. The aircraft within the stacking space are modeled as the customers waiting for the service and being served. Aircrafts leaving the queueing system are viewed as customers completing service at a single server.

4.2.1. Assumptions and Parameters of the Model

1. The space available for stacking aircraft arrivals in the TMA is considered as the maximum number of aircrafts (customers) that are permitted in the queue, and
2. the longitudinal separation distance between aircrafts conducting CDA are greater than the distance between aircrafts not conducting CDA

Our model assumes that the number of aircraft arrivals over the period of time considered follows the Poisson probability distribution. This distribution has high variability and, thus, is likely to lead to the model results being conservative. Moreover, the fleet mix is assumed to be homogeneous, that is, dominated by two aircraft wake turbulence classes. The following parameters represent the fundamental components of our model: the space available to stack aircraft arrivals, the minimum allowable horizontal separation distance between a pair of consecutive same-weight-class aircraft arrivals, and the number of aircrafts that can be stacked for the approach. Figure 6 illustrates these components in our model. In this regard, while stacking space can be viewed as three dimensional, we model it as one dimensional, reflecting the longitudinal separation of stacked aircrafts as if the aircrafts within the space are all positioned in a single line.

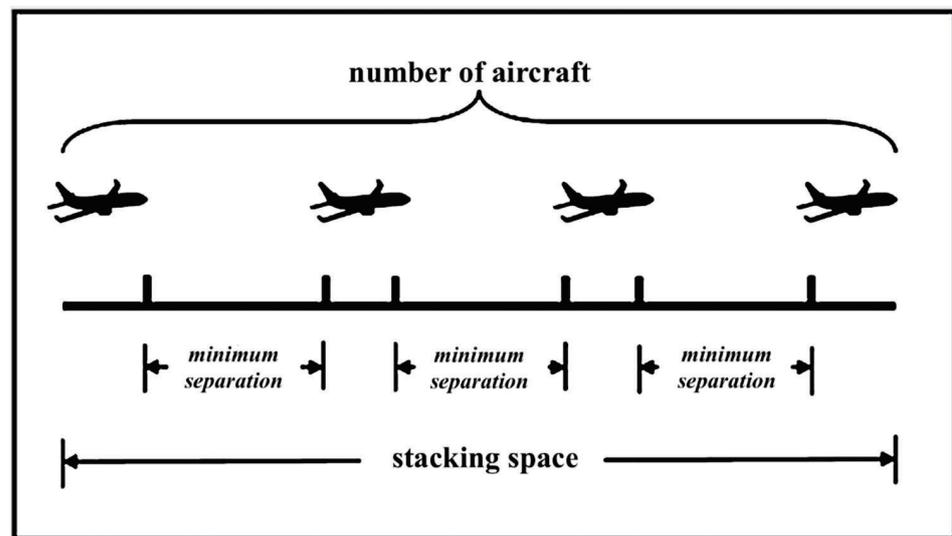


Figure 6. Parameters of our model for aircrafts approaching airports with CDA.

4.2.2. Capacity of the Stacking Space for Aircraft Arrivals

To maximize airport capacity, especially during periods of high demand, the controller longitudinally aligns and separates approaching aircrafts (i.e., positions arriving aircrafts in the queue) for landing on the same runway in a predetermined airspace, according to a predefined requirement for minimum separation between aircrafts that typically operates under IFR [36].

Optimal spacing refers to the efficient implementation of separation requirements by the controller, such that spacing delivers seamless and efficient air traffic control services while maintaining safety. As the controller often emphasizes sequencing (ordering of aircrafts approaching based on their sizes), this should not be the case with CDA operations. During CDA operations, the optimal spacing between aircrafts is more important than optimal sequencing [45]. Thus, we principally assume that the separation distance (mapped as the horizontal distance in Figure 6) between two, same-weight-class, consecutively

arriving aircrafts conducting CDA is *greater* than when these two consecutive arriving aircrafts are conducting SDA, thus

$$d_{CDA} > d_{SDA} \quad (16)$$

where:

d_{CDA} = minimum separation distance between aircrafts conducting CDA; and

d_{SDA} = minimum separation distance between aircrafts conducting SDA.

Assuming that the space available to stack aircraft arrivals (i.e., the maximum number of aircrafts in the queuing system) at an airport is S_p , and the minimum allowable horizontal separation distance between a pair of same-weight-class aircraft arrivals is d , then the number of aircrafts stacked for approach, k , must fit safely within the allowable stack space, as follows:

$$k \leq \frac{S_p}{d} \quad (17)$$

The largest integer value of k that satisfies Equation (17) is a key output of our modeling. When CDA is used, that largest integer value of k is the maximum number of aircrafts that can fit within the stacking space under the CDA assumption. The air traffic controller may be able to compare this largest integer with the number of CDA aircrafts presently in the tracking space to ascertain whether there is available stacking space to permit the next arriving aircraft to use CDA or not.

Assuming that the aircraft approach speed, measured in knots, on average, is V_{app} and that the distance the aircraft covers during descent from the TOD point to touchdown, measured in nautical miles, is d_{des} , then the *average* descent time, t_{des} , could be estimated as

$$t_{des} = \frac{d_{des}}{V_{app}} \quad (18)$$

We assume that an airport's nominal capacity, AAR , is sufficiently large to handle all aircrafts that can fit safely within the stacking space. Therefore, when implementing CDA, this assumption is represented as follows:

$$AAR \times t_{des} \geq k \quad (19)$$

Essentially, stacking space is a contained airspace with predefined boundaries based on traffic and/or obstacles limitations with the purpose to stack aircraft arrivals up to a certain capacity. As the separation distance between aircrafts increases, the stacking space capacity, in terms of the number of aircrafts that could be stacked, k , will decrease. Moreover, as the airport arrival rate increases, typically during periods of high demand when many airport staff are working and the airport operates at near capacity, stacking space capacity may decrease as well. This is due to the high level of traffic causing stress and cognitive pressures on the air traffic controllers who, thus, may decide to increase the minimum separation distance between aircrafts as a safety buffer to reduce stress and the possibility of a safety error.

Furthermore, we assume that almost all aircraft arrivals at the airport are expected to successfully land on a runway, regardless of their descent profile type. To attain this operationally, the runway, as a critical element in ATM and airports operations, is assumed to have an arrival capacity that is at least as large as the AAR . The maximum runway arrival hourly capacity is calculated by dividing the average aircraft ground speed, GS , in knots, crossing the runway threshold by the longitudinal separation distance, d , in nautical miles, required between successive arrivals, as follows:

$$RwyCap = \frac{GS}{d} \quad (20)$$

Observe above that the stacking space capacity, k , was determined above based on the minimum allowable separation distances between a pair of aircrafts of a similar weight class. Thus, the value of k represents a bound on the capacity of aircrafts that may safely fit within the stacking space. Consequently, an upper bound of k should be determined based on the worst-case sequencing of aircrafts within the stacking space. That worst case can be determined by assuming a sequence in which the lightest aircraft scheduled to land that day is followed by the heaviest aircraft scheduled that day followed by the lightest aircraft scheduled and so forth. Following these calculations, the air traffic controller may be provided with the lower and upper bounds of the stacking space capacity under the assumption that all aircrafts arriving during that time period use CDA. When a particular aircraft arrives during the period, the controller may compare the number of aircrafts presently in the system versus the lower and upper bounds. If the number in the system is less than the lower bound, then that is a favorable indicator that the arriving aircraft may be admitted using CDA. If the number is above the upper bound, then that indicates that CDA should be denied. If the number of aircrafts within the stack is as much as the lower bound or greater and yet below the upper bound, then the controller will need to consider other factors in reaching a decision on whether CDA is safe for that arriving aircraft. Even if the number in the system is below the lower bound, those other factors may need to be considered by the controller to ensure there is sufficient separation distance between the arriving aircraft and the aircraft it follows most closely within the stacking space. Providing the controller with lower and upper bounds on stacking space is an optional tactic. Doing so depends on whether it would be too much additional information for the, often busy, controller to absorb.

5. The Applied Queueing Model

In this section, we present our model and its key output, the Probability of CDA Blocking, and by blocking, we mean that an aircraft would be denied conducting CDA by the controller. Therefore, the probability of CDA blocking is the probability that (assuming all aircrafts arriving during the brief period of time considered (e.g., 15 min) are assumed to use CDA) an aircraft would need to revert to SDA even though the initial plan was for all aircrafts arriving in that time period to use CDA. However, we first discuss how the concept of traffic intensity, which is borrowed from the queueing theory, applies in the context of our model.

5.1. Traffic Intensity

Queueing theory presents a key parameter known as the traffic intensity, also referred to as the utilization factor, which is denoted by the Greek letter ρ_{ss} ("rho"), which is defined here as the average hourly demand rate of the stacking space divided by the average hourly capacity (or *service*) of the stacking space. If the average demand rate (the rate at which aircrafts arrive at the stacking space, i.e., the aircraft arrival rate) is denoted by λ_{ss} and the average service rate is denoted by μ_{ss} , then the *utilization of airspace factor*, ρ_{ss} , for the stacking space within the TMA is as follows:

$$\rho_{ss} = \frac{\lambda_{ss}}{\mu_{ss}} \quad (21)$$

where the demand rate is expressed in terms of the number of aircrafts that arrive per hour at the stacking space, and service rate is expressed in terms of the number aircrafts per hour that may enter the stacking space. The value of the service rate, μ_{ss} , is conceptually equivalent to the airport arrival rate (AAR). However, because λ_{ss} is expressed in our examples per 15-min time period, the value of the service rate, μ_{ss} , represents the number of aircrafts that can be processed (served) per 15-min time period, and thus, it is one fourth of the value of AAR. The symbol μ_{ss} is typically used in publications on queueing theory, while AAR is commonly used in air traffic management. Observe that the service rate,

μ_{ss} , will be lower when CDA is assumed than when SDA is assumed due to the longer separation distances required between aircrafts using CDA.

5.2. Probability of Aircraft Blocking

In a queueing system of finite capacity, the probability of “blocking” is the probability that an arriving customer arriving at the queue finds it full and thus exits this system. In our context where aircrafts are the customers, the interpretation of the probability of blocking depends on whether CDA or SDA is assumed for all aircrafts during the brief time period being analyzed. If CDA is assumed for the time period, then a full queue suggests that CDA may not be used, and thus, SDA will be used instead. If SDA is assumed for the time period, then a full queue suggests that the aircraft will enter a holding pattern until the queue has available space to accommodate it. The probability of blocking is the percentage of time an aircraft’s request to embark on CDA (or SDA) is denied principally due to safety and because the stacking space within the TMA is busy and congested. This probability is denoted by P_k and could be specified for an airport and its TMA to define a threshold beyond which CDA is unsafe to implement, or in the case of SDA, it is the probability that an aircraft will need to enter a holding pattern. Since the approach operations would be limited to the two profiles, namely CDA and SDA, then P_k should be determined for these two approach profiles. Theoretically, P_k is expressed based on the $M/M/1/k$ queueing model, in which the arrival process is Poisson with rate λ_s and the service process is Poisson with rate μ_s , a single server (that is, the stacking space), and finite system capacity at k aircraft, as follows:

$$P_k = \frac{1 - \rho_{ss}}{1 - \rho_{ss}^{k+1}} \rho_{ss}^k \quad (22)$$

Because the Markov (Poisson) process distribution has high variability and is assumed for both arrivals and service, the resulting probability is likely to be higher than the value experienced in practice, and thus, the Markov assumption may be viewed as conservative.

6. Numerical Results to Illustrate the Model

This section illustrates the application of our model through the use of industry standard data (e.g., minimum separation distance rules) and a stream of aircrafts arriving for landing at a mid-sized international airport during an afternoon level of demand. In particular, we used actual flight data from flights operated at Nashville International Airport (BNA) from 1200 to 1759 local time, on 17 June 2015.

Using BNA flight data and standard industry data, we calculated the probability of blocking for each type of descent profile, namely, CDA and SDA. These probabilities are denoted as P_{kCDA} and P_{kSDA} . P_{kCDA} is the probability of blocking if all the aircrafts in the stacking space will conduct CDA, and P_{kSDA} is defined as the probability of blocking if all the aircrafts in the stacking space will conduct SDA. As previously mentioned, the probability of blocking is the percentage of time an aircraft request to embark on CDA is denied for safety considerations and due to the stacking space being congested and busy and similarly the fraction of time an aircraft landing with SDA would need to enter a holding pattern.

The relevant parameters in building the model are described as follows. The rate of aircrafts arriving per time period, λ_{ss} , is determined based upon the actual number of arrivals at BNA within each of the 24 15-min time periods analyzed. The stacking space, S_p , was estimated by visually examining the data and assuming its value as a constant value throughout the six hours. The wind speed, W_s , is the average value at BNA during each time period. The aircraft approach speed, V_{app} , is determined based on averages, across all arriving aircrafts, of their initial approach velocity and final approach velocity. The fleet mix of heavy-, medium-, and light-weight aircrafts arriving in each time period and their sequence of arrivals is determined from what actually occurred in each time period.

The minimum separation distances under SDA and CDA assumptions, d_{SDA} and d_{CDA} , respectively, were chosen based on ICAO’s wake turbulence application [36] and matrix

model to calculate the separation distance between a mix of aircrafts proposed in [35] in each time period. As mentioned previously, these distances vary based on the wake turbulence categories of the leading and trailing aircrafts and, thus, depend on the fleet mix in each time period. Furthermore, the values of d_{SDA} and d_{CDA} are identical when the leading and trailing aircrafts have the same weight class and differ otherwise.

Figure 7 shows the values of the aircraft arrival rate, λ_{ss} , over time for BNA.

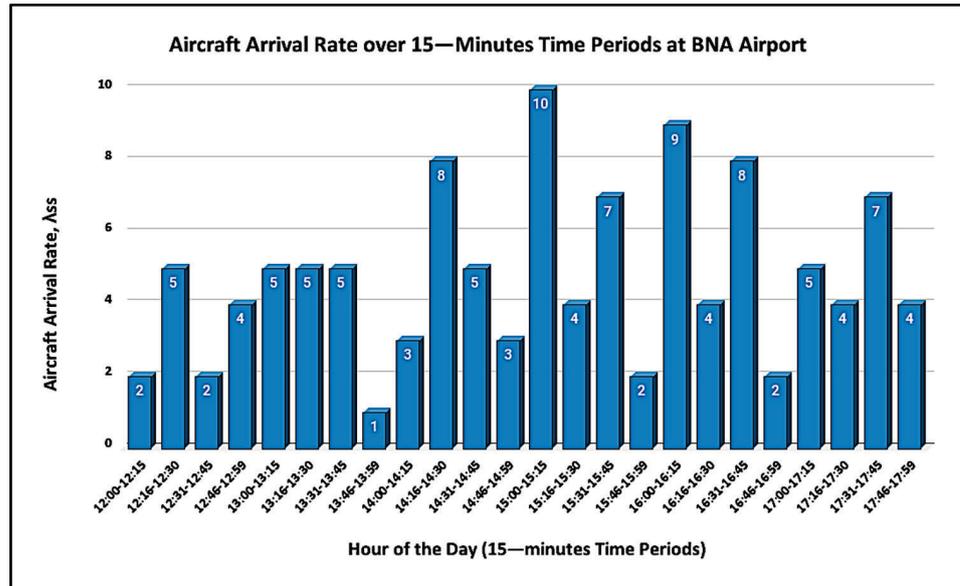


Figure 7. The number of aircrafts arriving in each time period at BNA airport.

Consistent with Equation (17), we calculated the number of aircrafts that can fit in the stacking space if CDA is used, k_{CDA} , and the number of aircraft that can fit in the stacking space if SDA is used, k_{SDA} , as follows:

$$k_{CDA} = S_p / d_{CDA} \tag{23}$$

$$k_{SDA} = S_p / d_{SDA} \tag{24}$$

Those values of k_{CDA} and k_{SDA} for BNA are shown in Figure 8. This figure may be provided as systematic output for the air traffic controllers, thus enabling them to know the bounds on the maximum number of aircrafts that may be permitted within the stacking space for each descent method. Because the minimum separation distances for CDA are longer than with SDA, the value of k_{CDA} is always lower than k_{SDA} , while the difference between the two varies due to the varying fleet mix and arrival sequences in each period.

Using Equation (18) to determine T_{des} , we determined the service rate, μ_{CDA} and μ_{SDA} , under the CDA and SDA conditions in the following manner, using a calculation similar to Equation (19), described earlier:

$$\mu_{CDA} = k_{CDA} / T_{desCDA} \tag{25}$$

$$\mu_{SDA} = k_{SDA} / T_{desSDA} \tag{26}$$

Because CDA requires greater minimum separation distances between aircrafts than SDA, this reduces the rate at which CDA arrivals can be processed through the stacking space and, thus, results in μ_{CDA} being lower than μ_{SDA} .

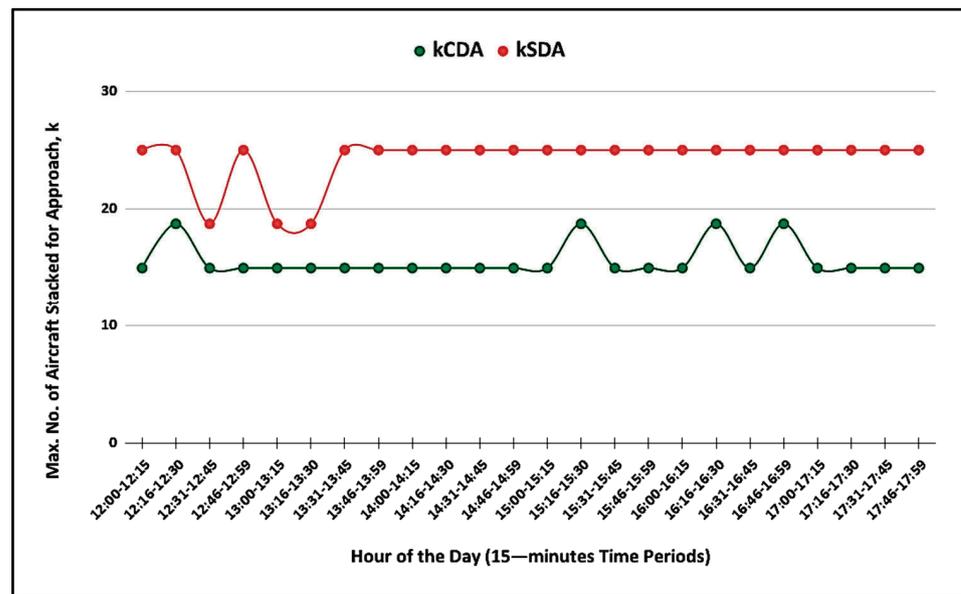


Figure 8. Capacity of the TMA’s stacking space k .

Using the above parameters and Equation (22), we calculated the values of the probabilities of blocking, P_{kCDA} and P_{kSDA} , for CDA and SDA, respectively, and show the results in Figure 9.

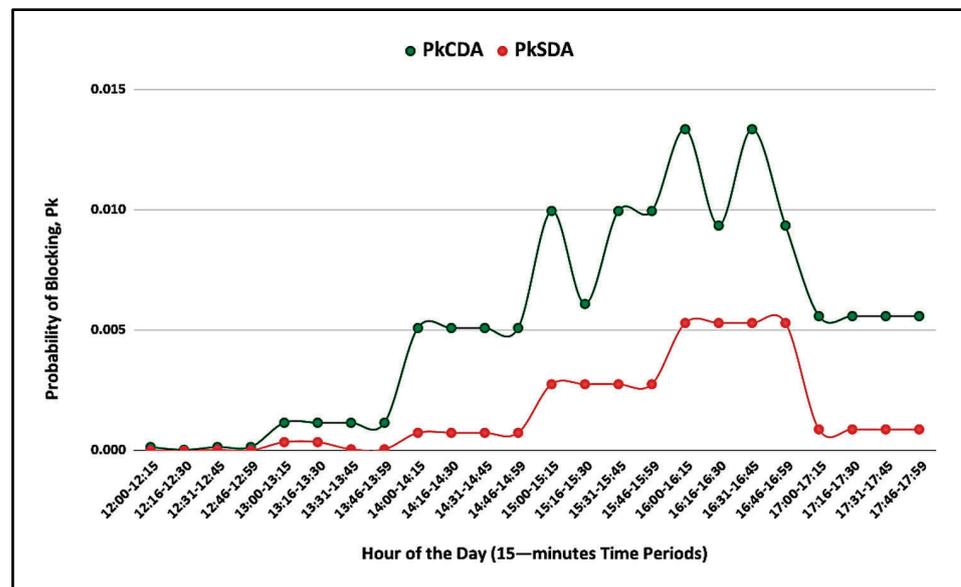


Figure 9. Probability of blocking of the two descent methods.

Observe in Figure 9 that the probability of CDA blocking (i.e., CDA requests being denied), P_{kCDA} , is consistently less than the probability of an SDA request requiring the aircraft to enter the holding pattern, P_{kSDA} . By looking at graphs such as Figure 9, the air traffic controller can see those periods of time when the vast majority of aircrafts will be able to use CDA, i.e., when the probability of CDA blocking, P_{kCDA} , is sufficiently low. For example, except for the late afternoon periods between 15:00 to 15:15 and from 15:30 to 16:45, there is less than one percent probability of CDA blocking. Aside from those late afternoon exceptions, the other time periods would be favorable for permitting CDA arrivals for all (or nearly all) arriving aircrafts. Even during the late afternoon time

periods, the probability of CDA blocking is low enough that the air traffic controller may be able to permit many of the aircrafts to land using CDA. Different air traffic controllers will have different probabilities at which they will be comfortable with using CDA. Their thresholds in such decision-making may change over time as they gain experience using the model. Overall, the information should be helpful to the air traffic controllers by focusing their attention on those opportunities to increase the use of CDA and thus lead to more sustainable air transportation operations.

After deciding on the periods of time when CDA may be used for all (or nearly all) arriving aircrafts, knowing the bounds on the number of aircrafts that may fit safely in the stacking space, k_{CDA} and k_{SDA} , as shown in Figure 8, may help the air traffic controller to decide on an aircraft-by-aircraft basis whether to admit a particular aircraft's descent using CDA. For instance, if an aircraft arrives during a period of time when the controller plans on permitting CDA and when the actual number of aircrafts within the stacking space upon a particular aircraft's arrival at the TMA is less than k_{CDA} , then the controller knows it is likely fine to accept that particular aircraft's request to use CDA, and otherwise, it is not.

7. Conclusions

Based on the analysis of the parameters that govern CDA implementation during high traffic levels, such as the terminal maneuvering area (TMA) and the size of stacking space to arrange aircraft arrivals, an analytical model has been developed that aims at addressing the accommodation of more CDA operations than is presently done. Our analysis shows that the parameters that have significant impacts on CDA usage include airport arrival rate, capacity of the stacking space, and the minimum separation distance between aircraft arrivals. Although CDA usage is also affected by other parameters, such as wind speed, types of arriving aircrafts, and traffic levels at contiguous airports, we were able to capture the underlying relationship between these parameters and CDA usage, potentially helping air traffic controllers to decide on whether to adopt more CDA more often and using more quantitative information. In particular, we calculated the probability that an aircraft arriving within a brief period of time (say 15 or 30 min) would be denied CDA as a function of airport conditions. This may enable controllers to identify those periods of time when CDA should be anticipated, in other words, when the controller should plan on permitting CDA. This may lead to an increased use of CDA and, thus, result in lower noise, fuel consumption, and pollution. Furthermore, we established bounds for the maximum number of CDA descending aircrafts would fit within the stacking space, providing insight to the controllers on whether to permit a particular aircraft to use CDA or not. Finally, we illustrated our model using actual data from flights operated at Nashville International Airport (BNA).

Future research opportunities include better support for air traffic controllers in analytically determining which arriving aircraft may be able to use CDA for a portion of their descent and when they should switch from SDA to CDA and vice versa. Another research possibility would be to analytically identify opportunities to put arriving aircrafts into a holding pattern for a short while if so doing would be result in it being able to use CDA after holding. Calculations could be performed to identify whether the additional fuel burned from holding (or adjustments to en route speed) would be more than compensated by the saved fuel from using CDA. In addition to the work presented in this paper, these research opportunities and others should be explored to improve aviation green operations and ensure air transportation sustainability.

Author Contributions: Conceptualization, L.L.A.-M., E.A.A. and R.J.M.; methodology, L.L.A.-M., E.A.A. and R.J.M.; software, E.A.A. and A.M.W.; validation, L.L.A.-M., R.J.M. and E.A.A.; formal analysis, E.A.A., R.J.M. and A.M.W.; investigation, E.A.A., R.J.M. and A.M.W.; resources, E.A.A.; data curation E.A.A. and A.M.W.; writing—original draft preparation, E.A.A. and R.J.M.; writing—review and editing, R.J.M.; visualization, E.A.A. and A.M.W.; supervision, L.L.A.-M.; project administration, E.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. IATA. IATA Forecasts Passenger Demand to Double Over 20 Years. 2016. Available online: <http://www.iata.org/pressroom/pr/Pages/2016-10-18-02.aspx> (accessed on 10 March 2020).
2. EPA. EPA Determines That Aircraft Emissions Contribute to Climate Change Endangering Public Health and the Environment. 2016. Available online: <https://archive.epa.gov/epa/newsreleases/epa-determines-aircraft-emissions-contribute-climate-change-endangering-public-health-0.html> (accessed on 10 March 2020).
3. European, C. Reducing Emissions from Aviation. 2017. Available online: https://ec.europa.eu/clima/policies/transport/aviation_en (accessed on 11 March 2020).
4. Dillingham, G.L.; Martin, B. *Aviation and the Environment: Results from a Survey of the Nation's 50 Busiest Commercial Service Airports*; US General Accounting Office: Washington, DC, USA, 2000.
5. Basner, M.; Babisch, W.; Davis, A.; Brink, M.; Clark, C.; Janssen, S.; Stansfeld, S. Auditory and non-auditory effects of noise on health. *Lancet* **2013**, *383*, 1325–1332. [[CrossRef](#)]
6. Clarke, J.-P.B.; Ho, N.T.; Ren, L.; Brown, J.A.; Elmer, K.R.; Tong, K.-O.; Wat, J.K. Continuous Descent Approach: Design and Flight Test for Louisville International Airport. *J. Aircr.* **2004**, *41*, 1054–1066. [[CrossRef](#)]
7. Coppenbarger, R.A.; Mead, R.W.; Sweet, D.N. Field Evaluation of the Tailored Arrivals Concept for Datalink-Enabled Continuous Descent Approach. *J. Aircr.* **2009**, *46*, 1200–1209. [[CrossRef](#)]
8. McNulty, M. Airbus Says Jets Flying Close Behind Each Other Could Reduce Fuel Costs by 10 Percent. 2019. Available online: <https://www.foxbusiness.com/lifestyle/airbus-jets-flying-close-reduce-fuel-costs> (accessed on 29 November 2019).
9. Turgut, E.T.; Usanmaz, O.; Canarslanlar, A.O.; Sahin, O. Energy and emission assessments of continuous descent approach. *Aircr. Eng. Aerosp. Technol.* **2010**, *82*, 32–38. [[CrossRef](#)]
10. Morrell, P.S. *Moving Boxes by Air: The Economics of International Air Cargo*; Ashgate: Farnham, UK, 2011.
11. Girvin, R. Aircraft noise-abatement and mitigation strategies. *J. Air Transp. Manag.* **2009**, *15*, 14–22. [[CrossRef](#)]
12. Postorino, M.N.; Mantecchini, L. A systematic approach to assess the effectiveness of airport noise mitigation strategies. *J. Air Transp. Manag.* **2016**, *50*, 71–82. [[CrossRef](#)]
13. Schroeder, J. A perspective on NASA Ames air traffic management research. In Proceedings of the 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO) and Aircraft Noise and Emissions Reduction Symposium (ANERS), Hilton Head, SC, USA, 21–23 September 2009.
14. Stibor, J.; Nyberg, A. Implementation of continuous descent approaches at Stockholm Arlanda airport, Sweden. In Proceedings of the 2009 IEEE/AIAA 28th Digital Avionics Systems Conference, Orlando, FL, USA, 23–29 October 2009; pp. 2.C.1-1–2.C.1-13. [[CrossRef](#)]
15. Arnaldo Valdés, R.; Pérez Sanz, V.G.C.L. Modeling Advanced Continuous Descent Approach Procedures in Congested Airports. *Int. Rev. Aerosp. Eng.* **2009**, *2*, 304–314.
16. Cao, Y.; Rathinam, S.; Sun, D. A Rescheduling Method for Conflict-free Continuous Descent Approach. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Portland, OR, USA, 8–11 August 2011. [[CrossRef](#)]
17. Robinson, J.; Kamgarpour, M. Benefits of Continuous Descent Operations in High-Density Terminal Airspace Under Scheduling Constraints. In Proceedings of the 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Fort Worth, TX, USA, 13–15 September 2010.
18. Weitz, L.A.; Hurtado, J.E.; Barmore, B.E.; Krishnamurthy, K. An Analysis of Merging and Spacing Operations with Continuous Descent Approaches. In Proceedings of the IEEE 24th Digital Avionics Systems Conference, Washington, DC, USA, 30 October–3 November 2005. [[CrossRef](#)]
19. Wilson, I.; Hafner, F. Benefit Assessment of Using Continuous Descent Approaches at Atlanta. In Proceedings of the IEEE 24th Digital Avionics Systems Conference, Washington, DC, USA, 30 October–3 November 2005. [[CrossRef](#)]
20. Khardi, S. Mathematical Model for Advanced CDA and Takeoff Procedures Minimizing Aircraft Environmental Impact. *Int. Math. Forum* **2010**, *5*, 1747–1774.
21. Sáez, R.; Polishchuk, T.; Schmidt, C.; Hardell, H.; Smetanová, L.; Polishchuk, V.; Prats, X. Automated sequencing and merging with dynamic aircraft arrival routes and speed management for continuous descent operations. *Transp. Res. Part C Emerg. Technol.* **2021**, *132*, 103402. [[CrossRef](#)]
22. De Jong, P.M.A.; de Gelder, N.; Verhoeven, R.P.M.; Bussink, F.J.L.; Kohrs, R.; van Paassen, R.; Mulder, M. Time and Energy Management During Descent and Approach: Batch Simulation Study. *J. Aircr.* **2015**, *52*, 190–203. [[CrossRef](#)]
23. Itoh, E.; Wickramasinghe, N.K.; Hirabayashi, H.; Uejima, K.; Fukushima, S. Analyzing Feasibility of Continuous Descent Operation Following Fixed-flight Path Angle from Oceanic Route to Tokyo International Airport. In Proceedings of the AIAA Modeling and Simulation Technologies Conference, San Diego, CA, USA, 4–8 January 2016; p. 0168. [[CrossRef](#)]

24. Fricke, H.; Seiß, C.; Herrmann, R. Fuel and Energy Benchmark Analysis of Continuous Descent Operations. *Air Traffic Control Q.* **2015**, *23*, 83–108. [CrossRef]
25. Zhao, W.; Alam, S.; Abbass, H.A. Evaluating ground–air network vulnerabilities in an integrated terminal maneuvering area using co-evolutionary computational red teaming. *Transp. Res. Part C Emerg. Technol.* **2013**, *29*, 32–54. [CrossRef]
26. Jafari, S.; Nikolaidis, T. Thermal Management Systems for Civil Aircraft Engines: Review, Challenges and Exploring the Future. *Appl. Sci.* **2018**, *8*, 2044. [CrossRef]
27. Lai, Y.-C.; Ting, W.O. Design and Implementation of an Optimal Energy Control System for Fixed-Wing Unmanned Aerial Vehicles. *Appl. Sci.* **2016**, *6*, 369. [CrossRef]
28. Wubben, F.; Busink, J. Environmental Benefits of Continuous Descent Approaches at Schiphol Airport Compared with Conventional Approach Procedures. 2000. Available online: <http://www.conforg.fr/internoise2000/cdrom/data/articles/000281.pdf> (accessed on 21 January 2022).
29. Alam, S.; Nguyen, M.H.; Abbass, H.A.; Lokan, C.; Ellejmi, M.; Kirby, S. A dynamic continuous descent approach methodology for low noise and emission. In Proceedings of the 29th Digital Avionics Systems Conference, Salt Lake City, UT, USA, 3–7 October 2010.
30. Belobaba, P.; Odoni, A.; Barnhart, C. *The Global Airline Industry*; John Wiley & Sons: West Sussex, UK, 2015.
31. Nolan, M.S. *Fundamentals of Air Traffic Control*; Brooks/Cole Company: Pacific Grove, CA, USA, 1999.
32. FAA. *Instrument Procedures Handbook*; U.S. Department of Transportation: Washington, DC, USA, 2015.
33. Alharbi, E. *Airspace Analysis for Greener Operations: Towards More Adoptability and Predictability of Continuous Descent Approach (CDA)*; Mechanical and Industrial Engineering Department, New Jersey Institute of Technology: Newark, NJ, USA, 2017.
34. FAA. *FAA Order JO 7210.3Z—Facility Operation and Administration*; U.S. Department of Transportation: Washington, DC, USA, 2015.
35. De Neufville, R.; Odoni, A.R.; Belobaba, P.P.; Reynolds, T.G. *Airport Systems: Planning, Design and Management*; McGraw Hill Education: New York, NY, USA, 2013.
36. ICAO (International Civil Aviation Organization). *Documentation 4444—Procedures for Air Navigation Services: Air Traffic Management*, 16th ed.; ICAO: Montreal, QC, Canada, 2016.
37. DeLaura, R.A.; Ferris, R.F.; Robasky, F.M.; Troxel, S.W.; Underhill, N.K. *Initial Assessment of Wind Forecasts for Airport Acceptance Rate (AAR) and Ground Delay Program (GDP) Planning*; Project Reprt ATC-414; MIT Lincoln Laboratory: Lexington, MA, USA, 2014.
38. Wei, P.; Chen, J.-T.; Andrisani, D.; Sun, D. Routing Flexible Traffic into Metroplex. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Portland, OR, USA, 8–11 August 2011.
39. Ramanujam, V.; Balakrishnan, H. Data-Driven Modeling of the Airport Configuration Selection Process. *IEEE Trans. Hum.-Mach. Syst.* **2015**, *45*, 490–499. [CrossRef]
40. Nuic, A.; Poles, D.; Mouillet, V. BADA: An advanced aircraft performance model for present and future ATM systems. *Int. J. Adapt. Control Signal Process.* **2010**, *24*, 850–866. [CrossRef]
41. Nuic, A. User Manual for the Base of aircraft data (BADA) revision 3.11. *Atmosphere* **2010**, *2010*, 1.
42. Hillier, F.S.; Lieberman, G.J. *Introduction to Operations Research*; McGraw-Hill Higher Education: New York, NY, USA, 2010.
43. Shortle, J.F.; Thompson, J.M.; Gross, D.; Harris, C.M. *Fundamentals of Queueing Theory*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2018. [CrossRef]
44. Taha, H.A. *Operations Research: An Introduction*; Pearson Education, Inc.: Hoboken, NJ, USA, 2017.
45. Chen, H.; Solak, S. Lower Cost Arrivals for Airlines: Optimal Policies for Managing Runway Operations under Optimized Profile Descent. *Prod. Oper. Manag.* **2014**, *24*, 402–420. [CrossRef]

Article

Mode Choice Effects on Bike Sharing Systems

Matthias Kowald *, Margarita Gutjar, Kai Röth, Christian Schiller and Till Dannewald

Department of Architecture and Civil Engineering, RheinMain University of Applied Sciences, 65197 Wiesbaden, Germany; margarita.gutjar@hs-rm.de (M.G.); kai.roeth@hs-rm.de (K.R.); christian.schiller@hs-rm.de (C.S.); till.dannewald@hs-rm.de (T.D.)

* Correspondence: matthias.kowald@hs-rm.de

Abstract: Bike-sharing systems (BSS) are offered in many cities and urban municipalities and urban areas without such systems are thinking about their introduction. In addition, many studies on BSS are available; however, neither mode nor route choice parameters are available for station-based BSS, which are required for the implementation of BSS in local and regional transport demand models. As a result, this makes it impossible to simulate demand model-based effects of these systems on other transport modes and e.g., calculate scenario-guided modal shifts. The paper presents results obtained from a survey study, which aims to estimate BSS-related choice parameters. The study combined computer-assisted telephone interviews (CATI) for a collection of revealed preferences (RP) on the use of BSS with a follow-up paper-and-pencil survey on stated preferences (SP) of 220 BSS users and non-users from the Rhine-Neckar area in mid-west Germany. Considering the three transport modes BSS, public transport (PT), and private motorized transport (PMT), results from this choice experiment and, according to behavioural parameters, allow integration of BSS in transport demand models and a simulation of modal shifts. Survey design, mode-choice experiment, and choice models are presented in this paper.

Keywords: bike-sharing system (BSS); mode choice; stated choice experiment; multinomial logit model; transport demand model

Citation: Kowald, M.; Gutjar, M.; Röth, K.; Schiller, C.; Dannewald, T. Mode Choice Effects on Bike Sharing Systems. *Appl. Sci.* **2022**, *12*, 4391. <https://doi.org/10.3390/app12094391>

Academic Editors: Giovanni Randazzo, Anselme Muzirafuti, Dimitrios S. Paraforos and Stefania Lanza

Received: 16 March 2022

Accepted: 25 April 2022

Published: 27 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The first bike-sharing systems (BSS) were introduced around five decades ago. Over the past two to three decades, the number of BSS has increased, and such systems are nowadays available in many cities around the globe. It is, however, surprising, that parameters for neither BSS-related mode nor route choices are currently available. This results in a lack of knowledge of behavioural patterns. Amongst other purposes, such parameters are needed for the implementation of BSS as a transport mode in transport demand models and to calculate, e.g., modal shifts.

To estimate such parameters, a survey study was conducted in the field in Germany and collected information on mode- and route choices from around 220 participants in an existing station-based BSS. This BSS was introduced in 2012, and is located in the Rhine-Neckar area in mid-west Germany, including 20 municipalities in total; 4 of them are major cities (Mannheim, Ludwigshafen, Heidelberg, Kaiserslautern), 11 are mid-size municipalities and 5 are smaller towns. The survey study combined computer-assisted telephone interviews (CATI) for a collection of revealed preferences (RP) on the use of BSS with a follow-up paper-and-pencil survey on stated preferences (SP) for BSS users and non-users. The choice experiment considers the three transport modes: BSS, PT, and PMT.

This effort resulted in a rich data set, which allows an analysis of behavioural patterns in terms of BSS-related mode and route choices and a quantification of the needed parameters; with that, the study closes a knowledge gap and allows the implementation of station-based BSS in local or regional transport demand models and according to simulations with these tools.

The paper focuses on mode choices exclusively. Section 2 presents results from literature analysis and aims to identify attributes of relevance for choices pro or against BSS-use. Section 3 introduces the study area and an existing BSS, that was used to recruit survey participants and collect information on BSS use. Furthermore, the recruitment strategy, survey protocol, and experimental design for the choice experiment are presented. Next, descriptive statistics on respondents' socio-demographics and their choices are provided in Section 4 together with the model formulation. Section 5 presents results on model development and the final multinomial logit models on choices between the alternatives BSS, public transport (PT), and private motorized transport (PMT). The models are estimated for mandatory and leisure trips and model results are accompanied by interpretations. Finally, conclusions and an outlook on future research are drawn in Section 6.

2. Literature Review

BSS were introduced in many cities around the globe in the last decade. In parallel, the number of studies on BSS has increased as well (for comparative meta-studies see e.g., Refs. [1–5]). Most of these studies indicate that BSS-use substitutes sustainable modes such as walking and PT; however, some report that BSS-use also reduces trips by car and other PMT means. The effect on PMT is closely related to the synergetic effects of multi- and intermodal combinations of BSS with PT and promoting cycling in general [5]. In order to incorporate BSS into transport demand models and thus estimate more comprehensively the potential of reducing car trips, it is necessary to understand the mode and route choice behaviour of (potential) BSS users; so far, only a few studies focused on mode and route choice behaviour of BSS-users see Refs. [6–9], while much more studies exist for cycling with private bikes. Filling this gap is the aim of the survey, which is presented in the following sections.

Several research studies investigated the influence of different attributes on cycling. Buehler and Dill [10] reviewed the effects of cycling infrastructure; among other things, they found that cyclists tend to prefer separated bike lanes, lower speed limits and volumes of motorized traffic, trees along the route, and routes with fewer intersections; further, they tend to avoid routes with on-street parking and many variations in altitude.

Other studies investigated the effects of the built environment to enhance walking and cycling levels [11,12]. Their main findings highlight the importance of short distances to destinations, mixed land use with high densities of population and facilities for groceries, retail, service, and recreation), charges for car parking, and a network of convenient cycling infrastructure for increased shares of bike-use. What cyclists perceive as convenient highly depends on the study location, but amongst other factors, they prefer segregation or protection from motor traffic, little or only small detours, avoidance of intersections with motor traffic, and bike parking facilities.

Studies have also investigated how the built environment affects the use of bike-sharing stations [13,14]. It was found that high densities and proximity to cycling-friendly routes and to PT stations play a crucial role; furthermore, a study from Lisbon described an algorithm-based approach for the identification of optimal locations for stations and fleet dimensions for BSS by taking into account user demand, renting-costs, and a mixed fleet of regular and electric bikes [15].

In addition, there are studies, which explicitly investigated the mode choice behaviour of cyclists on both private and shared bikes: Hamre and Buehler [16] studied the mode choice behaviour of around 4.600 commuters in Washington; their result is those bike parking facilities, non-free car parking, and showers at work increase the utility of commuting by bike. In another study, cycling was compared with bus riding and driving with respect to travel time reliability [17]; it was found that for habitually repeated trips, travellers rate reliability higher than travel time. Campbell et al. [6] investigated the factors influencing the choices for bike-sharing (conventional bike and e-bike) in Beijing [6]; they found that trip distance, rain, temperature, and poor air quality negatively impact the choice for non-electric BSS on the one side. On the other side, they reported that users'

socio-demographic characteristics only play a minor role. A study from Switzerland [18,19] compared several mode-specific effects on choices between PMT, PT, cycling, and walking under the inclusion of individuals' and area-specific characteristics, whereby for walking and cycling only the effect of travel time was considered as a mode-specific attribute; they found that car availability, younger age, low fuel, and parking costs, and low parking search times increase the utility of PMT, while low access and egress times, low ticket costs, and a low utilized capacity increase the utility of PT. An investigation on the Dutch mobility panel with almost 2000 participants focused on characteristics that affect choices between PMT, PT, cycling, and walking [20]. Among others, these characteristics include individual and household characteristics (such as socio-demographics and mobility ownership), weather, trip characteristics (such as distance and travel time), effects from the built environment, and work characteristics. Results indicate that, among other things, higher education, transit subscription, cycling to high school, weekdays, and certain trip purposes increase the utility of cycling, while owning a company car and travelling in larger groups decrease its utility. A recent study compared shared modes in Zurich, namely station-based BSS, e-bike sharing systems (e-BSS), and e-scooters [7]; they found that station density and morning hours increased the utility of BSS, while variation in topographical altitude and night hours decreased its utility. Ilahi et al., (2021) undertook an extensive survey with more than 5000 participants from the Greater Jakarta area in which they also incorporated less-established modes such as urban air mobility, including currently developed electric-based aircraft or autonomous vertical take-off and landing vehicles, on-demand transport, and bus rapid transit [21]; they found that motorcycles have the highest baseline utility, while bikes have the second-lowest.

In addition, numerous studies have focused on the route choices of cyclists. Although this paper is primarily focused on mode choices, it can be expected that route choice effects in general also affect mode choices. A recent study with 662 participants from Greece and Germany investigated the effects of cycling infrastructure, speed limit, surface, on-street parking, trees, and travel time on route choices [22]; they found that in both countries protected bike lanes are preferred over other forms of infrastructure. Furthermore, asphalt is preferred over cobblestone, while the utility of a speed limit depends on the country. Other attributes that were incorporated in older studies are the number of car lanes [23], stop signs and crowding of cyclists [24], number and type of crossings [25], the width of bike lanes and traffic volumes [26], as well as sharing space with pedestrians and the availability of secure parking and showers at the destination [27].

With reference to the presented studies, it can be assumed that relevant attributes for cycling with private bikes are relevant for BSS, too. Furthermore, additional attributes, such as renting costs, access and egress times, to renting stations are relevant in the case of BSS. Evaluating which effects are of relevance for BSS and quantifying their influence is the aim of our survey study. The employed survey design and choice experiment are introduced in the following section. The study aims to provide an overview of the most relevant effects to allow the implementation of BSS in multi-modal transport demand models, and, with that, to provide a basis for forecasts on urban transport, which consider BSS as an alternative transport mode.

3. Survey Design and Choice Experiment

In 2020, the transport association Rhine-Neckar (VRN) decided to evaluate the performance of its BSS named VRNnextbike and especially focus on users and their behavioural patterns such as e.g., origins and destinations of rental bike trips, trip purposes, average renting times, and distances. Figure 1 provides an overview of the supply area of VRNnextbike, which lies in Mid-West Germany; it includes 20 municipalities in total; 4 of these municipalities are major cities (Mannheim, Ludwigshafen, Heidelberg, Kaiserslautern), 16 are minor cities with 11 mid-size municipalities and 5 smaller towns.



Figure 1. The supply area of VRNnextbike.

Figure 2 provides an overview of the absolute development of bike rents between 2015 and 2020; it documents a positive development trend and repeating temporal patterns between annual seasons. Furthermore, it shows the effect of the first Corona-Lockdown in the second quarter of 2020 and the quick recovery of the system in the third quarter.

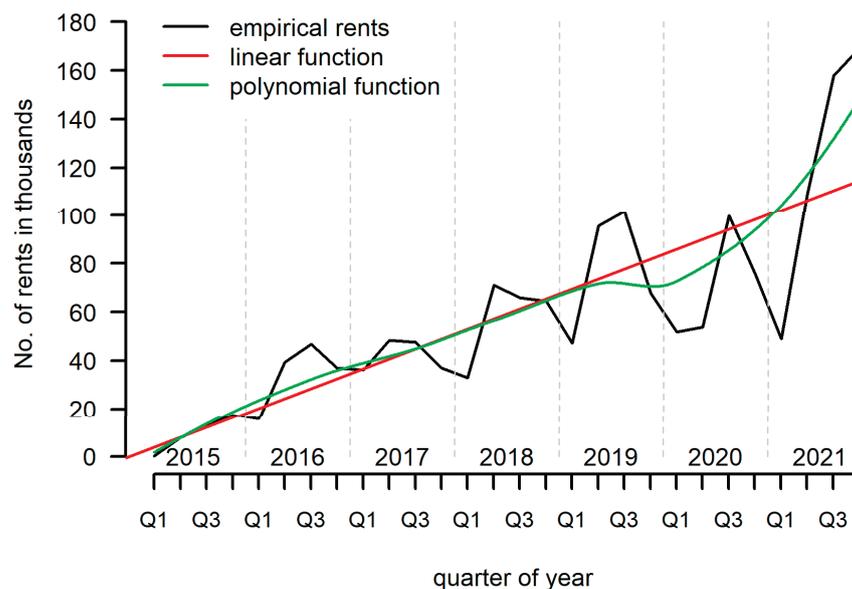


Figure 2. Quarterly development of the absolute number of rents.

Evaluating the BSS VRNnextbike and analyzing user behaviour in terms of, e.g., socio-demographics made it necessary to design a survey study. This provided the opportunity to include items on the effects of mode and route choices in terms of a BSS and to quantify mode and route choice parameters. Collecting data on BSS-related mode and route choices; however, made it necessary to extend the survey framework and include a stated choice experiment. In addition, the target population had to be extended. While the VRN-survey wanted to focus on BSS users exclusively, the choice experiment made it necessary to include non-users as well, to understand differences in the perception between these two groups and allow a later simulation of, e.g., modal shifts. For this reason, a hybrid recruitment strategy and a specific survey protocol for BSS users on the one side and non-users on the other side were employed and a stated preference experiment was developed and included in the survey.

The following descriptions and data analyses focus on the mode choice experiment exclusively. The route choice experiment and recording results will be presented in the future. Readers who are interested in the results are encouraged to contact the research team.

3.1. Recruitment Strategy

Table 1 presents an overview of the frame population and recruitment strategy of the survey study, which was split and specifically designed for BSS users and non-users. BSS users, on the one side, were recruited electronically either before or after renting a bike from VRNnextbike. During the electronic check-in or -out of the rented bike in the VRNnextbike smartphone app, they have presented an invitation to participate in a computer-assisted telephone interview (CATI) on their travel behaviour, bike-rent habits, and the BSS-trip they performed either before or after the recruitment. To participate, they were asked to mention their preferred daytime for a 30-min telephone interview within the next week and, in addition, to report their age, gender, and city of residence in an electronic recruitment questionnaire. Questions on socio-demographics aimed to keep control over the distribution of socio-demographic characteristics in the CATI sample. Participation in the electronic recruitment questionnaire took about two minutes on a smartphone. Accompanying the questions, information on data protection and the aim of the research project was provided on linked websites and an incentive of EUR 20 for participation in the CATI was mentioned. After participants completed the CATI, they were recruited for the subsequent paper-and-pencil questionnaire, which was sent via postal mail to the respondents. Enhancing convenient participation was the aim of this study, ensured by sending an addressed and postpaid mail-back envelope together with the questionnaire, including additional information on the study and data protection.

Table 1. Frame population and recruitment strategy.

	Subsample 1: BSS-Users	Subsample 2: BSS Non-User
Frame-population	Users of BSS VRNnextbike	BSS non-user residents from VRNnextbike supply area
Recruitment	Via the nextbike app at the start or end of bike-rent	Recruitment call via telephone
Survey mode	Computer-assisted telephone interviews (CATI) and paper-and-pencil questionnaire with a postpaid return envelope	Paper-and-pencil questionnaire with a postpaid return envelope
Incentive	EUR 20 after participation in the CATI and again after returning the filled-out questionnaire	EUR 20 after returning the filled-out questionnaire

BSS non-users, on the other side, were recruited from a sample of randomly generated phone numbers for the supply area of VRNnextbike. These phone numbers might in some cases have resulted in interviews with BSS-users; however, this approach was chosen as only 2.9% of all inhabitants in the supply area are BSS-users and sampling a user of VRNnextbike was for this reason rather unlikely; furthermore, a control questions whether respondents are BSS users was added to allow the identification in the analysis. Recruitment of this subsample was done via telephone by asking for gender, age, and city of residence to keep control over the distribution of socio-demographic characteristics. In addition, the telephone recruitment aimed to provide information on the study, and data protection issues and to discuss potential respondents’ questions. In addition, an incentive of EUR 20 was offered for a filled-out survey instrument. Like in the sample for BSS users, the printed paper-and-pencil questionnaire was sent via postal mail to the respondents, which included an addressed and postpaid mail-back envelope.

3.2. Revealed and Stated Preferences: Survey Protocol

Choices on, e.g., transport modes and routes can be observed either in real-life situations or in hypothetical choice situations. Real-life observations result in information on revealed preferences (RP) whereas observations of hypothetical choices result in data

on stated preferences (SP). RP data have the advantage of representing peoples' actual behaviour; they, however, also include disadvantages such as little variation in attributes, which could be used to explain choices (e.g., travel times by bus on a specific route), and issues of multicollinearity (e.g., travel times, distances, and costs are often highly correlated for a specific transport mean). In addition, future effects, demands, and supplies cannot be addressed with RP data. SP and according to hypothetical choices overcome these issues. Here, an experimental design is employed, that focuses on selected and potentially future attributes. It asks respondents to exclusively consider these selected effects when making their choices. Attribute variation is controlled by an experimental design, which allows overcoming the challenge of multicollinearity. The disadvantages are, however, the hypothetical character of the choice tasks and the reduction of complexity (for a more detailed discussion on RP- and SP choices see, e.g., Refs. [19,28,29]).

In transport planning, RP and SP data are often combined to overcome the mentioned limitations of the SP approach. One way to increase the reliability of SP choices is to employ an individual's RP decisions as the basis for her or his SP-choice situations [30] (for a discussion of combined RP-SP studies see Ref. [19]). This means, on the one hand, an increased complexity for the fieldwork, as choice situations are individually tailored for each participant of the SP survey. On the other hand, the RP-SP combination increases the quality of a survey as the choice situations are based on former choices of a respondent, transport familiarity and thus allow an easy imagination of the choice situation under observation. The more realistic and familiar a choice situation is, the less effort it takes to be contextualized resulting in more reliable respondents' answers (for choice situations see Ref. [31]; for a more general discussion on response burden see Ref. [32]).

In the CATI for BSS-users, information on mobility-tool ownership, trip characteristics of the BSS-use at the time of recruitment, attitudes on BSS, and socio-demographics were collected. During the interview, the chosen route of the rental bike trip was traced electronically by employing an online routing tool [33] to gather additional information for the trip, such as travel time and road surface. The collected information and RP data were employed as a basis for the mode choice experiment. Based on these BSS attributes, trip characteristics for the alternative modes PT and PMT were collected using an online routing provider [34] and an electronic PT-schedule service [35]. The choice experiment itself was designed as a follow-up survey and presented to those CATI participants, who agreed in filling out the paper-and-pencil questionnaire.

As RP data were not available for BSS non-users, aggregated RP characteristics on BSS usage were employed as the basis for the SP experiment. The aggregated figures for travel time were calculated on the automatically recorded data from the BSS VRNnextbike by discriminating between short, middle, and long trips for both major and minor cities (on average trips of the length of 0.8, 1.5, and 3.2 km for major cities, and 0.7, 1.4, and 4.4 km for minor cities), while averages for access and egress times were obtained from the BSS user-survey. In this case, data on mode alternatives were based on aggregated figures from the national travel survey in Germany [36], taking into account differences between major and minor cities. To design an individual questionnaire for this subsample, firstly, each respondent was assigned to a town size group based on his or her postal address. Secondly, every participant was sequentially assigned to a short, middle, or long trip distance and a trip purpose, either a leisure or mandatory activity at the trip destination. The resulting RP-values for each BSS non-user were employed as the basis for SP-experiment and its variations of attributes.

The recruitment procedure of both subsamples is visualized in Figure 3.

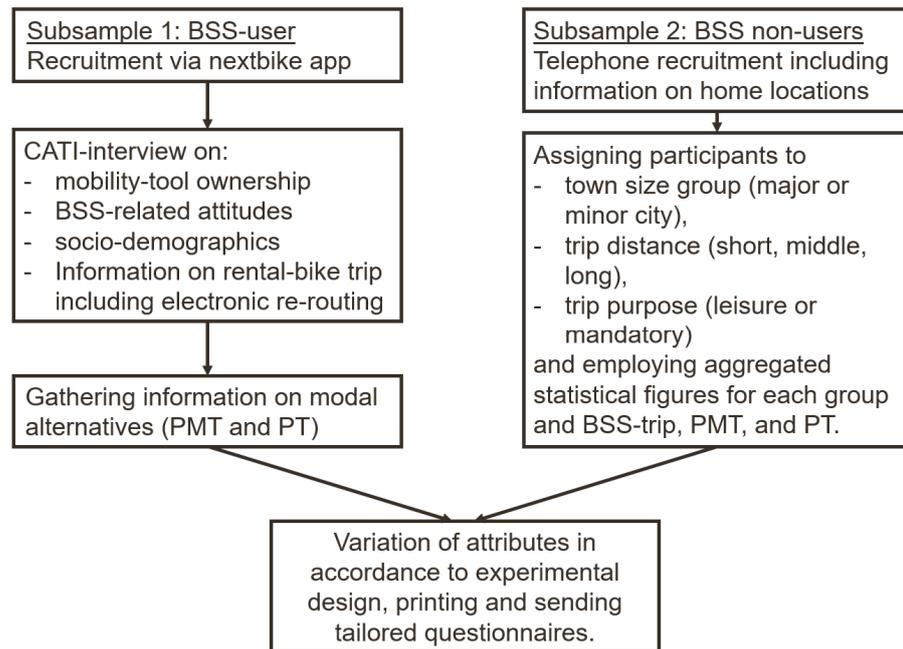


Figure 3. Schema of survey protocol.

3.3. Experimental Design and Choice Situations

Attributes of the mode choice experiment were taken from former studies on mode choices and studies on selected transport modes such as bikes, PMT, and PT (see Literature Review in Section 2). Furthermore, they were discussed with external academic partners and practitioners from transport planning offices.

As described above, RP information (for BSS users) or aggregated empirical figures (for BSS-non-users) were employed and empirical values for the alternative modes were collected. Next, an individually tailored SP questionnaire was created by varying the mode-specific characteristics in accordance with a predefined experimental design. An overview of the attributes and variation of attribute levels is provided in Table 2.

Table 2. Stated preference experiment: transport modes, attributes and variation.

Mode	Attribute	Variation of Reference Values/Levels
BSS	access and egress time	−50%/−10%/+40%
	travel time (TT)	−30%/−10%/+30%
	travel costs	−100%/−35%/+20%
	street type	cycleway/side street/arterial road
	surface type	asphalt/cobblestones/macadam
PMT	travel time (TT) incl. parking search	−20%/−10%/+30%
	fuel costs	−50%/+150%/+200%
	parking costs	−50%/+100%/+200%
PT	access and egress time	−50%/−10%/+40%
	travel time (TT)	−30%/−10%/+30%
	travel costs	−100%/−35%/+20%
	utilized capacity	middle/high/overloaded

To reduce the number of possible combinations of the presented variation levels, an efficient design [37] was generated with the software Ngene [38]; it resulted in 60 combinations to design the choice tasks, which were split into six blocks (the experimental design is available upon request). Each participant was assigned to one block and the empirical values were varied accordingly. Finally, each participant was asked to complete ten choice

tasks in the paper-and-pencil questionnaire. Participation in the study was restricted to adults (18 years and older) owning a driver’s license to make the alternative PMT realistic.

The individually tailored paper-and-pencil questionnaires were created, printed and sent within one week after recruitment to avoid fatigue effects. To increase reliability in terms of PT utilized capacity, an illustration accompanied the questionnaire (see Figure 4; for more details on reliability see Ref. [19]).

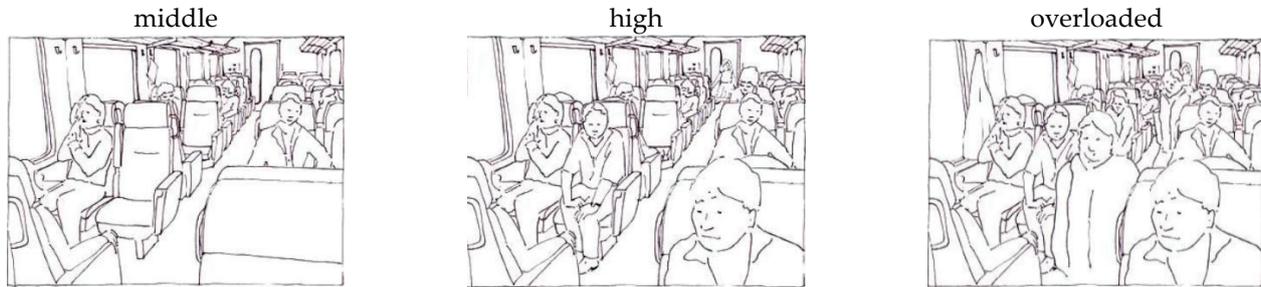


Figure 4. Illustration of capacity utilization (illustrations taken from Weis et al. [19]).

Both the survey protocol and instrument followed the suggestions by Dillmann [39]. After four weeks of non-response, reminders were sent with a new copy of the questionnaire. An exemplary choice situation is shown in Figure 5.

Shared bike		Car		Public transport	
• Access & egress time	3 min			• Access & egress time	2 min
• Travel time	8 min			• Travel time	4 min
Total travel time	11 min	Total travel time	11 min	Total travel time	8 min
Travel cost	1.80 €	Fuel costs	0.70 €	Travel cost	1.70 €
		Parking costs (€/hour)	4.80 €		
Street type	Arterial road			Utilized capacity	High
Surface type	Asphalt				
<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>	
Shared bike		Car		Public transport	

Figure 5. Example of a choice task.

4. Descriptive Statistics and Modelling Approach

The survey was in the field from September 2021 to February 2022. After data cleaning, information from 220 respondents, who filled out and returned the questionnaire, was collected. On average, respondents answered 9.93 (median = 10) mode choice tasks, which resulted in a total of 2184 observations for the analysis. 27 respondents (12.3%) showed non-trading behaviour, meaning they chose an identical transport mode in all presented choice tasks.

4.1. Descriptive Statistics

Respondents from the subsample recruited via randomly generated phone numbers (see subsample 2 in Table 3), who indicated to use a BSS, are considered BSS-users in the following analytical procedure ($n = 11$). Information on the distribution of selected

socio-demographic characteristics and the frequency of chosen transport modes in the choice experiment for both subsamples and the whole sample are presented in Table 3.

Table 3. Relative frequency distribution of selected sample characteristics.

Character	BSS-Users	BSS Non-User	Whole Sample
Sample size	128 (58%)	92 (42%)	220
Gender			
Male	64.1%	41.3%	54.5%
Female	35.9%	58.7%	45.5%
Age			
18–30 yr	71.1%	14.1%	47.3%
31–65 yr	28.9%	59.8%	41.8%
66–94 yr	0%	26.1%	10.9%
Home Municipality			
City	89.1%	64.1%	78.6%
Larger town	9.4%	25.0%	15.9%
Smaller town	1.6%	10.9%	5.5%
Chosen alternative			
BSS	760 (59.6%)	373 (41.1%)	1133 (51.9%)
PMT	111 (8.7%)	333 (36.7%)	444 (20.3%)
PT	405 (31.7%)	202 (22.2%)	607 (27.8%)

Around 58% of all participants were BSS users and 42% are non-users. Users were more often males (64.1%) than non-users (41.3%). For the whole sample, the gender proportion was more balanced with 54.5% males and 45.5% females. BSS users belonged remarkably more often to younger age groups than non-users. Again, the proportion between young adults (18–30 years; 47.3%) and middle-aged persons (31–65 years; 41.8%) was more balanced for the whole sample. There are only a few observations of people in the retired age group (66–94 years): 0% for BSS-users, 26% for non-users and around 11% for the whole sample. Furthermore, around 89% of the user-sample lives in cities, 9% in larger towns and around 2% in small municipalities. This fits well with the automatically tracked renting numbers of VRNnextbike [40]. The non-user sample includes more respondents from larger towns (25%) and small municipalities (around 11%). In summary, there is socio-demographic variation in the data, and it can be assumed that evaluations from people with different socio-demographic characteristics are considered in the analysis.

Concerning mode choice situations, respondents from the user sample most often chose the BSS (around 60%), followed by PT (around 32%) and PMT (9%). Non-users also preferred the BSS (64%), however, followed by PMT (37%) and PT (22%).

Covariates in the behavioural experiment are based on RP-data for subsample 1, respondents recruited during the CATI, and on aggregated empirical figures for subsample 2, respondents recruited from the random phone number sample. Table 4 presents the empirical distribution of these covariates as included in the experiment.

Table 4. Empirical distribution of covariates in a choice experiment.

Mode	Attribute	min	1st Qu.	Median	Mean	3.rd Qu.	Max
BSS	access and egress time [min]	1.0	4.0	6.0	6.3	10.0	49.0
	travel time [min]	1.0	6.0	10.0	12.4	15.0	108.0
	travel costs [EUR]	0.0	0.0	1.0	1.2	1.8	10.8
PMT	travel time incl. parking [min]	6.0	12.0	15.0	15.8	19.0	46.0
	fuel costs [EUR]	0.1	0.2	0.6	0.7	1.2	6.2
	parking costs [EUR]	0.6	0.8	3.2	2.7	3.3	4.8
PT	access and egress time [min]	0.0	4.0	6.0	7.0	9.0	41.0
	travel time [min]	1.0	6.0	9.0	10.5	14.0	64.0
	travel costs [EUR]	0.0	0.0	1.8	1.8	2.9	7.3

In terms of the BSS, access and egress time has a minimum of 1 min, a median of 6 min, a mean of around 6 min and a maximum of 49 min. This distribution is comparable to access and egress for PT. BSS trips are short with a median travel time of 10 min and 15 min for the 3rd quartile. A few trips, however, are long with a maximum of 108 min. Travel costs are low, with EUR 1.80 for the 3rd quartile.

The overall travel time (including parking search time) of PMT is substantially higher than the travel time for BSS and PT. The minimum travel time is 6 min, 1st quartile at 12 min, median at 15 and 3rd quartile at 19 min. The maximum travel time, however, is 46 min and substantially lower than the maximum travel times for BSS and PT. Fuel costs for the trip show a range between EUR 0.10 for a very short trip and EUR 6.20. Parking costs lay between EUR 0.60 and EUR 4.80 with a median of EUR 3.20.

Access and egress times and travel times for PT are overall comparable to BSS. In terms of travel costs, PT shows moderate values between BSS and PMT.

4.2. Model Formulation

Discrete choice data, where respondents choose between a limited number of alternatives, are commonly analyzed by applying random utility maximization theory. The theory assumes rational behaviour in which respondents choose the alternative with the highest utility [29,41–43]. Namely, an individual n faced with J alternatives in T choice tasks associates an indirect utility U_{njt} for an alternative j in a choice task t and chooses the alternative with the highest utility. The utility of an alternative j is therefore decomposed as

$$U_{njt} = V_{njt} + \varepsilon_{njt} = x'_{njt}\beta + \varepsilon_{njt} \quad (1)$$

where U_{njt} is not observed, but V_{njt} is the deterministic utility of alternative j , and ε_{njt} is a random component not included in V_{njt} . The deterministic utility V_{njt} can be specified by the term $x'_{njt}\beta$, where x is a vector of explanatory variables (e.g., attribute levels), and β is the corresponding coefficients to be estimated.

For each alternative, a utility function (V_{njt}) is specified, whereby the alternative-specific attributes, characteristics of the respondent or the choice situation are included as explanatory variables. When specifying the utility function, it is important to understand that only the differences in utility matter, while the scale of utility is arbitrary [29] (p. 19). Therefore, to capture the differences in the utility of the alternatives, $J-1$ alternative-specific constants (ASC) are specified, whereby the estimated ASCs are interpreted relative to the omitted alternative, which is normalized to zero [29,43]. For the categorical attributes, street type, surface type, and utilized capacity, the L levels of each attribute were transformed into $L - 1$ dummy variables. This means, the utility for one level per attribute is normalized to zero and serves as a reference category, while the parameter estimates for the $L - 1$ dummy variables capture the utility differences to this reference category [28,29,43].

5. Results

The 2184 observations (choice tasks) were analyzed by estimating multinomial logit models (MNL) [44] in R [45,46], whereby BSS was chosen as reference alternative when specifying the equations for estimation. Firstly, an initial MNL was estimated by including exclusively effects of attributes from the choice experiment (see Table 2; for a documentation of this work see Ref. [47]). With reference to previous studies on mode choice [18–21] effects of socio-demographics (age, gender, education, student status, car availability, PT season ticket availability), home municipality, and season (winter vs. autumn) were expected. Consequently, as recommended in methodological literature [29,44] the initial model was sequentially built up by including these effects as alternative-specific attributes in maximum $J-1$ alternatives (one alternative as reference category), testing the hypotheses, and comparing the models (restricted vs. unrestricted) to omit parameters without significant effects and/or substantial improvement in the model fit. Further, it was assumed that the effects of travel time and travel costs depend on household income and the distance of the trip, and this is why corresponding continuous interactions were specified [18,19]; however,

these interactions neither had a significant effect nor made a substantial improvement of the model and thus are not presented. All analytical steps along with estimated models can be made available on request. Table 5 provides an overview on model fit between the initial model, which exclusively included effects from attributes of the choice experiment, and the extended, final model, which is presented below. The likelihood ratio-test indicates a significantly better fit for the final model, which is supported by the increase in adjusted Rho-square, and by the decrease in AIC and BIC (for evaluation of model fit indices and model comparison, please review methodological literature, Refs. [29,44]).

Table 5. Model fit comparison between the initial and the final model.

Model Indices	Initial Model	Final Model
n estimated parameters	12	33
LL(0)	−2399.37	−2399.37
LL(final)	−1639.5	−1468.97
Adj. Rho-square	0.312	0.374
AIC	3303	3004
BIC	3371	3192
Likelihood ratio test		336.62; $p < 0.01$

Employing the estimated parameters in transport demand models at a later stage of the project requires a distinction between mandatory and leisure trips. Mandatory trips are those with destinations for purposes such as education, work, business, or home. Leisure trips are those with destinations such as shopping, private activities and tasks, or any leisure activities. Usually, people have more degrees of freedom in destination choice for leisure than for mandatory activities. Therefore, the trip purpose was included as alternative-specific attribute in the final model on the total sample. In addition, segregated models were estimated on a subsample for mandatory and a subsample for leisure trips. The results of all three models, the overall (total) model on all observations, and the segregated model on mandatory and leisure trips, are presented in Table 6.

All parameters for the final model show the expected sign and reasonable differences in parameter values. For all modes (BS, PMT, and PT), the estimated parameters for travel time and time for access and egress show a negative effect. Hereby the negative effect is stronger for access and egress, which was expected as ride-times in or on a vehicle are often considered less negatively than waiting times or access and egress-times [19]. Travel costs demonstrate a negatively associated utility for all modes. In addition, parking costs and fuel costs for PMT are negative, too.

For BS, the data do not support any significant difference in utility for the street type; however, with reference category arterial road, the cycleway has a higher positive estimated utility ($\beta = 0.195$, t -value = 1.339) and the side street has a negative utility ($\beta = -0.013$, t -value = 0.089). This negative utility of side streets can be explained with a detour-association in comparison to the probably more direct and thus shorter route on an arterial road. Relatively to macadam surface, cobblestones do not show differences in utility ($\beta = 0.004$, t -value = 0.028), while asphalt is a more preferred surface type; however, the effect is also not significant ($\beta = 0.239$, t -value = 1.634).

For PMT and PT, the estimated ASCs show the differences in utility of a given alternative from the reference BS when everything else is equal [44]. The utility of PMT is higher than for BS ($\beta = 0.669$, t -value = 1.059), whereby the direction of the effect changes when comparing mandatory trips ($\beta = -1.677$, t -value = 1.389) to leisure trips ($\beta = 0.980$, t -value = 1.233). This can be explained with the high share of commuters and students mainly using the system for trips to work and education. For these people, BS has a higher utility as PMT. This interpretation is also supported by the negative sign for mandatory trips in the overall (total) model ($\beta = -0.608$, t -value = -3.739). In addition, PT has a higher positive utility than BS, too ($\beta = 0.756$, t -value = 1.552), whereby there is no change in sign between mandatory and leisure trips. Influences from the spatial typology are limited to PMT, where the utility for PMT decreases with an increasing size of the home municipality.

For PT, a mode-specific effect results from capacity utilization. An increasing utilization results in a decreasing utility for PT.

Table 6. Results of MNL models for all observations (total) and trip specific models.

	Parameter	Total			Mandatory			Leisure		
		β	s.e.	t-Value	β	s.e.	t-Value	β	s.e.	t-Value
BS (Ref.)	travel time [min]	-0.108	0.009	-11.637	-0.127	0.017	-7.674	-0.105	0.012	-9.039
	access and egress [min]	-0.199	0.017	-11.938	-0.211	0.026	-8.250	-0.205	0.023	-8.733
	travel costs [EUR]	-0.689	0.054	-12.778	-0.880	0.091	-9.652	-0.586	0.068	-8.557
	street type (Ref.: arterial)									
	side street	-0.013	0.146	-0.089	-0.030	0.243	-0.124	0.071	0.187	0.380
	cycleway	0.195	0.145	1.339	0.337	0.243	1.387	0.102	0.185	0.552
	surface type (Ref.: macadam)									
	cobblestones	0.004	0.143	0.028	0.322	0.231	1.395	-0.142	0.187	-0.757
	asphalt	0.239	0.146	1.634	0.377	0.234	1.610	0.208	0.192	1.087
	PMT	ASC	0.669	0.632	1.059	-1.677	1.207	-1.389	0.980	0.795
travel time [min]		-0.116	0.014	-8.299	-0.085	0.028	-3.059	-0.111	0.017	-6.607
fuel costs [EUR]		-0.474	0.129	-3.684	-0.325	0.244	-1.336	-0.530	0.158	-3.346
parking costs [EUR]		-0.532	0.048	-11.037	-0.717	0.094	-7.587	-0.489	0.058	-8.380
age		-0.065	0.023	-2.829	-0.002	0.049	-0.042	-0.085	0.029	-2.941
age squared		0.001	0.000	3.722	0.000	0.001	-0.034	0.001	0.000	3.739
female (Ref.: male)		0.745	0.150	4.976	1.541	0.297	5.192	0.510	0.187	2.724
BS user (Ref.: non-user)		-0.606	0.190	-3.194	-0.743	0.487	-1.524	-0.807	0.224	-3.608
car available always (Ref.: sometimes/never)		0.627	0.205	3.056	0.874	0.344	2.538	0.718	0.272	2.638
PT season ticket (Ref.: none)		-0.494	0.171	-2.897	-0.108	0.337	-0.321	-0.644	0.215	-2.998
winter (Ref.: autumn)		0.360	0.162	2.230	1.209	0.348	3.480	0.215	0.188	1.144
home municipality (Ref.: small)										
larger town		-1.318	0.263	-5.010	-2.222	0.562	-3.956	-1.111	0.310	-3.586
city		-1.062	0.240	-4.429	-1.561	0.525	-2.972	-0.866	0.282	-3.066
trip mandatory (Ref.: leisure)		-0.608	0.163	-3.739						
PT	ASC	0.756	0.487	1.552	1.080	0.791	1.365	0.413	0.652	0.634
	travel time [min]	-0.101	0.012	-8.382	-0.136	0.021	-6.410	-0.093	0.015	-6.088
	access and egress [min]	-0.208	0.017	-12.014	-0.254	0.027	-9.572	-0.174	0.024	-7.184
	travel costs [EUR]	-0.779	0.049	-15.885	-0.900	0.079	-11.369	-0.737	0.065	-11.391
	capacity (Ref.: middle)									
	high	-0.498	0.141	-3.536	-0.501	0.226	-2.214	-0.509	0.185	-2.752
	overloaded	-1.147	0.147	-7.775	-1.115	0.226	-4.936	-1.221	0.201	-6.068
	age	-0.056	0.021	-2.694	-0.046	0.038	-1.187	-0.048	0.027	-1.787
	age squared	0.001	0.000	3.073	0.001	0.000	1.246	0.001	0.000	2.177
	female (Ref.: male)	-0.109	0.129	-0.840	0.404	0.206	1.958	-0.464	0.175	-2.657
	PT season ticket (Ref.: none)	0.433	0.134	3.241	0.164	0.205	0.803	0.600	0.184	3.268
	winter (Ref.: autumn)	0.643	0.149	4.322	0.832	0.265	3.143	0.573	0.188	3.045
trip mandatory (Ref.: leisure)	0.052	0.133	0.390							
n individuals		220			96			124		
n choice tasks		2184			954			1230		
LL(0)		-2399.37			-1048.08			-1351.29		
LL(final)		-1469.00			-530.55			-907.30		
Adj. Rho-square		0.374			0.464			0.306		

In terms of socio-demographics, the effect of age and age-squared shows a u-shaped distribution of utility for both, PMT and PT (see Figure 6). Choosing PMT has a negative utility from 18 to 64 years, whereby the smallest value is reached between 32 and 33 years. From this age on the utility of PMT increases again. A somehow similar picture is observed for PT, where PT has a negative utility in comparison to BS between 18 and 55 years. The lowest utility is calculated for an age of 28 years. From this age on the utility of PT increases in comparison to BS.

For women, PMT has a higher utility than BS ($\beta = 0.745$, t-value= 4.976). This effect is different for PT, where the utility for women is negative ($\beta = -0.109$, t-value= 0.840). This pattern fits the results of other studies, which show that women appreciate the privacy of cars (for a general discussion on car use and gender see, e.g., Ref. [48]) and perhaps BSS in comparison to PT.

Furthermore, in comparison to BS, having a car always available increases the use of PMT ($\beta = 0.627$, t-value = 3.056), while owning a PT season ticket increases the utility of PT ($\beta = 0.433$, t-value = 3.241) and decreases the utility of PMT ($\beta = -0.494$, t-value = -2.897). In winter, both, PMT ($\beta = 0.360$, t-value = 2.230) and PT ($\beta = 0.643$, t-value = 4.322) are more preferred than BS.

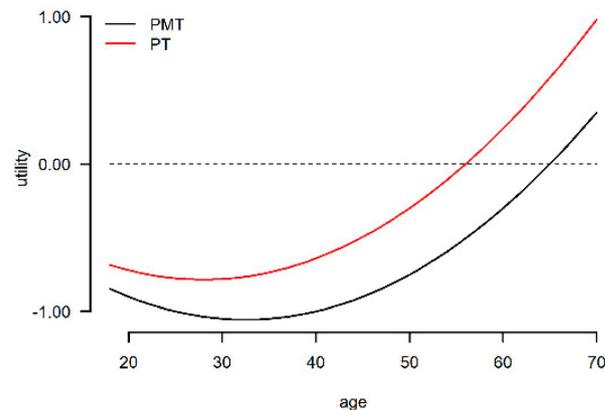


Figure 6. Example of a choice task.

6. Conclusions

The survey resulted in behavioral parameters, which show the expected signs and allow a straightforward interpretation. It has to be kept in mind that the survey was in the field between September 2020 and February 2021. In this rather cold season of the year, BSS-using figures are low, and it can be assumed that the share of experienced users is overrepresented, whilst occasional users are underrepresented in comparison to the warmer season. This, however, does not necessarily lead to bias in the data.

In addition, our study has a regional character. Topographically seen, the supply area of VRNnextbike is rather flat with some hills. Mountains and large altitudinal differences are rare if present at all. Even though altitude was not considered in the choice experiment, there is a correlation with travel time and with travel time related costs; this has to be considered when statistical results are employed in other regions.

In general, results can be used to implement BSS in transport demand models. The main empirical findings are:

- All parameters for the final model show the expected sign and reasonable differences in parameter values.
- With the reference category arterial road, the cycleway has a higher positive estimated utility and the side street has a negative utility (although both effects are not significant).
- In terms of socio-demographics, the non-linear effect of age shows a u-shaped distribution of utility for both, PMT and PT.
- For women, PMT has a higher utility than BS. This effect is different for PT, where the utility for women is negative.
- Having a car always available increases the use of PMT, while owning a PT season ticket increases the utility of PT and decreases its utility.
- In winter, both, PMT and PT are more preferred than BS.

Analyses, however, are not finished yet. Future work will be on a calculation of willingness to pay values (WTP) as well as values for travel time savings (VTTS). These values will allow a comparison to similar studies for PT and PMT and will show to what extent the above presented results are similar and reasonable. In addition, parameters for route choices have to be estimated. Once this is done, BSS-parameters will be implemented in an existing regional transport demand model and three scenarios will be simulated:

1. Lower access and egress times for BSS. A scenario where stations are more densely distributed in the research area and therefore the use of BSS becomes more comfortable;
2. Lower quality of PMT-supply. In this scenario, travel time, parking search time, and parking costs are increased to analyze effects on modal shift from PMT to PT and BSS;
3. Better quality in PT. Access and egress times for PT are reduced in this scenario and potential effects in terms of modal shift on PMT and BSS will be analyzed.

4. Estimation of route-choice parameters, implementation of BSS in a transport demand model and calculation of modal shifts along the above-mentioned scenarios will be documented later and published elsewhere. The present work, however, represents one necessary next step for a better understanding of a good established transport mode in cities and urban areas. In terms of data collection and analysis, it would be good to combine survey data on cycling in general with sensor-based data on, e.g., cycling safety [49].

Author Contributions: Conceptualization and methodology: M.K., T.D., C.S.; writing: M.K.; data acquisition: K.R., M.G., M.K.; data analysis: M.G.; literature review: K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the state of Hessian and the House of Logistics (HOLM) in the program “Innovations in logistics and mobility” of the Hessian Ministry for Economic Affairs, Energy, Transport and Housing, grant number HA-No. 1013/21-15. Open Access funding provided by the Open Access Publication Fund of RheinMain University of Applied Sciences.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The survey data can be obtained from the corresponding author. Interested researchers are asked to send a data request via e-mail.

Acknowledgments: This project (HA-No. 1013/21-15) is funded by the state of Hessian and the House of Logistics (HOLM) in the program “Innovations in logistics and mobility” of the Hessian Ministry for Economic Affairs, Energy, Transport and Housing.



HessenAgentur

HA Hessen Agentur GmbH

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Fishman, E. Bikeshare: A Review of Recent Literature. *Transp. Rev.* **2016**, *36*, 92–113. [CrossRef]
2. Fishman, E.; Washington, S.; Haworth, N. Bike Share: A Synthesis of the Literature. *Transp. Rev.* **2013**, *33*, 148–165. [CrossRef]
3. Ricci, M. Bike sharing: A review of evidence on impacts and processes of implementation and operation. *Res. Transp. Bus. Manag.* **2015**, *15*, 28–38. [CrossRef]
4. Si, H.; Shi, J.; Wu, G.; Chen, J.; Zhao, X. Mapping the bike sharing research published from 2010 to 2018: A scientometric review. *J. Clean. Prod.* **2019**, *213*, 415–427. [CrossRef]
5. Teixeira, J.F.; Silva, C.; Moura e Sá, F. Empirical evidence on the impacts of bikesharing: A literature review. *Transp. Rev.* **2021**, *41*, 329–351. [CrossRef]
6. Campbell, A.A.; Cherry, C.R.; Ryerson, M.S.; Yang, X. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transp. Res. Part C Emerg. Technol.* **2016**, *67*, 399–414. [CrossRef]
7. Reck, D.J.; Haitao, H.; Guidon, S.; Axhausen, K.W. Explaining shared micromobility usage, competition and mode choice by modelling empirical data from Zurich, Switzerland. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102947. [CrossRef]
8. Reck, D.J.; Martin, H.; Axhausen, K.W. Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transp. Res. Part Transp. Environ.* **2022**, *102*, 103134. [CrossRef]
9. Wiegel, A. Unser Auto—Kann das weg? Die Zukunft der Mobilität: Verkehrsnachfragemodell ermittelt Verlagerungspotenziale auf umweltfreundlichere Angebote. *Nahverkehr* **2021**, *39*, 66–67.
10. Buehler, R.; Dill, J. Bikeway Networks: A Review of Effects on Cycling. *Transp. Rev.* **2016**, *36*, 9–27. [CrossRef]
11. Forsyth, A.; Krizek, K.J. Promoting Walking and Bicycling: Assessing the Evidence to Assist Planners. *Built Environ.* **2010**, *36*, 429–446. [CrossRef]

12. Wang, Y.; Chau, C.K.; Ng, W.Y.; Leung, T.M. A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighborhoods. *Cities* **2016**, *50*, 1–15. [CrossRef]
13. Banerjee, S.; Kabir, M.; Khadem, N.K.; Chavis, C. Optimal locations for bikeshare stations: A new GIS based spatial approach. *Transp. Res. Interdiscip. Perspect.* **2020**, *4*, 100101. [CrossRef]
14. Croci, E.; Rossi, D. *Optimizing the Position of Bike Sharing Stations. The Milan Case*; (SSRN Scholarly Paper No. ID 2461179); Social Science Research Network: Rochester, NY, USA, 2014. [CrossRef]
15. Martinez, L.M.; Caetano, L.; Eiró, T.; Cruz, F. An Optimisation Algorithm to Establish the Location of Stations of a Mixed Fleet Biking System: An Application to the City of Lisbon. *Procedia-Soc. Behav. Sci.* **2012**, *54*, 513–524. [CrossRef]
16. Hamre, A.; Buehler, R. Commuter Mode Choice and Free Car Parking, Public Transportation Benefits, Showers/Lockers, and Bike Parking at Work: Evidence from the Washington, DC Region. *J. Public Transp.* **2014**, *17*, 67–91. [CrossRef]
17. Lu, M. *RP and SP Data-Based Travel Time Reliability Analysis*; ETH Zurich: Zürich, Switzerland, 2013. [CrossRef]
18. Axhausen, K.W.; Hess, S.; König, A.; Abay, G.; Bates, J.J.; Bierlaire, M. Income and distance elasticities of values of travel time savings: New Swiss results. *Transp. Policy* **2008**, *15*, 173–185. [CrossRef]
19. Weis, C.; Kowald, M.; Danalet, A.; Schmid, B.; Vrtic, M.; Axhausen, K.W.; Mathys, N. Surveying and analysing mode and route choices in Switzerland 2010–2015. *Travel Behav. Soc.* **2021**, *22*, 10–12. [CrossRef] [PubMed]
20. Ton, D.; Duives, D.C.; Cats, O.; Hoogendoorn-Lanser, S.; Hoogendoorn, S.P. Cycling or walking? Determinants of mode choice in the Netherlands. *Transp. Res. Part Policy Pract.* **2019**, *123*, 7–23. [CrossRef]
21. Ilahi, A.; Belgiawan, P.F.; Balac, M.; Axhausen, K.W. Understanding travel and mode choice with emerging modes; a pooled SP and RP model in Greater Jakarta, Indonesia. *Transp. Res. Part Policy Pract.* **2021**, *150*, 398–422. [CrossRef]
22. Hardinghaus, M.; Papantoniou, P. Evaluating Cyclists' Route Preferences with Respect to Infrastructure. *Sustainability* **2020**, *12*, 3375. [CrossRef]
23. Clark, C.; Mokhtarian, P.; Circella, G.; Watkins, K. User preferences for bicycle infrastructure in communities with emerging cycling cultures. *Transp. Res. Rec.* **2019**, *2673*, 89–102. [CrossRef]
24. Vedel, S.E.; Jacobsen, J.B.; Skov-Petersen, H. Bicyclists' preferences for route characteristics and crowding in Copenhagen—A choice experiment study of commuters. *Transp. Res. Part Policy Pract.* **2017**, *100*, 53–64. [CrossRef]
25. Stinson, M.A.; Bhat, C.R. Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey. *Transp. Res. Rec. J. Transp. Res. Board* **2003**, *1828*, 107–115. [CrossRef]
26. Sener, I.N.; Eluru, N.; Bhat, C.R. An analysis of bicycle route choice preferences in Texas, US. *Transportation* **2009**, *36*, 511–539. [CrossRef]
27. Hunt, J.D.; Abraham, J.E. Influences on bicycle use. *Transportation* **2007**, *34*, 453–470. [CrossRef]
28. Louviere, J.J.; Hensher, D.A.; Swait, J.D. *Stated Choice Methods—Analysis and Application*; Cambridge University Press: Cambridge, UK, 2000.
29. Train, K.E. *Discrete Choice Methods with Simulation*; Cambridge University Press: Cambridge, UK, 2009.
30. Hensher, D.A. Stated preference analysis of travel choices: The state of practice. *Transportation* **1994**, *21*, 107–133. [CrossRef]
31. Camerer, C.; Mobbs, D. Differences in behaviour and brain activity during hypothetical and real choices. *Trend. Cognit. Sci.* **2017**, *21*, 46–56. [CrossRef]
32. Groves, R.M. *Survey Errors and Survey Costs*; Wiley: Hoboken, NJ, USA, 2004.
33. Komoot. Available online: <https://www.komoot.de/> (accessed on 4 January 2022).
34. Google Maps. Available online: <https://www.google.com/maps/> (accessed on 1 September 2021).
35. VRN—Verkehrsbund Rhein-Neckar. Available online: <https://www.vrn.de/> (accessed on 5 January 2022).
36. Mobilität in Deutschland 2017 (MiD 2017). Available online: <http://www.mobilitaet-in-deutschland.de/> (accessed on 16 March 2022).
37. Rose, J.M.; Bliemer, M.C.J. Stated choice experimental design theory: The who, the what and the why. In *Handbook of Choice Modelling*; Hess, S., Daily, A., Eds.; Edward Elgar Publishing: Cheltenham, UK, 2014; pp. 152–177.
38. ChoiceMetrics. Ngene 1.2: User Manual & Guide Reference. 2018. Available online: <http://www.choice-metrics.com/download.html#manual> (accessed on 15 March 2022).
39. Dillman, D.A. *Mail and Internet Surveys—The Tailored Design Method*; Wiley: New York, NY, USA, 2000.
40. Pautzke, C.; Kowald, M.; Dannewald, T.; Eireiner, M.; Bles, V. Die Entwicklung des Fahrradvermietsystems VRNnextbike 2015–2021: Monitoringbericht Q2/2021, Arbeitsbericht Fachgruppe Mobilitätsmanagement, Nr. 012. 2021. Available online: https://www.hs-rm.de/fileadmin/Home/Fachbereiche/Architektur_und_Bauingenieurwesen/Studiengaenge/Mobilitaetsmanagement_B.Eng./Publikationen/Arbeitsbericht_Q2_2021_VRNnextbike.pdf (accessed on 15 March 2022).
41. Adamowicz, W.; Louviere, J.; Williams, M. Combining revealed and stated preference methods for valuing environmental amenities. *J. Environ. Econ. Manag.* **1994**, *26*, 271–292. [CrossRef]
42. Louviere, J.J.; Flynn, T.N.; Carson, R.T. Discrete choice experiments are not conjoint analysis. *J. Choice Model.* **2010**, *3*, 57–72. [CrossRef]
43. Mariel, P.; Hoyos, D.; Meyerhoff, J.; Czajkowski, M.; Dekker, T.; Glenk, K.; Jacobsen, J.B.; Liebe, U.; Olsen, S.B.; Sagebiel, J.; et al. *Environmental Valuation with Discrete Choice Experiments: Guidance on Design, Implementation and Data Analysis*; Springer International Publishing: Cham, Switzerland, 2021. [CrossRef]
44. Ben-Akiva, M.; Lerman, S.R. *Discrete Choice Analysis: Theory and Application to Travel Demand*; MIT Press Series in Transportation Studies; MIT Press: London, UK, 1985.

45. Hess, S.; Palma, D. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *J. Choice Model.* **2019**, *32*, 100170. [[CrossRef](#)]
46. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
47. Gutjar, M.; Röth, K.; Kowald, M. A stated preferences experiment to analyze bike sharing as an alternative transport mode. In Proceedings of the 10th Symposium of the European Association for Research in Transportation, Leuven, Belgium, 1–3 June 2022.
48. Polk, M. The influence of gender on daily car use and on willingness to reduce car use in Sweden. *J. Transp. Geogr.* **2004**, *12*, 185–195. [[CrossRef](#)]
49. Murgano, E.; Caponetto, R.; Pappalardo, G.; Cafiso, S.D.; Severino, A. A novel acceleration signal processing procedure for cycling safety assessment. *Sensors* **2021**, *21*, 4183. [[CrossRef](#)] [[PubMed](#)]

Article

Multi-Factor Rear-End Collision Avoidance in Connected Autonomous Vehicles

Sheeba Razzaq¹, Amil Roohani Dar¹, Munam Ali Shah¹, Hasan Ali Khattak^{2,*}, Ejaz Ahmed³, Ahmed M. El-Sherbeeney⁴, Seongkwan Mark Lee⁵, Khaled Alkhaledi⁶ and Hafiz Tayyab Rauf⁷

- ¹ Department of Computer Science, COMSATS University Islamabad, Park Road, Tarlai Kalan, Islamabad 44500, Pakistan; sheeba.razzaq@uokajk.edu.pk (S.R.); amil.rohani@uokajk.edu.pk (A.R.D.); mshah@comsats.edu.pk (M.A.S.)
 - ² School of Electrical Engineering and Computer Science (SEECs), National University of Sciences and Technology (NUST), Islamabad 44500, Pakistan
 - ³ Computer Science Department, National University of Computer and Emerging Sciences (NUCES-FAST), Islamabad 44000, Pakistan; ejaz.ahmed@nu.edu.pk
 - ⁴ Industrial Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia; aelsherbeeney@ksu.edu.sa
 - ⁵ Department of Civil & Environmental Engineering, UAE University, Al-Ain 15551, United Arab Emirates; marklee@uaeu.ac.ae
 - ⁶ Industrial Management and Systems Engineering Department, College of Engineering and Petroleum, Kuwait University, P.O. Box 5969, Safat, Kuwait City 13060, Kuwait; hf.s@ku.edu.kw
 - ⁷ Department of Computer Science, Faculty of Engineering & Informatics, University of Bradford, Bradford BD7 1DP, UK; h.rauf4@bradford.ac.uk
- * Correspondence: hasan.alikhattak@seecs.edu.pk

Citation: Razzaq, S.; Dar, A.R.; Shah, M.A.; Khattak, H.A.; Ahmed, E.; El-Sherbeeney, A.M.; Lee, S.M.; Alkhaledi, K.; Rauf, H.T. Multi-Factor Rear-End Collision Avoidance in Connected Autonomous Vehicles. *Appl. Sci.* **2022**, *12*, 1049. <https://doi.org/10.3390/app12031049>

Academic Editors: Giovanni Randazzo, Anselme Muzirafuti and Dimitrios S. Paraforos

Received: 18 December 2021

Accepted: 13 January 2022

Published: 20 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: According to World Health Organization (WHO), the leading cause of fatalities and injuries is rear-ending collision in vehicles. The critical challenge of the technologically rich transportation system is to reduce the chances of accidents between vehicles. For this purpose, it is especially important to analyze the factors that are the cause of accidents. Based on these factors' results, this paper presents a driver assistance system for collision avoidance. There are many factors involved in collisions in the existing literature from which we identified some factors which can affect the accident occurrence probability. However, with advancements in the technologies of autonomous vehicles, these factors can be controlled using an onboard driver assistance system. We used MATLAB's Fuzzy Inference System Tool to analyze the categories of accident contributing factors. Fuzzy results are validated using the VOMAS agent in the NetLogo simulation model. The proposed system can inform the vehicle's automated system when chances of an accident are higher so that the vehicle may take control from the driver. The proposed research is extremely helpful in handling various kinds of factors involved in accidents. The results of the experiments demonstrated that multi-factor-enabled vehicles could better avoid collision as compared to other vehicles.

Keywords: collision avoidance; fuzzy logic; on board driver assistance; semi-autonomous; multi-factor; VANET

1. Introduction

According to the WHO [1], around 20 to 50 million people suffer from severe injuries in road traffic crashes, with many experiencing disabilities because of their injury. Road traffic injuries were the leading cause of death for children and adults between the ages of 5 and 29 years [1]. To reduce fatalities and injuries from road traffic crashes, the World Health Organization (WHO) acts as a team with partners responsible for technical support to countries. The leading cause of fatalities and injuries is the rear-end collision, which make up 70% of all vehicle collisions [2]. Another report, according to the authors in [3], is that 1.078 million injuries in the USA are only due to rear-end collisions. So, an efficient collision

avoidance system in vehicles is needed to reduce the death rate [1]. In the existing literature, many researchers have proposed solutions for collision avoidance from the rear-end.

The authors in [4] proposed a rear-end collision avoidance controller based on proportional–integral–derivative. Another research was proposed by authors in [5] for vehicle rear-end collision avoidance using the linear quadratic optimal control technique. The problem with these solutions is that these are highly dependent on mathematical models [6]. The problem of mathematical-model-based solutions can be overcome using fuzzy logic [7].

Now we figure out important factors from literature which can be used in fuzzy logic to solve the problem of rear-end collision avoidance. The authors in [8] used the physical, environmental, and mental factors to reduce the chances of accidents (COA). The authors in [9] analyzed and discussed the road and weather condition factors in accident occurrence. The authors in [10] also used the environmental factors, such as road and weather conditions in rear-end crash avoidance. Driver characteristics can be added in reducing accident chances and increasing the flexibility of the algorithm. The authors in [11] proposed an algorithm in which they pass the characteristics of driver in the proposed algorithm and showed a significant improvement. Driver's characteristics are also important in decision making because the warning thresholds can be improved by adding driver experience [12], age, and time of accident, along with the factors that are used in [8]. Different factors can change the results of accident occurrence. The authors in [12] discuss different single factors and multi-factors involved in road accidents. The time of the accident can also play an important role in accident avoidance in all these factors discussed in [12].

The combinations of these factors, as discussed in [8,11,12], can improve decision making while driving. We will use fuzzy logic to check how these factors can increase the chances of accidents. First, we examine whether these fuzzy rules can be verified and validated or not. There will be serious problems, e.g., false warning when there is no need for that, if the fuzzy rules are not properly validated. There is an existing model in this regard that provides the Virtual Overlay for Multi-Agent System (VOMAS), which can test any kind of system for accurate results. VOMAS can be applied using NetLogo tool for the validation purpose of different simulations. The authors in [13] used the VOMAS in their proposed system.

The proposed system requires the output of the Fuzzy system to take actions for accident prevention. The actions can be simulated with the help of NetLogo Tool. This is possible if we fed the input of multi-levels of factors, such as environmental and physical conditions [8], driver [11], and weekday and time [12], into the simulation model.

The contributions of the proposed system are as follows:

- Providing simulation-based solution to the problem highlighted by authors in [12] of multiple factor analysis.
- Use of the Mamdani Fuzzy Inference system to compute the values of factors involved in collisions.
- Highlighting the combination of leading factors involved in rear-end collision.
- The combination of these factors has never been used before for collision avoidance using a Fuzzy system.
- The proposed model can assist drivers during different conditions by switching the control to the vehicle. The simulation will show the switching of control in the simulation section.

2. Related Work

This literature review consists of collision-avoidance-based research work between vehicles using different factors involved. We tried to find out these factors for our research purpose and how these factors help in accident occurrence or collision avoidance. We also tried to figure out tools and techniques used by researchers for this work. In recent times, many researchers have conducted research on collision avoidance between vehicles. Due to an increase in the number of vehicles, it has become a challenge to reduce the deaths in

accidents by vehicles collision [14]. Collision-avoidance-based warning systems [15] have a significant impact on road traffic safety. In the existing literature, many algorithms with the name collision avoidance, collision warning, collision assessment, collision prediction, or collision risk assessment have received massive research in reducing collisions in vehicular ad hoc network (VANET). Figure 1 depicts the VANET.

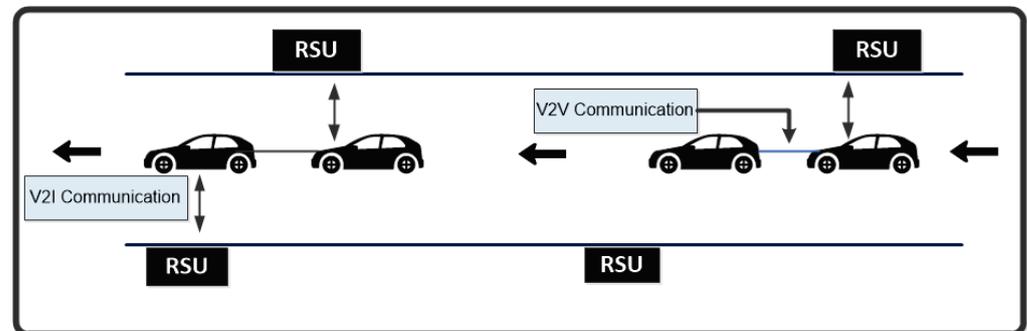


Figure 1. Ad-hoc Network using Vehicles (VANET) [16].

The authors in [11] proposed a safety collision avoidance algorithm. This algorithm takes the characteristics of environment and driver and assign weights to these factors. These characteristics (health, mental index, age, visual acuity, driver age, etc.) are the inputs to the algorithm, and an output is generated in the form of a warning. MATLAB and VISSUM were used for the implementation. The authors said that more experimental data are required for the improvement and optimization of the algorithm. Another collision warning system was proposed by the authors in [17]. In this collision warning model, the authors discussed the impact of weather factors on human-related factors. Their intent was to consider the low visibility factor and proposed a Visibility-based Collision Warning System. MATLAB was used for the implementation and the results. According to authors in [18], a collision warning system is necessary for avoiding collision as depicted in Figure 1. Their proposed model consisted of three steps. They used PreScan commercial software for simulation tests. The proposed system results were better than the time of the collision-based system [19].

The authors in [20] proposed a new methodology for finding crash risk. They studied the driver factors and time factors involved in accidents from police reports recorded during 2002–2012 in Great Britain. They found that drivers of different ages and the time of the accident has huge potential impact in accidents. Their study helps in finding new factors involved in accidents. The authors in [21] proposed an ANN-based self-learning control framework which can improve the strength of vehicle during collision avoidance with the increase in the experience of driver. Their study added new factor called driver experience. They performed the experiments using CarSim software. The authors in [22] assess the impact of V2V communication for road safety applications. Instead of DSRC devices, the authors used laptops as a test bed with the necessary equipment. Their main interest was in broadcasting messages between V2V for collision avoidance without DSRC. Tests were performed using a Linux-Based Laptop and a Scapy add-on.

Frontal obstacle detection is a challenging task in collision avoidance. The authors in [23] proposed Mamdani and Sugeno fuzzy logic methods to overcome this challenge. They said that Mamdani and Sugeno can obtain the same efficiency. Experiments were performed using MATLAB. The authors in [24] have proposed an obstacle avoiding system based on a fuzzy logic controller. It allows the vehicle to move independently while avoiding collision with an obstacle. The controller was implemented in real time with an underwater vehicle. Based on tests, the authors proved that fuzzy logic can be useful for collision avoidance. The authors in [9] proposed a methodology for avoiding rear-end collisions. In this study, the focus was on visibility and road alignment factors.

The researchers in [25] describe different factors involved in accident occurrence. According to their study, the most influential factors are environmental and human factors. In addition, high-speed driving, cell phone use, and use of substances also increases the risk of accidents. The authors suggested some strategies to reduce the chance of accidents. Faisal et al. [26] proposed a novel approach for collision avoidance between autonomous vehicles following social norms and emotions. The authors used the fuzzy logic to compute the results of factors involved. A simulation was created for the proposed model using NetLogo. Xiang et al. [27] proposed a forward collision avoidance algorithm where fuzzy logic rules were used for initiating critical brake control.

The researchers in [28] have proposed a rear-end collision avoidance scheme between vehicles. They have considered factors such as the road, vehicle type, driver, and external environment. For implementation purposes, Fuzzy Logic, VISSIM, and MATLAB were used. The authors in [29] deal with two key aspects of road transport: efficiency and safety. The proposed system detects the obstacles and generates the warnings and then sends them to the driver. In case the driver fails to perform an action, the control shifts to cruise control system. The authors in [30] proposed software-based collision avoidance systems using Dedicated Short-Range Communication (DSRC). They performed the timing analysis of events based on the DSRC detection range, communication latency, and road condition. Zhao et al. [31] also proposed a collision warning system based on DSRC.

The authors in [32] proposed a collision warning algorithm in which they analyze different factors (human, road condition, time, and position), which can affect the performance of collision. The authors in [33] proposed a collision avoidance system where traffic lights communicate with nearby smartphones. Then these smartphones share warning with other smartphones. Though they performed collision avoidance, but no factors were used in their proposed system. The authors in [34] described a rear-end collision system using a model called the Bayesian Network. The model depends on ego factors of drivers and the braking intention of the front vehicle. They will implement and test their work in the future. The authors in [35] proposed a framework for space-based collision avoidance. V2V communication and a machine learning approach were used to accurately detect the collision and avoid its occurrence.

According to the authors in [36], features of human drivers have been used to control the rear-end collision using fuzzy logic, and according to the authors in [37], fuzzy logic can resolve the rear-end collision avoidance mathematical issues. In our previous work [8], we have proposed a V2V rear-end collision avoidance algorithm with the help of fuzzy rules bearing in mind the environmental, physical, and mental factors. These factors contribute to road collisions. A Multi-Factor-Based Road Accident Prevention System (MFBRAPs) was proposed to avoid collisions. MATLAB and Net Logo were used for the implementation. In this paper, we are fetching new important factors which can help the collision avoidance algorithms in more effective way based on existing research. We put these new factors in MFBRAPs for a better collision warning system in V2V.

3. Proposed Methodology

There are 6 levels (level 0 to level 5) of autonomy defined by the Society of Automotive and Engineers (SAE) [38]. In level 0, all tasks are performed by driver. Level 1 assists drivers with an advanced driver assistance system (ADAS) [39]. In Level 2, the driver is still present and responsible for driving and monitoring the environment with the assistance of more than one ADAS. Level 2 is also known as partial driving automation [40]. Levels 3, 4, and 5 are under the system software responsibility in which an autonomous system monitors the environment continuously [41]. The driver is still required in level 3 and level 4. Level 5 is called fully autonomous. Due to legislative factors and technological limitations [42–44], the human driver is still mandatory in AVs. The proposed architecture is applicable to the semi-autonomous vehicles [41], in which the driver or vehicle can shift control. Figure 2 describes the proposed methodology and shows how control will be

shifted between human and driver after calculating the chances of accidents. Figure 3 describes the factors involved at a fuzzy level.

The system will apply fuzzy membership function rules for the fuzzification of input values with every possible combination. Time, environmental, physical, weekday, and driver factors are the inputs to the Mamdani Fuzzy Inference System of MATLAB. The combination of these factors' values will find the chances of a collision occurring. If the chance of collision is high, then the vehicle will take control from the driver and apply the brakes automatically with the help of an agent-based system SIM Connector. When the situation is in normal position, the control will be handed back to the driver. If chances of accident are low, then the vehicle's control will remain with the human driver.

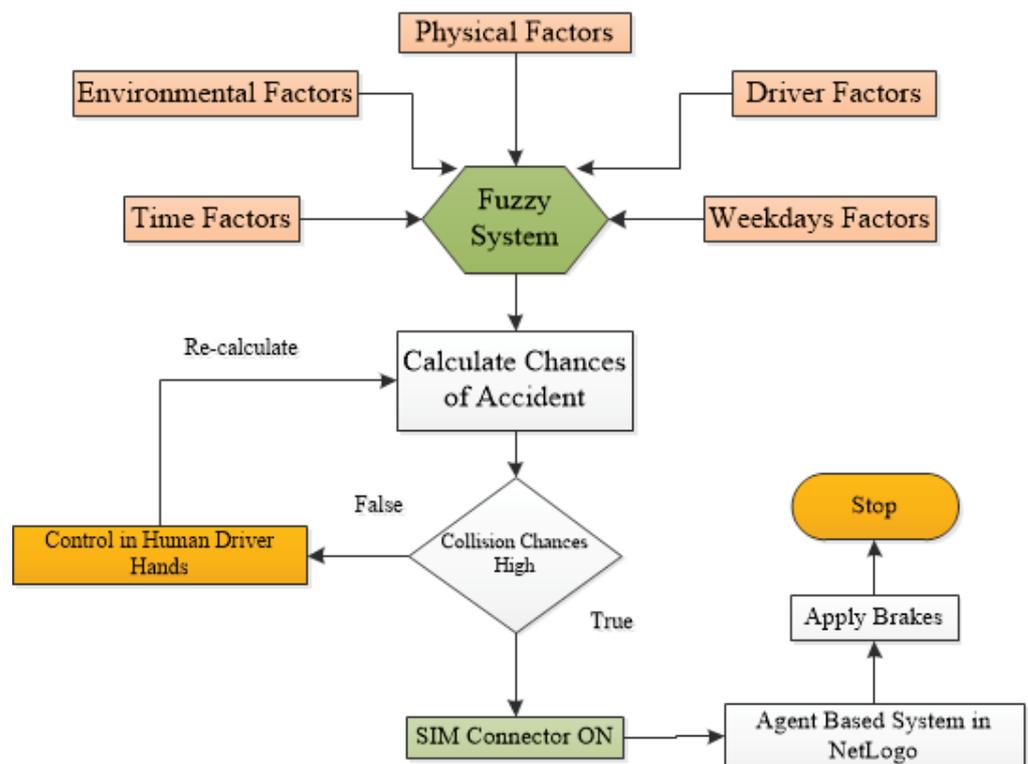


Figure 2. Architectural Diagram of Proposed Methodology.

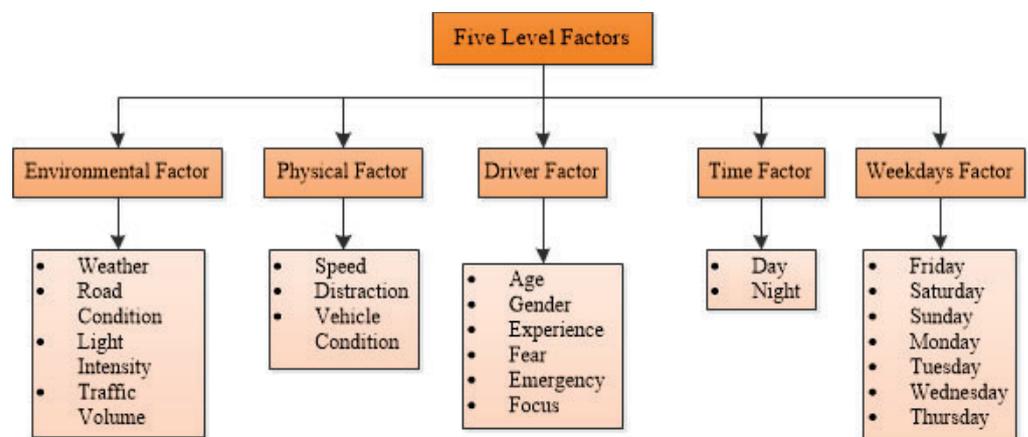


Figure 3. Five Level Factors of Collision in V2V.

3.1. Five Factors Description

We have classified accident reasons into five categories, i.e., environmental, physical, driver characteristics, time, and weekday factors. Each category has further sub-member

functions. These are described in past studies but not combined in one algorithm. For the selected factors, the fuzzy values range from low to high according to the authors in [8,37] and as shown in Figure 3. Five intensity levels from very low, meaning the value is zero, to very high, meaning the value is one. This nature of the variation in values from low to high proves the importance of the fuzzy system. Table 1 defines the numeric range division for factors described in Figure 3.

Table 1. Quantitative values of intensity levels for All Factors.

Very Low	Low	Average	High	Very High
0 to 0.1	0.11 to 0.25	0.251 to 0.5	0.51 to 0.75	0.751 to 1

3.1.1. Environmental Factors

Weather conditions [45], road conditions [46], light intensity [47], and traffic volume [48] are the contributing factors of accident. Each of these factors will be checked individually. Weather can be rainy, snowy, foggy, or a dust storm, in which the intensity of the chance of accident will vary. Good road conditions along with rain will produce different results as compared with bad road conditions with rain. Light intensity can be very good, average, or bad. Traffic volume can be high, average, or there can be few vehicles on the road. All these factors and sub-factors are the inputs to Environmental Factors. The result will show how much environmental factors will contribute to accidents.

3.1.2. Physical Factors

This factor includes the speed of the vehicle [49], distracting activities [50] of the driver, and the current vehicle condition [51]. Higher speeds of the vehicle has a higher risk. Distinctive activities can be the use of mobile phones during driving. The vehicle can be in very good condition, average condition, or in very bad condition. All these factors will contribute to accident risk.

3.1.3. Driver Factor

Focus on driving [12] due to alcohol or drugs, behavioral situation [52], fear [53], and behavior in an emergency [54] are the contributing factors that may cause accidents. These are all the inputs to the mental factors which affect driving. According to authors in [11,12], driving experience matters a lot in accident occurrence. The age and gender of the driver [12] also contribute to chances of accident. These are the inputs for the measurements of accident chances. The result will be the output of the driver factor.

3.1.4. Time Factor

The authors in [12] described the effects of day and nighttime on the chance of accident. From 18:00 to 20:00 h, there is a high accident rate. According to their study, accident rate during the daytime is high.

3.1.5. Weekdays Factor

According to authors in [12], weekdays, especially Friday and Saturday, have high chances of accidents because drivers behave differently on different days.

Our proposed system will check which category contributes the most to accident occurrence, and by combining the result of all the categories, the system will calculate the chance of an accident. Every category and its sub-functions will compute the fuzzy values from the vehicle's sensors and other pre-defined fuzzy values. These values will be processed in MATLAB, and if the chance of an accident is high, then the vehicle will apply an immediate brake and send messages to the relevant persons and organizations. The proposed simulation will show how brakes will be applied when chances of accidents are high. If chances of accidents are not high, then control remains with the human driver.

4. Proposed Algorithm

In semi-autonomous vehicles [55–57], the driving control can be shifted between vehicles and human. The proposed algorithm elaborates how the control of the vehicle will be shifted between the human driver and the vehicle's automatic driving system. Sensor values are the basis for this implementation. Variables store these values, and functions perform operation on these variables and generate the chance of accident value. According to this calculated value, the required function call takes place as shown in Algorithm 1.

Algorithm 1: Control Structure for transferring vehicle control to and from vehicle

```

Do{
  CD->GetSensorValues();
  Data = Fuzzification (CD);// calculate the current situation of each category
  Result = MFBRECCAS (Data)// combined/integrated result of every category and factors
  If (Result==chances of accidents)
  Control_brak() //Take control of the driver.
  If (EF is high)
  Control_speed_generate_alert();
  PY_F=result of physical factor obtained from Data variable
  If (PY_F is high)
  Control_Speed()
  DF=result of Day factor obtained from Data variable
  If (MF is high)
  Control_speed_send_message();
  Control_Speed();// control_speed() will be called within this function
  DF=result of driver factors obtained from Data variable
  If (DF is high)
  Control_brak()
  T_O_A=result of time factor obtained from Data variable
  If (T_O_A is high)
  {
  Apply Brakes ()
  Control Speed ()
  }
  Else
  Control_back_to_driver()
  End if
}
While (1);

```

The proposed algorithm variables and functions are described in Table 2 below. a brief description is provided of the variables and functions in the table. Table-based description helps in easily understanding the purpose of the variables and functions. The proposed algorithm presented in this section shows the collision avoidance mechanism in a simulation environment. The function which obtains sensor values in the proposed algorithm is actually related to different factor values. After calculating the integrated result of every factor, it is passed into a result variable. The selected vehicle in the proposed system takes the necessary action on the basis of the result variable's value. In Figure 10, in the upper right corner, we set the different factors' variables whose values passed into the obtain sensor values function used in the proposed algorithm. Figure 11 shows the selected vehicle with blue color.

Table 2. Proposed algorithm Variables & functions.

Variables Used	Meanings of Variables	Description of Variables	Function Name	Function Description
CD	Category Data	It will take input from sensors about every accident-causing factor	Get SensorValues()	Vehicle has sensors from which it can take information about different factors involved.
Data	Sensor data	Calculate the current situation of each category	Fuzzification();	Mamdani membership function will take sensor values as input and apply fuzzy logic.
Result	Chances of Accident (COA)	It contains the value of the chance of accident. If the value is high, then the required function will be called.	Controlbrak()	If fuzzy logic functions show higher chance of accident, then our proposed system will apply brakes.
EF	Environmental Factor	Result of environmental factor obtained from data variable	ControlSpeed() & generate alert	It will accelerate or decelerate the vehicle's speed (reduce/fast).
TOA	Time Of Accident	Result of time factor obtained from Data variable	ControlSpeed() & generate alert	It will accelerate or decelerate the vehicle's speed (reduce/fast).
PYF	Physical Factor	Result of physical factor obtained from Data variable	Control speed send message()	It will accelerate or decelerate the vehicle's speed (reduce/fast).
WD	Weekday Factor	Result of weekday factor obtained from Data variable List	ControlSpeed() & generate alert	It will accelerate or decelerate the vehicle's speed (reduce/fast).
DF	Driver Factor	Result of driver factors obtained from Data variable	ControlSpeed() & generate alert	It will accelerate or decelerate the vehicle's speed (reduce/fast).

5. Experiments

We divided this section in two parts: First, based on multi-factors, finding the chances of an accident using the Fuzzy Logic Tool Box and, second, Net-Logo-based Simulation experiments to show the effects of multi-factor-enabled vehicle on accidents results.

The Fuzzy Logic Toolbox provides MATLAB functions and a Simulink block for analyzing, designing, and simulating systems based on fuzzy logic, as described in Figure 4. The authors in [58–62] have also used the Fuzzy Logic in the modeling of their proposed work.

We have used the Fuzzy Inference System (FIS) as presented in [63] to apply membership functions on pre-defined input values ranging from 0 to 1. A chance of accident value approaching 0 is considered a low chance of an accident and 1 as a high chance.

Our proposed system is based on five categories of factors which are the causes of accidents and fatalities. We performed the experiments in MATLAB using FIS, and Figures 5–8 are the samples of experiments with input–output relationships.

In Figure 5, multi-factors are given as inputs, and the chances of accidents are calculated. Every factor is evaluated using membership functions of the Mamdani Fuzzy inference system. Whenever the integrated factors values are greater than 0.75, then the system will take control from the driver to avoid collision between vehicles. We are describing the switching of control in the second part of experiments in Net-Logo simulations based on the values of different factors values.

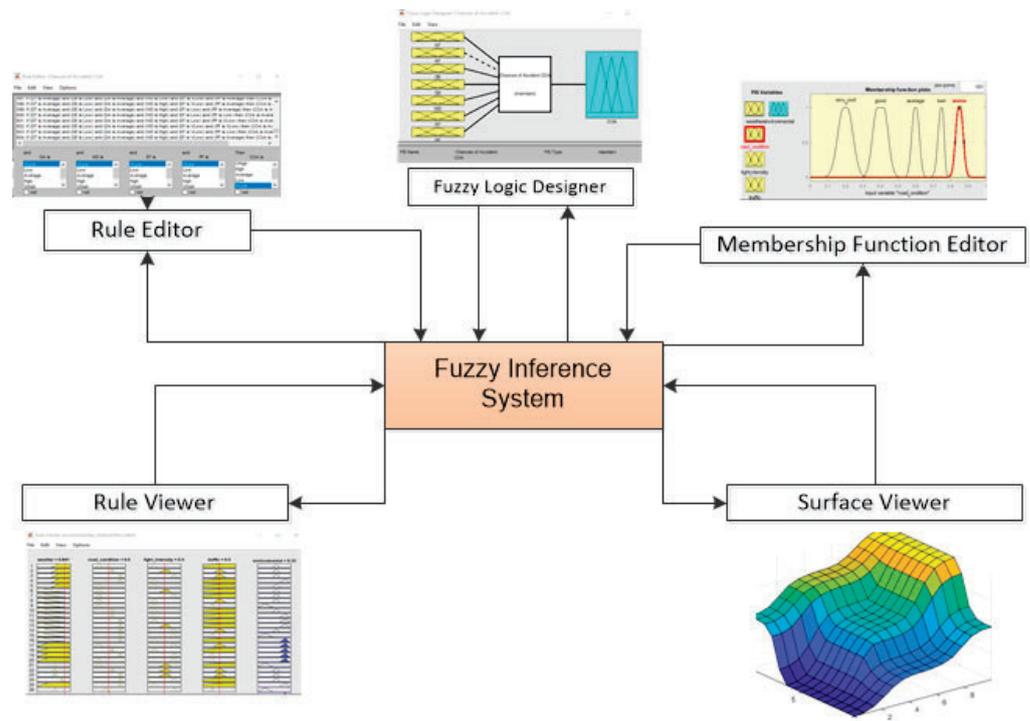


Figure 4. Fuzzy Inference System to apply membership functions.

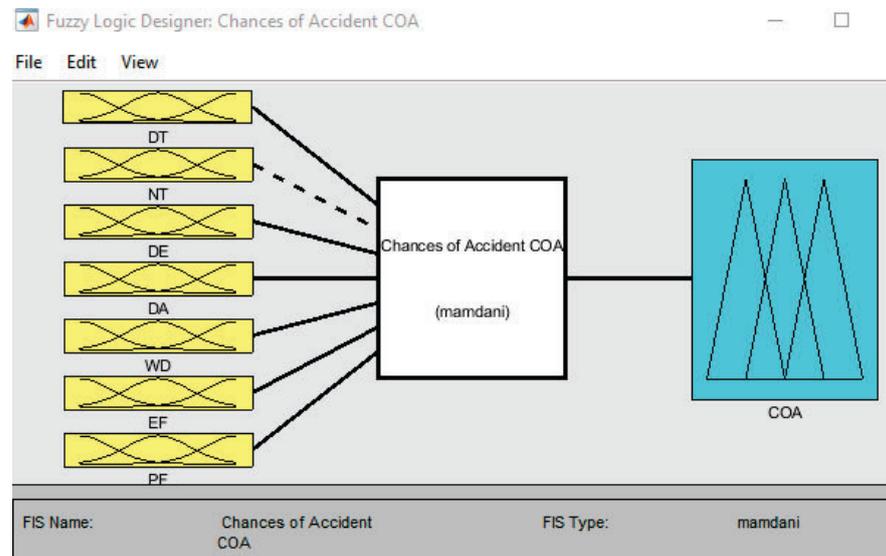


Figure 5. Input & Output Relationship of Integrated Factors.

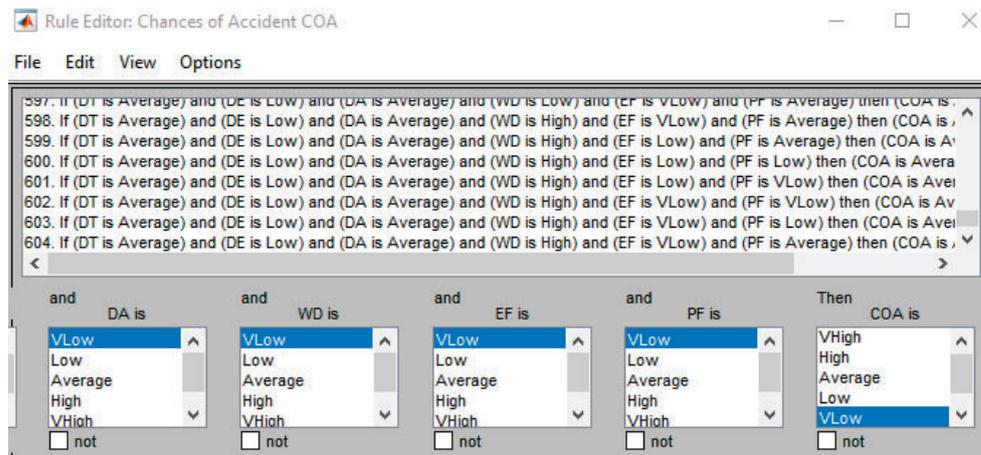


Figure 6. Rules Used in Fuzzy Logic.

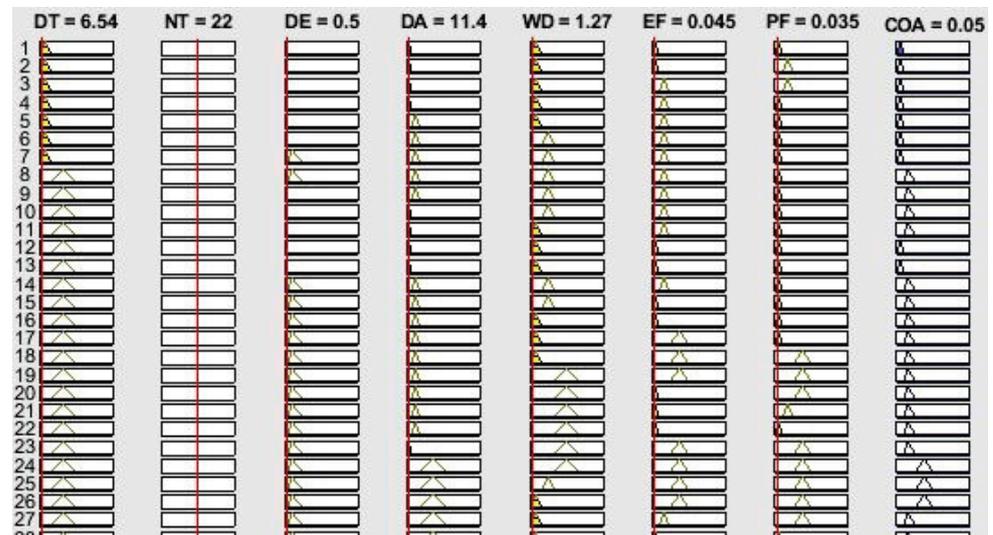


Figure 7. Testing the COA on Rule Viewer with Input as Day Time (DT), Night Time (NT), Driver’s Experience (DE), Driver’s Age (DA), Weekday (WD), Environmental Factors (EF), Physical Factors (PF).

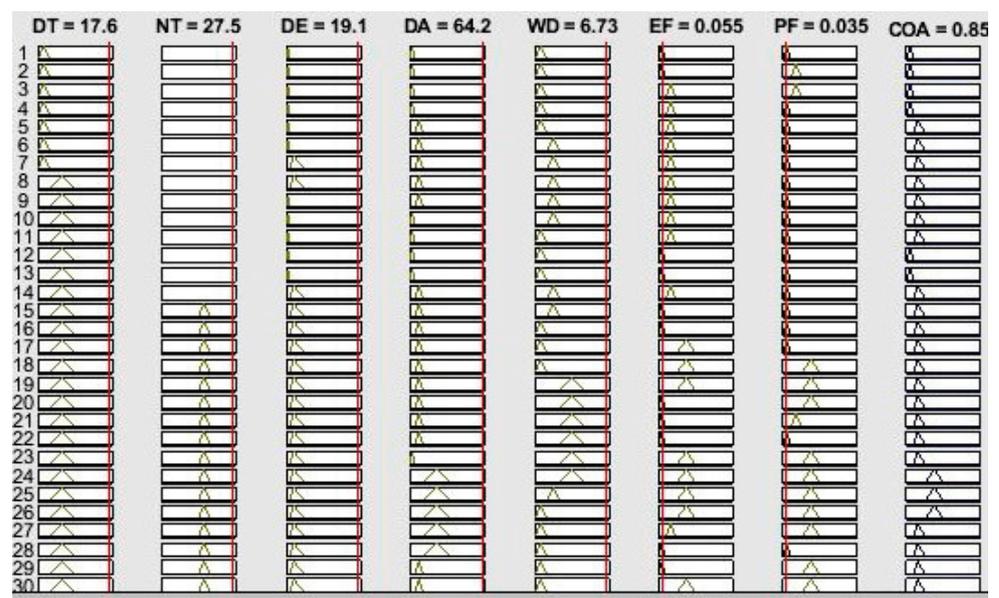


Figure 8. Testing the COA on Rule Viewer.

5.1. Fuzzy-Logic Based Experiments

Next, the Figures in this section show how inputs as rules are fed into the model and the output results between 0 and 1. Each input/factor consists of five member functions from very low to very high, as described in Table 1. The rules for computing the COA are defined in the FIS Rule editor as shown in Figure 6.

The rules viewer in Figure 7 shows how inputs are contributing to finding the COA. The nighttime input factor is selected as none in Figure 6 and has no effect on the COA. The authors in [12] described the importance of the time factor in collisions. Consequently, when we select the day time input as none, then this factor input will not effect the COA. We performed the experiments by scaling input variables, and the results are verified as per the rules defined in the FIS rule editor. Figures 7 and 8 show the experiment results performed in FIS.

The experiment in Figure 9 shows a very low chance of accident (COA = 0.05), which means control is in driver’s hand and there is no need of control switching. However, the experiment in Figure 9 shows a very high chance of accident (COA = 0.854), and there is need of control switching from the driver to the autonomous mode. This variation in the COA from very low to very high validates the proposed fuzzy logic for multi-factor inputs into the proposed system. Further experiments and their results are provided in the result section in Table 3.

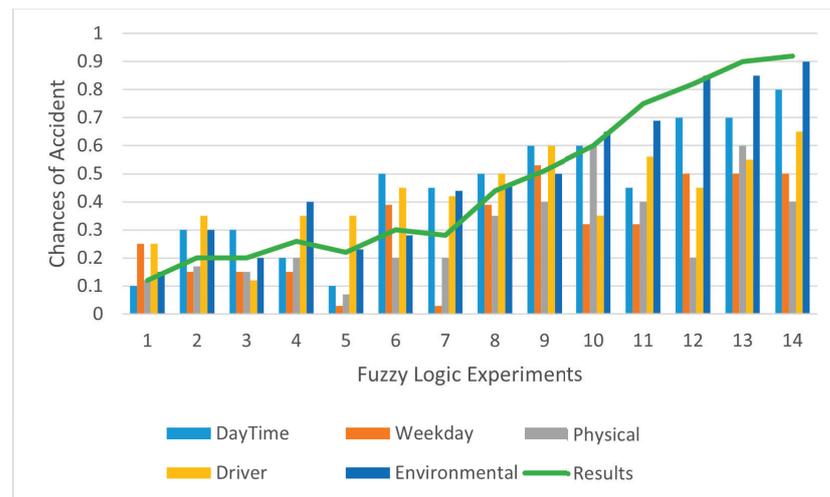


Figure 9. Sample Experiments Results from VLow to VHigh for COA.

Table 3. Different Experiments Results Computed Using Rule Viewer.

Sr.	DT	WD	PF	DE	DA	EF	COA
1	0.1(VL)	0.25(L)	0.12(L)	0.25(L)	0.15(L)	0.15(L)	0.12(L)
2	0.3(L)	0.15(L)	0.17(L)	0.35(AVG)	0.2(L)	0.3(AVG)	0.20(L)
3	0.3(L)	0.15(L)	0.15(L)	0.12(L)	0.3(AVG)	0.2(L)	0.20(L)
4	0.2(L)	0.15(L)	0.2(L)	0.35(AVG)	0.25(AVG)	0.40(AVG)	0.26(AVG)
5	0.1(VL)	0.03(VL)	0.07(VL)	0.35(L)	0.3(AVG)	0.23(L)	0.08(VL)
6	0.5(AVG)	0.39(AVG)	0.2(L)	0.45(AVG)	0.45(AVG)	0.28(AVG)	0.3(AVG)
7	0.45(AVG)	0.03(VL)	0.2(L)	0.42(AVG)	0.35(AVG)	0.44(AVG)	0.28(AVG)
8	0.5(AVG)	0.39(AVG)	0.35(AVG)	0.5(AVG)	0.5(AVG)	0.46(AVG)	0.44(AVG)
9	0.6(H)	0.53(H)	0.4(AVG)	0.6(H)	0.5(AVG)	0.5(AVG)	0.51(H)
10	0.6(H)	0.32(L)	0.6(H)	0.35(AVG)	0.7(H)	0.65(H)	0.6(H)

Table 3. Cont.

Sr.	DT	WD	PF	DE	DA	EF	COA
11	0.45(AVG)	0.32(L)	0.4(AVG)	0.55(H)	0.7(H)	0.69(H)	0.75(H)
12	0.7(H)	0.5(AVG)	0.2(L)	0.45(AVG)	0.9(VH)	0.85(H)	0.82(VH)
13	0.7(H)	0.5(AVG)	0.6(H)	0.55(H)	0.9(VH)	0.85(VH)	0.9(VH)
14	0.8(H)	0.5(AVG)	0.4(AVG)	0.65(H)	0.95(VH)	0.9(H)	0.97(VH)

5.2. Simulation-Based Experiments

The approach used to model the complex systems in engineering and technologies, etc., is known as agent-based modeling [64–66], and for this purpose, we used a Net-Logo simulation model. Net-Logo provides an observer which can monitor and validate the simulation scenario. The results achieved using the simulation validate the proposed algorithm. Figure 10 shows the simulation interface for the experiments’ setup. The selected blue car is enabled with multi-factors to avoid the collision, and the experiments’ results validate that the selected car collisions are much less as compared with other cars in the simulation environment.

This can be seen in the plot generated by the Net-Logo simulator which shows the number of collisions. An alert message box shows a beep when the vehicle is in danger or is in normal condition according to the reading of the factors involved to take necessary actions. The control box in the simulation shows whether the control of the vehicle is in the driver’s hand or in the autonomous mode. As the simulation runs, the values of the plot show the collision, alert box status, and the vehicle control changes according to multi-factor values, as seen in Figures 10 and 11. In Figure 10, the control is in the driver’s hands, but as the blue car finds very high chances of accident, the control in Figure 11 is then shifted to the autonomous mode. In the simulation, we performed eight experiments with different values of multi-factors ranging from very low to very high. The experiments and their results are presented in Table 4. To understand the effect of multi-factor-enabled vehicles, we discuss experiments 1 and 6 here. In experiment 1, when the daytime factor is very high and the weekday factor is also very high, the vehicle without factors’ collision count is 65 and the vehicle with factors’ count is 4. In experiment 6, when the environmental factor is high, the physical factor is very high, and daytime is also very high, then the vehicle without factors’ collision count is 314, which is very high, and the vehicle with factors enabled’s collision count is 23. There is a significant result difference between both vehicles due to the proposed system implementation in the simulation environment. The results of the experiments are given in Figure 12 in the Results section. Figures 10 and 11 show the experimental setup with results plot as number of collisions count.

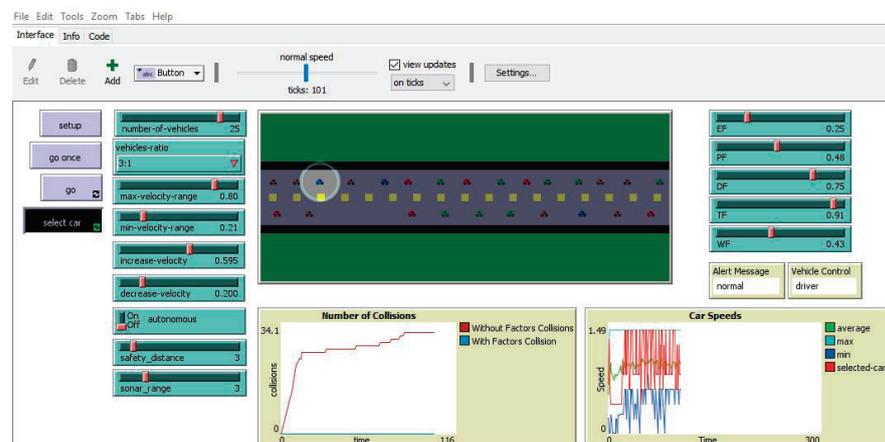


Figure 10. Main Interface of Simulation with Control in Driver Hand.

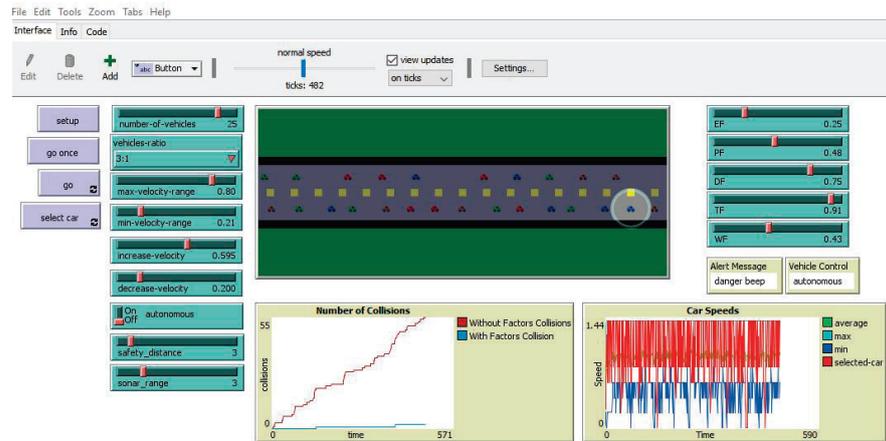


Figure 11. Main Simulation Interface and Control Shifted to AV.

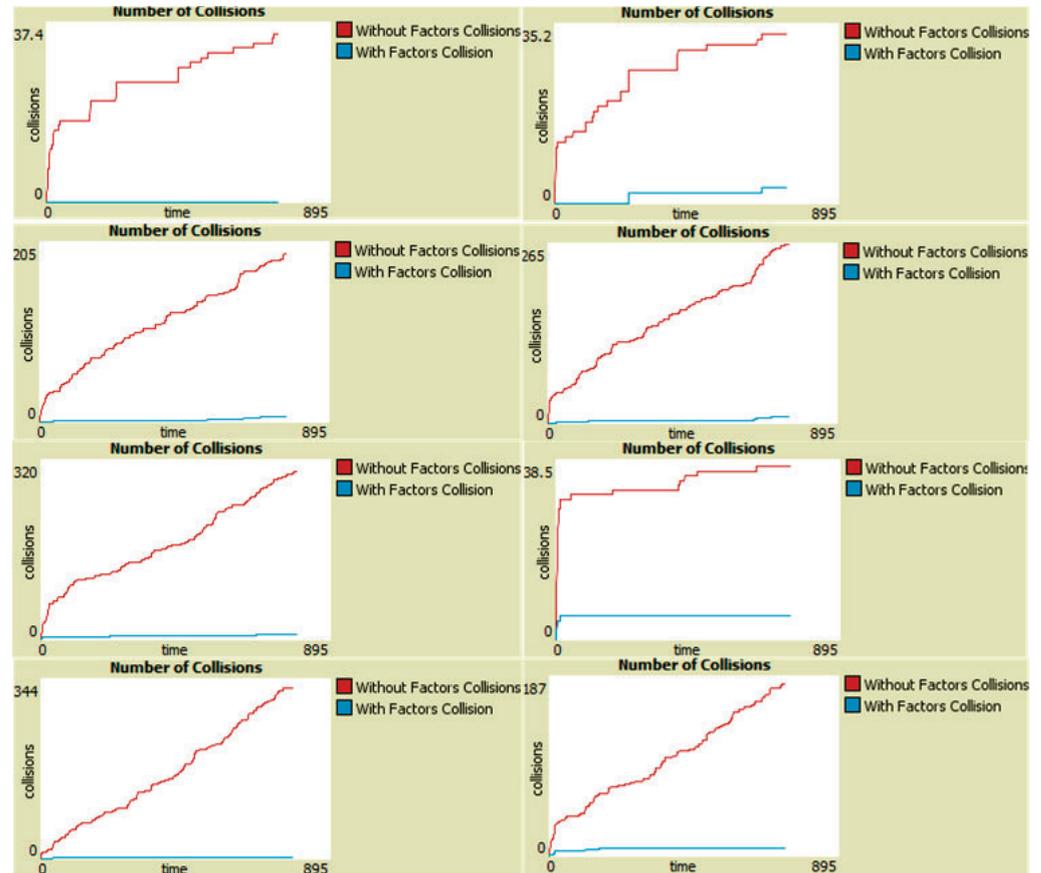


Figure 12. Simulation Results with and without Multi Factors.

Table 4. Experiments Results Using Net-Logo Simulation with Inputs (EF = Environmental Factor, PF = Physical Factor, DF = Daytime Factor, TF = Time Factor, WF = Weekday Factor).

Exp #	EF	PF	DF	TF	WF	Collisions without Factors	Collisions with Factors
1	0.3	0.4	0.8	0.45	0.91	65	4
2	0.46	0.8	0.7	0.55	0.48	17	0
3	0.3	0.4	0.4	0.51	0.45	26	2
4	0.7	0.8	0.7	0.85	0.48	46	3

Table 4. *Cont.*

Exp #	EF	PF	DF	TF	WF	Collisions without Factors	Collisions with Factors
5	0.7	0.8	0.7	0.25	0.44	47	7
6	0.7	0.8	0.8	0.15	0.54	314	23
7	0.3	0.4	0.8	0.95	0.90	404	15
8	0.7	0.6	0.8	0.55	0.45	28	0

6. Results

We divided the results section in two parts. In the first part, we explain the chances of accident calculation using Table 3 and combo chart. In the second part, we explained the results generated after simulation-based experiments in Net-Logo.

The fuzzy experimental results in Figure 7 show very low chances of accidents as compared to experiment in Figure 8, which shows high chances. Table 3 consists of the input factors' contribution to chances of accident, and Figure 9 shows the combo chart, which shows how the COA varies according to input values ranging from 0 to 1. The experiments 5 and 12 with very low and very high COA results in the Table 3 generated from fuzzy logic. The results ranging between very low and very high are validated by the proposed fuzzy logic. For graphical representation of the results in Table 3, we used the combo chart in this section.

Figure 12 shows number of collision with factors and without factors. Red line in the plot shows vehicles without multi factors and blue line in plot shows vehicle with multi factors enabled. The difference between number of collisions can be seen clearly in plots. We also listed the experiments in Table 4 with the results of collisions. The experiments show very good results in collision avoidance when the proposed algorithm is used in the simulation environment setup. The speed plot in the simulation setup shows the speed of both vehicles, and this plot shows clearly that the blue car with multi-factors enabled speed is under control due to speed control function as we used in our proposed algorithm and described in Table 2. The results of the simulation in Table 4 shows a significant difference in collisions between vehicles with and without multi-factors enabled.

7. Discussion

The primary focus of this research work is to highlight the importance of different factor combinations involved in the vehicle's collision and to avoid this collision with the help of driver assistance software. There are many factors that exist according to the existing research, but it is not possible to cover all of them here. We used some of them and achieve a satisfactory result. We used the Net-Logo tool for simulation and designed a multi-factor-based simulation environment. In the simulation section, when different factor values change from high to very high, the system shows how the control is shifted from the human driver to the vehicle's autonomous mode. In addition, the alert box shows danger beeps for the driver's assistance. Our proposed system computes the quantitative values and calculates the chances of accident. The driver assistance system then takes the necessary action and avoids the collision. After calculating the chances of accident, the driver assistance system with the proposed algorithm activates the different functions to control the speed and apply the brake. These functions are described in Table 2 with function name and function meaning. When required, the driver assistance system generates the alert. Basically, these alerts are warnings for the driver to take necessary action and control the vehicle. If the driver ignores the alert, then the control shift function takes place. When an alert is generated by the system, which is the danger beep, it means that the system is now ready to apply the brake and reduce the speed to avoid the collision if the driver does not take action to handle the emergency. Control shifts from the human to driver assistance systems after alert messages from "normal to danger beep". In Figures 10 and 11,

we can see the alert message status changes from normal to danger beep. This switching helps the proposed system to apply the brake and reduce the speed in time because it prepares the system to handle the situation. The status of the alert message changing from normal to danger beep means that the proposed logic is working, monitoring the environment, and calculating the chances of collision and is now ready to take action if the driver does not take the necessary action.

Experiments are performed with considering different values of factors involved in vehicle collision. We can change the values of factors from very low to very high in the simulation model. According to proposed model, the control is shifted from the human to the autonomous mode of the vehicle when chances of accident are high to avoid collision. The algorithm used in the proposed model implementation shows the shifting of control between human and autonomous mode. In the case of very high results computed by the system, an alert is also generated to inform the driver about the current situation. The results show the importance and correctness of the proposed model for collision avoidance. The generated graphs from simulation-based results show that a vehicle with multi-factors enabled shows significant improvement in collision avoidance and also prove the worthiness of the proposed model for a driver assistance system.

Limitations

There are some limitations in this study which can be addressed in future research. One limitation of the proposed system is that we performed the experiments in a simulation-based environment and not in a real-time environment. The second limitation is if the driver is not responding to alerts and the software is also not in working condition. Then, the proposed system will not be effective at avoiding the collision. The third limitation is related to hardware failure. If any of the hardware fails in the vehicle, it can affect the performance of the proposed system. In addition, when two drivers are driving very closely and the leading driver applies their brakes suddenly, then this scenario is also very difficult to handle to avoid collision.

8. Conclusions

Our contribution in this work is that we have taken a step towards the betterment of humans using modern techniques. If we timely calculate these factors' risks, we may be able to save one's life. Existing studies proved that different factors individually and combined can play their role in the collision between vehicles. However, no one countered them using simulation-based results to avoid collisions. We first demonstrated and calculated the chances of accident using fuzzy logic and showed the multi-factors' importance. The proposed system first calculates the chances of accident and then avoids the collision by shifting the control from the human driver to vehicle's automated system while generating an alert from the human driver. The simulation results designed in the Net-Logo tool demonstrated that the vehicle which is enabled with multi-factors can avoid the collision as compared to the other vehicles without multi-factors. The automakers in the near future can use this research for the improvement of collision avoidance because, in existing studies, many authors are working on finding factors which are the causes of accidents/collisions.

8.1. Future Work

The current work is performed on a semi-autonomous vehicle. In the future, fully autonomous vehicles can be accommodated with the fuzzification of accident-causing factors. Message passing in times of emergency can also be processed according to privacy protection rules. Emotional factors may also improve the results to avoid collision. In addition, this model can be enhanced to work on the T-junctions. Another important research direction is that time to collision avoidance can be incorporated in the future to enhance the proposed model.

8.2. Recommendations

Collision avoidance systems are the key for connected autonomous vehicles and are very helpful in reducing road traffic injuries and fatalities. In this regard, the proposed system is providing a solid foundation to handle different factors involved in the collision of vehicles. This system can be extended to many other issues which are the causes of collision. These are planning and deciding factors, e.g., illegal maneuvers, following too closely, stopping suddenly, or accelerating very rapidly from stop. Factors which are unavoidable by the driver include brakes failing, suspension failing, steering failing, wheels failing, and transmission failing. Incapacitation issues include heart attack or physical impairment of the ability to act. These highlighted factors in the recommendations can be overcome with the proposed system, or new techniques can be applied in the future.

Author Contributions: Conceptualization, M.A.S., S.R. and A.R.D.; methodology, S.R.; software, H.T.R.; validation, S.R., A.R.D. and H.A.K.; formal analysis, H.A.K. and E.A.; investigation, A.M.E.-S.; resources, S.M.L.; data curation, H.A.K.; writing—original draft preparation, S.R., H.A.K. and A.R.D.; writing—review and editing, A.R.D. and H.A.K.; visualization, K.A.; supervision, M.A.S. and H.A.K.; project administration, H.A.K.; funding acquisition, K.A., A.M.E.-S. and S.M.L. All authors have read and agreed to the published version of the manuscript.

Funding: Researchers Supporting Project number (RSP-2021/133), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to King Saud University for funding this work through Researchers supporting project number (RSP-2021/133), King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

- World Health Organization. *Global Status Report on Road Safety 2018: Summary*; Technical Report; World Health Organization: Geneva, Switzerland, 2018.
- Meng, Q.; Qu, X. Estimation of rear-end vehicle crash frequencies in urban road tunnels. *Accid. Anal. Prev.* **2012**, *48*, 254–263. [\[CrossRef\]](#)
- Chen, C.; Zhang, G.; Tarefder, R.; Ma, J.; Wei, H.; Guan, H. A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. *Accid. Anal. Prev.* **2015**, *80*, 76–88. [\[CrossRef\]](#)
- Moon, S.; Moon, I.; Yi, K. Design, tuning, and evaluation of a full-range adaptive cruise control system with collision avoidance. *Control Eng. Pract.* **2009**, *17*, 442–455. [\[CrossRef\]](#)
- Van Den Berg, J.; Wilkie, D.; Guy, S.J.; Niethammer, M.; Manocha, D. LQG-obstacles: Feedback control with collision avoidance for mobile robots with motion and sensing uncertainty. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 346–353.
- Chen, C.; Li, M.; Sui, J.; Wei, K.; Pei, Q. A genetic algorithm-optimized fuzzy logic controller to avoid rear-end collisions. *J. Adv. Transp.* **2016**, *50*, 1735–1753. [\[CrossRef\]](#)
- Tron, E.; Margalot, M. Mathematical modeling of observed natural behavior: A fuzzy logic approach. *Fuzzy Sets Syst.* **2004**, *146*, 437–450. [\[CrossRef\]](#)
- Razzaq, S.; Riaz, F.; Mehmood, T.; Ratyal, N.I. Multi-factors based road accident prevention system. In Proceedings of the 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan, 11–12 April 2016; pp. 190–195.
- Malin, F.; Norros, I.; Innamaa, S. Accident risk of road and weather conditions on different road types. *Accid. Anal. Prev.* **2019**, *122*, 181–188. [\[CrossRef\]](#)
- Shangguan, Q.; Fu, T.; Liu, S. Investigating rear-end collision avoidance behavior under varied foggy weather conditions: A study using advanced driving simulator and survival analysis. *Accid. Anal. Prev.* **2020**, *139*, 105499. [\[CrossRef\]](#)
- Zhang, Y.; Du, F.; Wang, J.; Ke, L.; Wang, M.; Hu, Y.; Yu, M.; Li, G.; Zhan, A. A Safety Collision Avoidance Algorithm Based on Comprehensive Characteristics. *Complexity* **2020**, *2020*, 1616420. [\[CrossRef\]](#)
- Ashraf, I.; Hur, S.; Shafiq, M.; Park, Y. Catastrophic factors involved in road accidents: Underlying causes and descriptive analysis. *PLoS ONE* **2019**, *14*, e0223473. [\[CrossRef\]](#)

13. Czubenko, M.; Kowalczyk, Z.; Ordys, A. Autonomous driver based on an intelligent system of decision-making. *Cogn. Comput.* **2015**, *7*, 569–581. [[CrossRef](#)]
14. Sharma, A.; Zheng, Z.; Bhaskar, A.; Haque, M.M. Modelling car-following behaviour of connected vehicles with a focus on driver compliance. *Transp. Res. Part B Methodol.* **2019**, *126*, 256–279. [[CrossRef](#)]
15. Córdoba, A.; Astrain, J.J.; Villadangos, J.; Azpilicueta, L.; López-Iturri, P.; Aguirre, E.; Falcone, F. Sestocross: Semantic expert system to manage single-lane road crossing. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 1221–1233. [[CrossRef](#)]
16. Abdulshaheed, H.R.; Yaseen, Z.T.; Salman, A.M.; Al_Barazanchi, I. A survey on the use of wimax and wi-fi on vehicular ad-hoc networks (vanets). In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2020; p. 012122.
17. Chen, K.P.; Hsiung, P.A. Vehicle collision prediction under reduced visibility conditions. *Sensors* **2018**, *18*, 3026. [[CrossRef](#)] [[PubMed](#)]
18. Yang, W.; Wan, B.; Qu, X. A forward collision warning system using driving intention recognition of the front vehicle and V2V communication. *IEEE Access* **2020**, *8*, 11268–11278. [[CrossRef](#)]
19. Yan, X.; Xue, Q.; Ma, L.; Xu, Y. Driving-simulator-based test on the effectiveness of auditory red-light running vehicle warning system based on time-to-collision sensor. *Sensors* **2014**, *14*, 3631–3651. [[CrossRef](#)]
20. Regev, S.; Rolison, J.J.; Moutari, S. Crash risk by driver age, gender, and time of day using a new exposure methodology. *J. Saf. Res.* **2018**, *66*, 131–140. [[CrossRef](#)]
21. Wang, Y.; Yin, G.; Li, Y.; Ullah, S.; Zhuang, W.; Wang, J.; Zhang, N.; Geng, K. Self-learning control for coordinated collision avoidance of automated vehicles. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2021**, *235*, 1149–1163. [[CrossRef](#)]
22. Flanagan, S.K.; He, J.; Peng, X.H. Improving emergency collision avoidance with vehicle to vehicle communications. In Proceedings of the 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; pp. 1322–1329.
23. Fauadi, H.F.; Nordin, M.H.; Zainon, Z.M. Frontal obstacle avoidance of an autonomous subsurface vehicle (ASV) using fuzzy logic method. In Proceedings of the 2007 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, 25–28 November 2007; pp. 125–128.
24. David, K.K.A.; Vicerra, R.R.; Bandala, A.A.; Lim, L.A.G.; Dadios, E.P. Unmanned underwater vehicle navigation and collision avoidance using fuzzy logic. In Proceedings of the 2013 IEEE/SICE International Symposium on System Integration, Kobe, Japan, 15–17 December 2013; pp. 126–131.
25. Gicquel, L.; Ordonneau, P.; Blot, E.; Toillon, C.; Ingrand, P.; Romo, L. Description of various factors contributing to traffic accidents in youth and measures proposed to alleviate recurrence. *Front. Psychiatry* **2017**, *8*, 94. [[CrossRef](#)] [[PubMed](#)]
26. Riaz, F.; Jabbar, S.; Sajid, M.; Ahmad, M.; Naseer, K.; Ali, N. A collision avoidance scheme for autonomous vehicles inspired by human social norms. *Comput. Electr. Eng.* **2018**, *69*, 690–704. [[CrossRef](#)]
27. Xiong, X.; Wang, M.; Cai, Y.; Chen, L.; Farah, H.; Hagenzieker, M. A forward collision avoidance algorithm based on driver braking behavior. *Accid. Anal. Prev.* **2019**, *129*, 30–43. [[CrossRef](#)] [[PubMed](#)]
28. Chen, C.; Xiang, H.; Qiu, T.; Wang, C.; Zhou, Y.; Chang, V. A rear-end collision prediction scheme based on deep learning in the Internet of Vehicles. *J. Parallel Distrib. Comput.* **2018**, *117*, 192–204. [[CrossRef](#)]
29. Jiménez, F.; Naranjo, J.E.; Anaya, J.J.; García, F.; Ponz, A.; Armingol, J.M. Advanced driver assistance system for road environments to improve safety and efficiency. *Transp. Res. Procedia* **2016**, *14*, 2245–2254. [[CrossRef](#)]
30. Tang, A.; Yip, A. Collision avoidance timing analysis of DSRC-based vehicles. *Accid. Anal. Prev.* **2010**, *42*, 182–195. [[CrossRef](#)]
31. Zhao, X.; Jing, S.; Hui, F.; Liu, R.; Khattak, A.J. DSRC-based rear-end collision warning system—An error-component safety distance model and field test. *Transp. Res. Part C Emerg. Technol.* **2019**, *107*, 92–104. [[CrossRef](#)]
32. Fu, Y.; Li, C.; Luan, T.H.; Zhang, Y.; Mao, G. Infrastructure-cooperative algorithm for effective intersection collision avoidance. *Transp. Res. Part C Emerg. Technol.* **2018**, *89*, 188–204. [[CrossRef](#)]
33. Santos-González, I.; Caballero-Gil, P.; Rivero-García, A.; Caballero-Gil, C. Priority and collision avoidance system for traffic lights. *Ad Hoc Netw.* **2019**, *94*, 101931. [[CrossRef](#)]
34. Russo, J.N.; Sproesser, T.; Drouhin, F.; Basset, M. Risk level assessment for rear-end collision with Bayesian network. *IFAC-PapersOnLine* **2017**, *50*, 12514–12519. [[CrossRef](#)]
35. Yu, J.; Petnga, L. Space-based collision avoidance framework for autonomous vehicles. *Procedia Comput. Sci.* **2018**, *140*, 37–45. [[CrossRef](#)]
36. Riaz, F.; Niazi, M.A. Towards social autonomous vehicles: Efficient collision avoidance scheme using Richardson’s arms race model. *PLoS ONE* **2017**, *12*, e0186103. [[CrossRef](#)]
37. Riaz, F.; Niazi, M.A. Enhanced emotion enabled cognitive agent-based rear-end collision avoidance controller for autonomous vehicles. *Simulation* **2018**, *94*, 957–977. [[CrossRef](#)]
38. SAE. Definitions for terms related to driving automation systems for on-road motor vehicles. *SAE Stand. J.* **2016**, *3016*, 2016.
39. Ziebinski, A.; Cupek, R.; Grzechca, D.; Chruszczyk, L. Review of advanced driver assistance systems (ADAS). In *AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2017; Volume 1906, p. 120002.
40. Llorca, D.F. From driving automation systems to autonomous vehicles: Clarifying the terminology. *arXiv* **2021**, arXiv:2103.10844.
41. van Wyk, F.; Khojandi, A.; Masoud, N. Optimal switching policy between driving entities in semi-autonomous vehicles. *Transp. Res. Part C Emerg. Technol.* **2020**, *114*, 517–531. [[CrossRef](#)]

42. Goodrich, M.A.; Boer, E.R. Designing human-centered automation: Trade-offs in collision avoidance system design. *IEEE Trans. Intell. Transp. Syst.* **2000**, *1*, 40–54. [[CrossRef](#)]
43. Kyriakidis, M.; Happee, R.; de Winter, J.C. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *32*, 127–140. [[CrossRef](#)]
44. Parasuraman, R.; Manzey, D.H. Complacency and bias in human use of automation: An attentional integration. *Hum. Factors* **2010**, *52*, 381–410. [[CrossRef](#)] [[PubMed](#)]
45. Strong, C.K.; Ye, Z.; Shi, X. Safety effects of winter weather: The state of knowledge and remaining challenges. *Transp. Rev.* **2010**, *30*, 677–699. [[CrossRef](#)]
46. Usman, T.; Fu, L.; Miranda-Moreno, L.F. Quantifying safety benefit of winter road maintenance: Accident frequency modeling. *Accid. Anal. Prev.* **2010**, *42*, 1878–1887. [[CrossRef](#)] [[PubMed](#)]
47. Wanvik, P.O. Effects of road lighting: An analysis based on Dutch accident statistics 1987–2006. *Accid. Anal. Prev.* **2009**, *41*, 123–128. [[CrossRef](#)]
48. Wang, C.; Quddus, M.; Ison, S. A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK. *Transp. A Transp. Sci.* **2013**, *9*, 124–148. [[CrossRef](#)]
49. Clarke, D.D.; Ward, P.; Bartle, C.; Truman, W. Killer crashes: Fatal road traffic accidents in the UK. *Accid. Anal. Prev.* **2010**, *42*, 764–770. [[CrossRef](#)] [[PubMed](#)]
50. Susilo, Y.O.; Joewono, T.B.; Vandebona, U. Reasons underlying behaviour of motorcyclists disregarding traffic regulations in urban areas of Indonesia. *Accid. Anal. Prev.* **2015**, *75*, 272–284. [[CrossRef](#)] [[PubMed](#)]
51. Zhang, M.; Kecojevic, V.; Komljenovic, D. Investigation of haul truck-related fatal accidents in surface mining using fault tree analysis. *Saf. Sci.* **2014**, *65*, 106–117. [[CrossRef](#)]
52. Parker, D.; West, R.; Stradling, S.; Manstead, A.S. Behavioural characteristics and involvement in different types of traffic accident. *Accid. Anal. Prev.* **1995**, *27*, 571–581. [[CrossRef](#)]
53. Walshe, D.G.; Lewis, E.J.; Kim, S.I.; O’Sullivan, K.; Wiederhold, B.K. Exploring the use of computer games and virtual reality in exposure therapy for fear of driving following a motor vehicle accident. *CyberPsychol. Behav.* **2003**, *6*, 329–334. [[CrossRef](#)]
54. Gunnarsson, B.M.; Stomberg, M.W. Factors influencing decision making among ambulance nurses in emergency care situations. *Int. Emerg. Nurs.* **2009**, *17*, 83–89. [[CrossRef](#)] [[PubMed](#)]
55. Brown, J.W.; Revell, K.M.; Richardson, J.; Clark, J.R.; Caber, N.; Amanatidis, T.; Langdon, P.; Bradley, M.; Thompson, S.; Skrypchuk, L.; et al. Customisation of Takeover Guidance in Semi-Autonomous Vehicles. In *Designing Interaction and Interfaces for Automated Vehicles*; CRC Press: Boca Raton, FL, USA, 2021; pp. 231–250.
56. Hopkins, D.; Schwanen, T. Talking about automated vehicles: What do levels of automation do? *Technol. Soc.* **2021**, *64*, 101488. [[CrossRef](#)]
57. Nikitas, A.; Vitel, A.E.; Cotet, C. Autonomous vehicles and employment: An urban futures revolution or catastrophe? *Cities* **2021**, *114*, 103203. [[CrossRef](#)]
58. Riaz, F.; Khadim, S.; Rauf, R.; Ahmad, M.; Jabbar, S.; Chaudhry, J. A validated fuzzy logic inspired driver distraction evaluation system for road safety using artificial human driver emotion. *Comput. Netw.* **2018**, *143*, 62–73. [[CrossRef](#)]
59. Upadhya, S.M.; Vinothina, V. Fuzzy logic based approach for possibility of road accidents. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2020; Volume 1427, p. 012011.
60. Poloni, M.; Ulivi, G.; Vendittelli, M. Fuzzy logic and autonomous vehicles: Experiments in ultrasonic vision. *Fuzzy Sets Syst.* **1995**, *69*, 15–27. [[CrossRef](#)]
61. Pawlus, W.; Karimi, H.; Robbersmyr, K. A fuzzy logic approach to modeling a vehicle crash test. *Open Eng.* **2013**, *3*, 67–79. [[CrossRef](#)]
62. Kurnaz, S.; Cetin, O.; Kaynak, O. Fuzzy logic based approach to design of flight control and navigation tasks for autonomous unmanned aerial vehicles. *J. Intell. Robot. Syst.* **2009**, *54*, 229–244. [[CrossRef](#)]
63. Aduthaya, T.C.N. Master of Applied Science in Industrial Systems Engineering. Ph.D. Thesis, University of Regina, Regina, SK, USA, 2006.
64. Wilensky, U.; Rand, W. *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*; MIT Press: Cambridge, MA, USA, 2015.
65. Hjorth, A.; Head, B.; Brady, C.; Wilensky, U. Levelspace: A netlogo extension for multi-level agent-based modeling. *J. Artif. Soc. Soc. Simul.* **2020**, *23*, doi:10.18564/jasss.4130. [[CrossRef](#)]
66. Kahn, K. An introduction to agent-based modeling. *Phys. Today* **2015**, *68*, 55. [[CrossRef](#)]

Article

ROAD Statistics-Based Noise Detection for DME Mitigation in LDACS

Miziya Keshkar ¹, Raja Muthalagu ^{2,*} and Abdul Rajak ¹

¹ Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science Dubai Campus, Dubai 345055, United Arab Emirates; miziyak@gmail.com (M.K.); abdulrazak@dubai.bits-pilani.ac.in (A.R.)

² Department of Computer Science, Birla Institute of Technology and Science Dubai Campus, Dubai 345055, United Arab Emirates

* Correspondence: raja.m@dubai.bits-pilani.ac.in

Abstract: Interference mitigation in L-band Digital Aeronautic Communication Systems (LDACS) from legacy users is extremely important as any error in data retrieval of aeronautic communication can adversely affect flight safety. This paper proposes an LDACS receiver prototype which uses rank-ordered absolute differences (ROAD) statistics to detect the distance measuring equipment (DME) interference. The detected DME interference is reduced in the next stage by pulse blanking. The performance of the proposed ROAD pulse blanking method (ROAD PB) is compared with the existing interference mitigation methods which use the amplitude of the received signal for the detection of DME interference. In depth analysis of the obtained results affirms that the proposed ROAD value-based interference detection excels amplitude-based detection. For an SNR value of 0 dB, the proposed method of detection could achieve a 3% increase in terms of accuracy with a reduction of 4% in false alarms. With the advantage of ROAD statistics detection, the proposed ROAD PB could achieve an SNR saving of 2.7, 1.1, 0.7, 0.25 and 0.2 dBs at BER 10^{-1} in comparison with pulse blanking, Genie-aided estimation enhanced pulse peak attenuator (GAEPPA), GAE enhanced pulse peak limiter (GAEPPPL), optimum Bayesian estimator enhanced pulse peak attenuator (OBEPPA) and OBE enhanced pulse peak limiter (OBEPPPL). The comparative results show that the proposed ROAD pulse blanking outperformed the other techniques for the optimum threshold value of the operation.

Keywords: OFDM; LDACS; aeronautical communication; impulse noise; pulse blanking; ROAD statistics

Citation: Keshkar, M.; Muthalagu, R.; Rajak, A. ROAD Statistics-Based Noise Detection for DME Mitigation in LDACS. *Appl. Sci.* **2022**, *12*, 3774. <https://doi.org/10.3390/app12083774>

Academic Editors: Giovanni Randazzo, Dimitrios S. Paraforos, Stefania Lanza and Anselme Muzirafuti

Received: 17 January 2022

Accepted: 1 April 2022

Published: 8 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Air traffic growth is happening at a very rapid rate. As per Eurocontrol's latest study report about European aviation in 2040, the air traffic growth will be limited by the available capacity at the airports. This can lead to a rapid increase in congestion at the airport, which in turn can cause extra pressure on the network and more delays [1]. To accommodate this huge increase in air traffic, an efficient air traffic management system (ATM) supported by a secure and spectrum-efficient Communications, Navigation and Surveillance (CNS) framework is needed.

The existing air traffic management system is supported by voice and data communications systems. The main voice communication media for air to ground communication is still analog. The existing analog VHF double sideband amplitude modulation (DSB-AM) will remain in service for many more years as it ensures safe and reliable communication with the use of low-cost communication equipment. However, this technology becomes a hindrance in deploying new ATM applications, such as flight centric operation with point-to-point communications [2].

Similar to voice communication, data communication to the cockpit is also ensured by ground-based equipment operating within HF or VHF radio bands. The communication is through narrowband radio channels, which limits the data throughput to some kilobits

per second. These data links are insufficient to provide broadband services now or in the future with the existing VHF and HF spectrum [2]. Hence, the International Civil Aviation Organization (ICAO) has recommended Future Communications Infrastructure (FCI) to modernize the existing communication links with new spectrum-efficient and reliable infrastructure, which can support new ATM applications and broadband services. This led to the development of the L-band (960–1164 MHz) Digital Aeronautical Communication Systems (LDACS) [3].

In 2009, the specification of LDACS was proposed. ICAO suggested two possible standards: LDACS1 derived from the IEEE 802.16 wireless system [3] and LDACS2 derived from the global system for mobile communication (GSM) [4]. LDACS1 uses advanced network protocols of current commercial standards. It is a broadband multicarrier system based on orthogonal frequency division multiplexing (OFDM). LDACS2 uses protocols that offer high QoS communications. It is a narrow band single carrier system based on Gaussian minimum shift keying modulation (GMSK). It is expected to accommodate the huge increase in air traffic with the deployment of any of these subsystems of LDACS.

As shown in Figure 1, L-band is already providing services to legacy users such as distance measuring equipment (DME), military tactical air navigation (TACAN) system and joint tactical information distribution system (JTIDS), which are used for navigation aids. Apart from these, universal access transceiver (UAT) at 978 MHz and secondary surveillance radar (SSR) and airborne collision avoidance system at 1030 and 1090 MHz are also allotted with fixed channels [5–8]. However, studies about spectrum occupancy revealed that large portions of the L-band spectrum are used less frequently or underutilized [3,4]. Hence, the LDACS system is deployed in the L-band either as an inlay system between the legacy users or as an overlay system in the unoccupied spectrum [9]. The overlay method is selected for LDACS2 (960–975 MHz). Though the method is less complex, spectrum scarcity is a noticeable challenge [10–12]. The inlay approach, chosen in LDACS1, is expected to overcome the challenge of spectrum scarcity by utilizing the 1 MHz spectral gap between legacy user DME, thereby increasing the spectrum utilization.

The comparative studies between LDACS1 and LDACS2 affirmed that LDACS1 is the preferred choice over LDACS2. LDACS1 is highly capable of supporting high-speed delay-sensitive multimedia services and is also compatible with the cellular communication standards. LDACS1 is further referred to as LDACS [13,14]. Hence, the work presented in this paper uses LDACS to refer to LDACS1 hereafter.

LDACS involves two way communication: a forward link (FL) from the ground station (GS) to air station (AS) and a reverse link (RL) from AS to GS. It provides frequency division duplexing (FDD) of 63 MHz spacing between FL (962–1213 MHz) and RL (1025–1150 MHz) with the opportunistic access of paired spectrum. The deployment of LDACS in L-band gives rise to interferences to license users and vice versa. The possible interference scenario for LDACS is portrayed in Figure 2. In words, it can be affirmed as follows: (a) LDACS FL is impaired by DME GS (FL), not by DME airborne station (AS), as (RL) is not active in this part of spectrum, (b) LDACS RL is impaired by both the DME GS (FL) and DME AS (RL), (c) DME FL is impaired by interference from both the LDACS GS (FL) and LDACS AS (RL) and (d) DME RL is impaired with interference from LDACS AS (RL), not from LDACS GS (FL), as it is not active in this part of the spectrum.

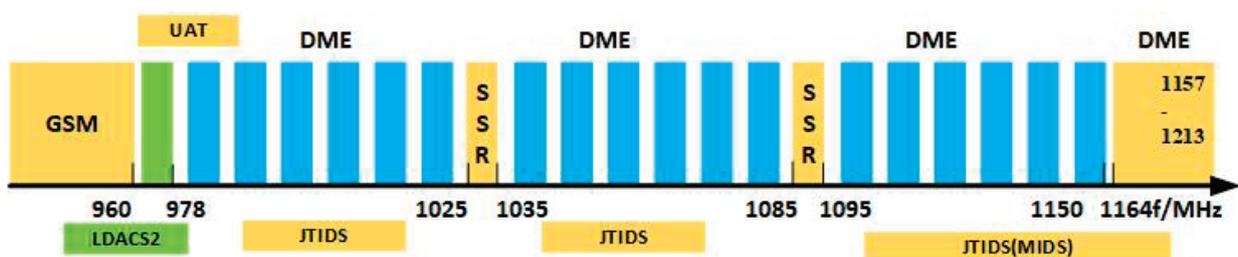


Figure 1. L-band Spectrum Occupancy [9].

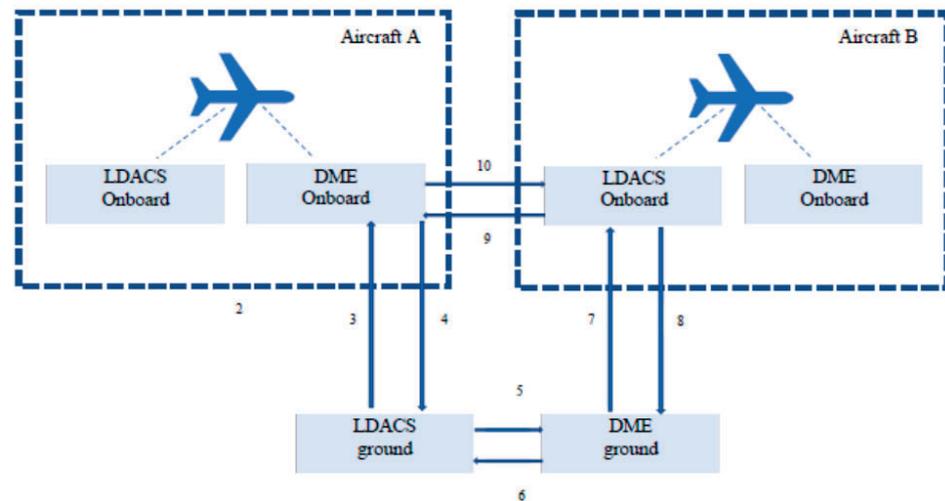


Figure 2. Interference Between DME and LDACS [15].

Identifying the possible interference scenario, the role of interference mitigation is recognized as critically important as any malfunctioning of the licensed system can affect flight safety. The rest of the paper is oriented as follows: Section 2 lists the literature survey of existing DME interference mitigation techniques in LDACS with their advantages and drawbacks. Section 3 expounds on the system and noise model used for this work. Section 4 elaborates the theory and functioning of the proposed method in spotting and reducing DME interference. The section also includes the detailing of nonlinear methods used for the comparison of the proposed method in DME mitigation. Section 5 elucidates the results obtained for the proposed method in terms of reducing the bit error rate in received data with other DME interference reduction techniques. The following are the list of abbreviations used in this paper.

2. Related Work

Several interference mitigation techniques for LDACS are available in the literature; among these, most of the proposed schemes focus on mitigating DME interference. In 2011, two methods capable of detecting and mitigating DME interference were put forward. The monotonous structure and spectral shape of the DME pulse are exploited in proposing these two methods. The methods achieve the merit of simpleness at the cost of losing a portion of data resulting from pulse blanking [16]. In 2012, Hailiang Wang et al. put forward a concoction of pulse blanking and notch filter to reduce the interference. Though the method gives leeway in the time and frequency domain of operation, it is more safe and effective for the B2 band (1900 MHz) signal [17]. Further, Yun Bai et al. proposed a DME mitigation scheme that advantages complementary code keying (CCK). Though the CCK encoding has the merit of better gain, it requires low phase distortion and a wideband channel. The latency of the system is high due to the large acquisition time. Moreover, the modulation employed here is not a power-efficient modulation [18].

In 2014, Q. Li et al. proposed an iterative receiver design [19]. The design employs iterative decoding between the demodulator and decoder based on the Turbo principle. Another type of selective pulse blanking method to curtail DME interference is put forward in [20]. The method bestowed a designed fast filter bank for this purpose. In 2016, Li Douzheetal et al. propounded a method based on deformed pulse pair spotting and its subtraction from the actual signal [21]. Later, Khodr A. Saaifan and Werner Henkel insinuated lattice signal sets to resist DME interference. Precoding based on lattice signal set at the transmitter changes the shape of the DME signal spectrum. A simple clipping technique is then applied for DME mitigation [22].

The major flaw of pulse blanking is recognized as intercarrier interference. In [9], decision-directed noise estimation is put forward to eliminate the intercarrier interference.

Reduced throughput of transmission and wastage of power are the drawbacks of this scheme. LDACS-OFDM based on discrete wavelet transform (DWT) is discussed in [23]. The scheme makes use of the real nature of DME. The signal affected with DME is selectively transmitted through the quadrature channel. With the effective utilization of direct sequence spread spectrum to combine the in phase and quadrature phase, the method proclaims the elimination of DME. The computational complexity and resource requirement are the flaws present in this scheme. In [24], an energy-based DME detector has been proposed. The detector works with adaptive threshold value in order to obtain the best trade-off between DME signal detection and false alarm. In 2021, a deep clipping-based DME noise reduction technique is propounded in [25]. It is a linear clipping method that uses two threshold levels for recognition and reduction in DME. Detailed study of the existing interference mitigation techniques in LDACS exposes active research is going on in this area.

The Genie-aided estimator (GAE) taps the statistical description of the side information to generate the design parameters to accomplish lower bounds on the bit error at the receiver [26]. However, details of the side information are required to accomplish this lower bound performance. The correlation of the impulsive noise or the frequency in impulsive noise arrival time is some other side information [27]. When a Gaussian source is influenced by uncorrelated impulse noise, it is possible to attain optimum system performance with the use of a Bayesian signal estimator. In 2013, P. Banelli proposed an optimal Bayesian estimator (OBE) mainly for real-valued Gaussian mixture noise [28]. Later, the method was further upgraded for complex signals in 2015 [29]. It is possible to propose different types of pulse peak attenuators and pulse peak limiters for DME mitigation with the estimation knowledge obtained from GAE and OBE [30]. In this paper, we have used GAE and OBE enhanced pulse peak attenuators and limiters to compare the performance of the proposed ROAD pulse blanking.

ROAD statistics-based impulse detector was proposed in 2005 to detect the impulse pixels in an image. The idea can be extended to remove any mix of Gaussian and impulse noise [31]. In [32], the ROAD value of the received signal is used as one of the inputs to train the deep neural network for the detection of signal instances corrupted with impulse noise. The most affected or least acceptable data are present on those subcarriers whose powers are much different from neighboring subcarriers at each time epoch. Hence, it is possible to use ROAD statistics to identify the subcarriers affected with impulse noise. To the best of our knowledge, no work has been reported that employs ROAD statistics for the detection of DME interference to date.

In this paper, ROAD pulse blanking is proposed which uses ROAD statistics for the detection of DME interference and pulse blanking for noise mitigation. The performance of the proposed method is compared with absolute value-based DME interference detection methods such as pulse blanking, GAE enhanced pulse peak attenuators/limiters and OBE enhanced pulse peak attenuators/limiters.

All these methods include two basic operations:

1. Detection of DME interference.
2. Mitigation of DME interference.

The advantage of the proposed method compared to other methods is that it could identify affected subcarriers more accurately and hence could eliminate noise more effectively. The performance of pulse blanking is observed to be improved when ROAD statistics-based noise detection has been employed. The improvement in performance is such that it outperformed GAE enhanced pulse peak attenuators/limiters and could stand with OBE enhanced pulse peak attenuator/limiter.

3. System and Noise Model

LDACS System Model

The system model includes the LDACS transmitter, the channel which imparts additive white Gaussian (AWGN) noise and DME interference and the LDACS receiver. Figure 3

elaborates the detailed block diagram of the LDACS ground station transmitter. Data source creates random data of 91 bytes and passes to Reed–Solomon (RS) coder resulting in an extra 10 bytes of redundant data for error correction and detection. A 6-bit zero padding is performed on the output of the RS coder before passing to the convolutional coder. The coded output of the convolutional coder is made to pass through permutation interleaver for reducing burst errors. The output of permutation interleaver is arranged into a standard LDACS data format (F) after symbol mapping, modulating with Quadrature Phase Shift Keying (QPSK) and frame composing. All the variables shown in Figures 3 and 4 are generated for standard LDACS data frame format (F). The same variables with suffix ‘t’ signify the same signal for an instant ‘t’ or the t^{th} OFDM symbol. In other words, $F_t = [F_t[0], F_t[1] \dots F_t[N - 1]]^T$ stands for the t^{th} symbol of LDACS forward link frame (F) with N orthogonal subcarriers. F_t carries the random data $F_t[m]_{m=0,1,\dots,N-1}$ with zero mean and variance σ_F^2 . The OFDM symbol $S_t = [S_t(0), S_t(1) \dots S_t(N - 1)]^T$ is generated in the time domain by calculating the 64-point IFFT of the data F_t . Further, N_{CP} number of cyclic prefix bits are added to the total N subcarriers resulting in transmitted vector $X'_t = [X'_t(0), X'_t(1) \dots, X'_t(N + N_{cp} - 1)]$.

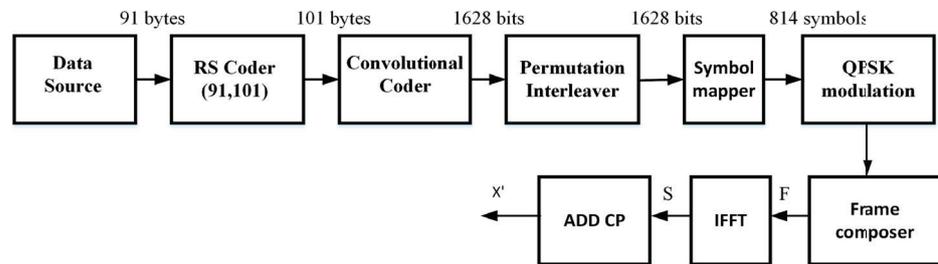


Figure 3. LDACS transmitter block diagram.

The transmitted signal X' changes to r' when it passes through the channel in the presence of DME. For each instant t , the transmitted vector X'_t is affected by a noise component $i_t = [i_t(0), i_t(1), \dots, i_t(N + N_{cp-1})]^T$ which is a mixture of the additive white Gaussian noise $A_t = [A_t(0), A_t(1), \dots, A_t(N + N_{cp-1})]^T$ and the impulse noise (DME) $p_t = [p_t(0), p_t(1), \dots, p_t(N + N_{cp-1})]^T$. Thus, the received signal for an instant ‘t’ is $r'_t = [r'_t(0), r'_t(1), \dots, r'_t(N + N_{cp} - 1)]^T$ and can be denoted as in (1).

$$r'_t = X'_t + i_t \tag{1}$$

where

$$i_t = A_t + p_t \tag{2}$$

As discussed in Section 1, the prime contributor of interference to LDACS is DME. These signals are a pair of Gaussian-shaped pulses, separated by a duration of Δt . The transmission rate (30 pulse pairs per second or 50 ppps), as well as the duration Δt (12 or 36 μs) of DME signals, varies with the mode of operation of the distance measuring equipment. A pair of DME pulses in the baseband can be expressed as in (3) [33].

$$P_d(t) = e^{-\frac{\zeta t^2}{2}} - e^{-\frac{\zeta(t-\Delta t)^2}{2}} \tag{3}$$

where $\zeta = 4.5 \times 10^{11} \text{ s}^{-2}$.

It has a width of 3.5 μs at half of the maximum amplitude. The frequency domain representation of DME signal is as in (4). The spectrum is modulated with a cosine as the pulses are always happening pairwise [34].

$$I_{pd}(f) = \sqrt{\frac{8\pi}{\zeta}} e^{\frac{2\pi^2 f^2}{\zeta}} \cdot e^{-j\pi f \Delta t} \cos(\pi f \Delta t) \tag{4}$$

The base band DME pulse pairs are modulated to the relative carrier frequency of the channel to 0.5 MHz left and to 0.5 MHz right of the LDACS system bandwidth. The DME interfering signal that affects the LDACS system is expressed in (5). $I_{pd}(t)$ is the total interference signal for a time interval 't' caused by N number of DME stations that are operating on the 0.5 MHz offset to the center frequency of the LDACS system [35].

$$I_{pd}(t) = \sum_{i=0}^{N_{pd}-1} \sum_{l=0}^{N_i-1} \sqrt{P_{i,l}} P_d(t - t_{i,l}) e^{j2\pi f_{c,i}t + j\chi_{i,l}} \tag{5}$$

where N_{pd} is the total number of interfering DME stations, N_i is the total number of pulse pair in the particular time interval for the i th interfering DME station, $P_{(i,l)}, \chi_{(i,l)}$ are power and phase of the pulse pair, respectively, $f_{(c,i)}$ —the relative carrier frequency of the i th interfering DME station and $t_{(i,l)}$ is the starting time of the l th pulse pairs of the i th DME station. The methods used to reduce impulse noise work well to reduce DME noise also.

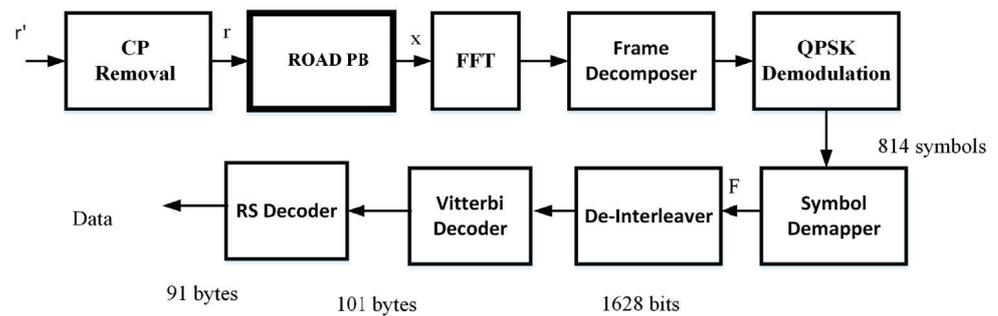


Figure 4. Proposed LDACS receiver block diagram.

The detailed block diagram of the proposed LDACS FL AS receiver is sketched in Figure 4. The first block in receiver removes the cyclic prefix bits associated with the received signal $r' = [r'_t(0), r'_t(1), \dots, r'_t(N + N_{Cp-1})]^T$ resulting in signal $r_t = [r_t(0), r_t(1), \dots, r_t(N - 1)]^T$. The nonlinear device ROAD PB detects the DME interference in LDACS signal r (with the clever use of ROAD statistics) and performs pulse blanking to reduce the bit error rate in received data. In general, the resulting vector can be defined as $x = f(r)$ or $x_t = f(r_t)$, where $f(\cdot)$ is the nonlinear function with enough intelligence to sense DME interference. The nonlinear devices discussed in this paper process the signal r_t in dissimilar ways to sense the DME interference. Moreover, the nonlinear estimators used for the performance comparison of the proposed method utilize one more vector π_t to estimate the received signal data. Hence, the definition of function $f(\cdot)$ varies with different nonlinear devices.

The nonlinear device is operated on the signal r before the DFT processing to block the dispersion of sparse time domain impulses $p_t[n]$ over all the OFDM carriers in the frequency domain.

4. Nonlinear Estimators

As discussed in the system model, the nonlinear device is designed to detect and eliminate DME interference from the LDACS AS receiver. In this paper, the proposed nonlinear device ROAD PB uses ROAD statistics for the detection of DME interference and pulse blanking for the mitigation of interference. The performance of the method is compared with the conventional pulse blanking, which uses the amplitude of the received signal for the detection of DME interference. In addition, nonlinear estimators such as GAE enhanced pulse peak processors and OBE enhanced pulse peak processors are also used for the mitigation of DME interference to compare the performance of ROAD PB.

The functioning of the nonlinear devices ROAD PB, pulse blanking, GAEPPA, GAEPPL, OBEPPA and OBEPL in the detection and elimination of DME noise are discussed below.

4.1. Proposed ROAD PB

When the LDACS signal is affected by DME interference, the amplitude of the received data exhibits a large variation in amplitude from neighboring data. The variation in amplitude due to bit change or additive white Gaussian noise is less due to DME interference. Therefore, the most affected data are those whose amplitude is much different from neighbors. The proposed ROAD PB uses ROAD statistics to quantify the variation in amplitude of particular LDACS (OFDM) data from the neighboring data for each time epoch. Further, the most basic threshold-based detection approach is utilized to identify the data signal affected by DME interference. The goal of fixing a suitable threshold is to recognize the OFDM data that are significant outliers.

The calculation of ROAD value [32] for a one-dimensional LDACS OFDM symbol involves the following steps:

1. The received OFDM symbol for each time epoch is considered as a one-dimensional vector. For a one-dimensional vector of size $(1 \times 2f + 1)$, the absolute difference between a center sample and a receiving sample for each time epoch $f_d(k)$ are calculated as in (6)

$$f_d(k) = |r_k - [r_{k-f}, \dots, r_{k+f}]| \tag{6}$$

2. The difference vector ($f_d(k)$) is sorted in increasing order.

$$Q(k) = \text{sort}(f_d(k)). \tag{7}$$

3. The ROAD value is calculated as the sum of first f values of $Q(k)$

$$\text{ROAD} = \sum_{k=1}^f Q(k). \tag{8}$$

The simple way to recognize the effectiveness of ROAD statistics is to incorporate the method into any existing DME mitigation method. Hence, the proposed ROAD PB incorporates ROAD statistics into pulse blanking. The mathematical depiction of ROAD pulse blanking is as in (9). Here, R_p is the lower threshold value used to discriminate LDACS signals affected with DME interference.

$$x_{|R_p|}(r) = \begin{cases} |r|e^{j\arg(r)} & \text{if } \text{ROAD}(r) \leq R_p \\ 0 & \text{Otherwise} \end{cases} \tag{9}$$

The reduction in bit error rate at the receiver is compared with normal pulse blanking in Section 5.

The extra computational complexity put forward by ROAD PB compared to conventional pulse blanking is the sum of the computational complexity put forward by the steps involved to calculate ROAD value, as in (6)–(8). As these three steps have no complex multiplication in calculating the ROAD value, it is evident that the extra complex multiplication contributed by ROAD PB is zero. Hence, there is no change in number of complex multiplication compared to LDACS OFDM receiver (without any mitigation) or to pulse blanking. As we have used fast Fourier transform in LDACS OFDM receiver, the total number of complex multiplications involved is $(N/2) \cdot \log_2(N)$, where N is the number of subcarriers in OFDM signal [36].

The steps in (6) introduce the extra complex additions $N \cdot (N - 1)$ or real number addition $2N \cdot (2N - 1)$. Hence, the total number of complex additions of LDACS OFDM receiver with ROAD PB is $N \cdot \log_2 N + N \cdot (N - 1)$. The number of real-time additions put forward by sorting depends on the type of sorting that is used. For instance, if selective sorting is used, it introduces $\frac{N^2 \cdot (N-1)}{2}$ number of real-time additions. Finally, the number of real value additions introduced by step (8) is $\frac{N \cdot (N-1)}{2}$. The proof is included in Appendix A.

4.2. Conventional Pulse Blanking

The method pulse blanking makes the signal zero if the absolute value of the received signal is above a particular threshold value α_p . The mathematical depiction of pulse blanking is as in (10),

$$x_{|p|}(r) = \begin{cases} |r|e^{j\arg(r)} & \text{if } |r| \leq \alpha_p \\ 0 & \text{Otherwise} \end{cases} \quad (10)$$

4.3. Pulse Peak Processors

As discussed in the system model, data estimation needs one more vector π_t along with r_t in performing the nonlinear function $f(\pi_t, y_t)$. The detailing of the nonlinear operations performed by pulse peak attenuators and limiters are depicted in block diagrams Figures 5 and 6, respectively. The received data (after CP removal) r are passed through a 2-GMM estimation block to extract the vector π_t . The parameters contained in vector π_t are further used to compute scaling factor μ_t for each time epoch. From Figure 5, it is clear that pulse peak attenuator uses this parameter μ for processing the signal r_t .

From Figure 6, it is to be noted that pulse peak limiters have one extra block compared to pulse peak attenuators. It is denoted as a decision device that holds the algorithm to change or update the scaling factor. The decision device determines if the scaling factor is needed to modify or not. When the scaling is performed with a modified scaling factor, pulse peak attenuators become pulse peak limiters. In this paper, we have included two types of pulse peak attenuators and four types of pulse peak limiters to compare the performance of proposed ROAD statistics.

It is inevitable to discuss K-GMM modeling to understand the vector π_t in more detail. This section outlines how GAE and OBE enhanced pulse peak processors exploit 2-GMM estimation (2-GMM) to scale the signals affected with DME.

Any random variable can be expressed as the combination of K-number mutually exclusive Gaussian variables with K-GMM modeling [37,38]. Hence, this model can be effectively applied to any ImpN distribution (Class A, S- α -S noises, etc.), either estimated [39–41] and approximated by a K-GMM [42] or modeled with the actual equation [43]. The K-GMM model is mathematically expressed with pdf,

$$f_W(i) = \sum_{k=0}^{K-1} P_k \cdot G(i, \sigma_k^2) \quad (11)$$

where $\{P_k\}_{k=0,1,\dots,K-1}$ with $\sum_{k=0}^{K-1} p_k = 1$ are the probability of occurrence of each Gaussian component k . For the value $k = 0$, the component $i_0 \approx G(i_0, \sigma_0^2)$ represents the thermal noise with variance σ_0^2 and with the probability of occurrence p_0 . For values $k = 1$ to $K - 1$, the statistical combinations of components characterize the impulse noise with the probability $p_I = 1 - p_0$ and noise power σ_I^2 . The ratio of thermal noise to impulse noise is expressed as $\Gamma = \frac{\sigma_0^2}{\sigma_I^2}$. For the value $K = 2$, this model will reduce to 2-GMM with a thermal noise component $i_0 \approx G(i_0, \sigma_0^2)$ and an impulse noise component $i_1 \approx G(i_1, \sigma_1^2)$.

The 2-GMM model is simple and assumes the presence of a strong impulsive noise as the recognition of only two mutually exclusive events, with probability p_0 and p_1 . Hence, 2-GMM is exploited to employ GAE and OBE enhanced pulse peak processors as DME interference mitigators in LDACS receivers. The received signal r , when passed through 2-GMM estimation, results in parameters thermal noise component and impulse noise component with a probability of p_0 and p_1 for each time epoch. Thus, the vector π holds parameters $\sigma_0^2, \sigma_1^2, p_0$ and p_1 obtained from 2-GMM estimation. These parameters are used to calculate the instant scaling parameter μ_t to apply instant nonlinearity to the affected subcarriers in the time domain. The parameter μ varies with different types of pulse peak processors.

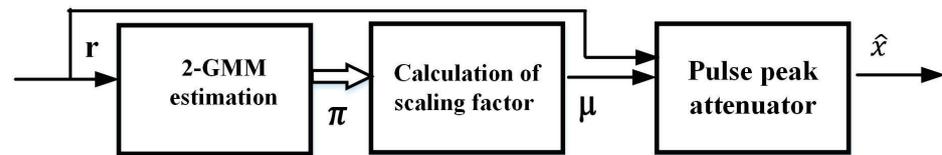


Figure 5. General block diagram for pulse peak attenuator.

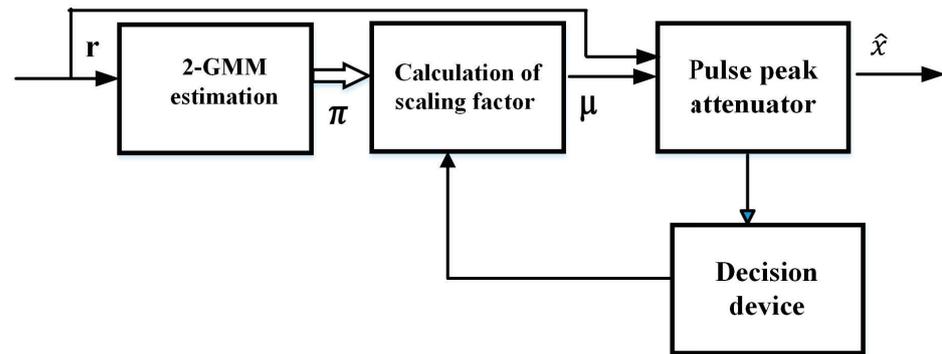


Figure 6. General block diagram for pulse peak limiter.

GAE Enhanced Pulse Peak Processors

The pulse peak processors included in this section are GAE enhanced PPA and GAE enhanced PPLs (Type I and Type II). These pulse peak processors have better performance than GAE by utilizing other side information, such as impulsive noise arrival time or relationship of the impulsive noise [30].

The received signal r can be expressed as the sum of transmitted signal X and noise $i|_k$, where X and $i|_k$ are two zero-mean independent Gaussian random variables with variances σ_X^2 and σ_i^2 . With any impulse noise, when modeled as properly weighted mutually exclusive Gaussian events, the GAE claims to know which is the k th Gaussian component of the (pdf), which actually affects the transmitted signal at each time epoch. Once the signal affected with DME interference is identified, all three GAE enhanced pulse peak processors use the same knowledge of GAE for pulse peak processing. The amplitude of the received signal is used for the detection of DME as in pulse blanking.

The GAE enhanced PPA attenuates the nonlinear input r when the amplitude of the received signal exceeds the threshold value. The operation of GAE enhanced PPA can be expressed as follows:

$$\hat{x}_{|kPA|}(r) = \begin{cases} |r| & \text{if } |r| \leq \alpha_p \\ \rho_k \cdot |r| & \text{otherwise.} \end{cases} \tag{12}$$

where $\rho_k = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_k^2}$

$$\text{And, } \sigma_k^2 = (1 + \frac{k}{A\Gamma})\sigma_0^2 = \frac{k/A + \Gamma}{1 + \Gamma} \sigma_i^2 = \frac{k}{A\Gamma} \sigma_i^2 + \sigma_0^2 \tag{13}$$

As the device is a pulse peak attenuator, the scaling factor μ is ρ . The scaling factor changes at each instant t as it is a function of r_t and π_t .

GAE enhanced PPA has better performance than the pulse blanking method as it attenuates the DME affected signal rather than losing the data by blanking. The device is well suited to process complex data signals as in LDACS with the modified equation as follows:

$$\hat{x}_{|kPA|}^*(r) = \begin{cases} |r|e^{j\arg(r)} & \text{if } |r| \leq \alpha_p \\ \rho_k \cdot |r|e^{j\arg(r)} & \text{otherwise.} \end{cases} \tag{14}$$

GAE enhanced PPL is a modified or improved form of GAE enhanced PPA. It employs a decision device to update the scaling factor to attenuate the received signal continuously until the amplitude of the received signal reaches the threshold value (α_p). The repeated attenuation will not affect other subcarriers which are not affected with DME noise. This nonlinear device has better performance at high SNR values compared to the GAE enhanced PPA. The GAE enhanced PPL processes the input signal r and delivers output $\hat{x}_{kPL}(r)$ as stated in (15).

$$\hat{x}_{|kPL|}(r) = \begin{cases} r & \text{if } |r| \leq \alpha_p \\ \rho_{mod} \cdot r & \text{otherwise.} \end{cases} \quad (15)$$

where $\rho_{mod} = \frac{\sigma_x^2}{\sigma_x^2 + N \cdot \sigma_k^2}$.

Here, the value of N varies directly with the difference in power of received signal and threshold peak detection value at each instant. The maximum value of N occurs when the resulting signal holds no values greater than α_p . With this knowledge, the value of N_{max} is as in (16). The algorithm to support this derivation is from [30]

$$N_{max} = \frac{\sigma_x^2 (|r|_{max} - \alpha_p)}{\alpha_p \sigma_k^2} \quad (16)$$

Similar to GAE enhanced PPA, GAE enhanced pulse peak limiters also reduce the drawback of the pulse blanking method with less complexity. The same operation can be performed in another way as in (17)

$$\hat{x}_{|kPLs|}(r) = \begin{cases} r & \text{if } |r| \leq \alpha_p \\ M \cdot \rho_k \cdot r & \text{otherwise.} \end{cases} \quad (17)$$

where the maximum value of value of M is derived as in (18) [30]

$$M_{max} = \frac{\alpha_p}{\rho_k \cdot |r|_{max}} \quad (18)$$

From Equations (15) and (17), the updated scaling factors (μ) for GAEPPL Type I and Type II are identified as ρ_{mod} and $M \cdot \rho_k$. Both the methods are applicable to perform scaling of complex valued data with a slight change in Equations (15) and (17) resulting in (19) and (20), respectively.

$$\hat{x}_{|kPL|}^*(r) = \begin{cases} |r|e^{jarg(r)} & \text{if } |r| \leq \alpha_p \\ \rho_{mod} \cdot |r|e^{jarg(r)} & \text{otherwise.} \end{cases} \quad (19)$$

where $\rho_{mod} = \frac{\sigma_x^2}{\sigma_x^2 + N \cdot \sigma_k^2}$.

$$\hat{x}_{|kPLs|}^*(r) = \begin{cases} |r|e^{jarg(r)} & \text{if } |r| \leq \alpha_p \\ M \cdot \rho_k \cdot |r|e^{jarg(r)} & \text{otherwise.} \end{cases} \quad (20)$$

In both cases, the definition for ρ_{mod} and M remains the same as that used in Equations (15) and (17).

4.4. OBE Enhanced Pulse Peak Processors

Bayesian estimators are useful in any Gaussian source affected by any Gaussian-mixture noise [28]. The time domain OFDM signal x can be approximated by Gaussian pdf, $f_X(x) = G(x; \sigma_x^2) = \frac{x^2/2\sigma_x^2}{\sqrt{2}}$. The complex valued received signal $r_t[n]$ at the receiver side has real and imaginary parts $r_{t,R}[n]$ and $r_{t,I}[n]$, respectively. Consider that r represents distinctly either the real or the imaginary part of $r_t[n]$. When the received signal of interest

is modeled or approximated as a Gaussian pdf or K-component pdf, the minimum mean square error Bayesian estimators can be effectively utilized along with the knowledge of the signal x_t . By exploiting the statistical dependency between X and i , it is possible to write $f_{(r|X)}(y) = f_i(r - x)$ and $G(r; \sigma_{rk}^2) = G(r; \sigma_x^2) * G(r; \sigma_k^2)$. Here, $*$ stands for convolution operation. Thus, the received noise power σ_{rk}^2 is the sum of the signal power σ_x^2 and k th Gaussian component noise power σ_k^2

$$\sigma_{rk}^2 = \sigma_r^2 + \sigma_k^2 \tag{21}$$

The OBE enhanced pulse peak processors perform Bayesian estimation only when the received signal is identified as DME affected signal. The pulse peak processors included in this section are OBE enhanced pulse peak attenuator and OBE enhanced pulse peak limiters (Type I and Type II).

The device attenuates the nonlinear input r when the amplitude of the received signal is above α_p . The mathematical expression of OBE enhanced pulse peak attenuator is as in (22).

$$\hat{x}_{|kOPA|}(r) = \begin{cases} |r| & \text{if } |r| \leq \alpha_p \\ \beta_o(r) \cdot r & \text{otherwise.} \end{cases} \tag{22}$$

where

$$\beta_o(r) \cdot r = \frac{\sum_{k=0}^{K-1} \rho_k p_k G(r; \sigma_{rk}^2)}{\sum_{k=0}^{K-1} p_k G(r; \sigma_{rk}^2)} \tag{23}$$

As the device is a pulse peak attenuator, the scaling factor μ is $\beta_o(r)$. It is possible to use this OBE enhanced PPA for processing complex valued signal as well. The mathematical statement for this operation is as given in (24).

$$\hat{x}_{|kOPA|}(r) = \begin{cases} |r| & \text{if } |r|e^{j\arg(r)} \leq \alpha_p \\ \beta_o(r) \cdot |r| \cdot e^{j\arg(r)} & \text{otherwise.} \end{cases} \tag{24}$$

OBE enhanced PPL is an altered or upgraded form of OBE enhanced PPL, where it reduces the amplitude of the received signal unceasingly until the amplitude of the DME affected subcarrier reaches the threshold value. As the repeated attenuation is performed only for the subcarriers which exceed the threshold value, it will not disturb the subcarriers which are not affected with DME interference. The OBE enhanced PPL process the input signal r and deliver output $x_{kOPL}(r)$ as stated in (25).

$$\hat{x}_{|kOPL|}(r) = \begin{cases} |r| & \text{if } |r| \leq \alpha_p \\ \beta_{mod}(r) \cdot |r| & \text{otherwise.} \end{cases} \tag{25}$$

where

$$\beta_{mod}(r) = \frac{\alpha_p}{|r|} \tag{26}$$

Here, value of $\beta_{mod}(r)$ is the modified scaling factor μ . The modification can be performed in two ways so that $\beta_{mod}(r) \cdot |r|$ becomes equal to α_p .

In one method, P multiples of $\beta_o(r)$ is considered as $\beta_{mod}(r)$. In this situation, the maximum value of P for limiting the output (P_{max}) can be expressed as in (27) [30].

$$P_{max} = \frac{\alpha_p}{\beta_{mod}(r)|r|_{max}} \tag{27}$$

In the second method, the value of the noise power component σ_k^2 is boosted R times so that the output of the nonlinear device is limited to the threshold value α_p . In this case, the modified scaling factor $\beta_{mod}(r)$ can be expressed as in (28) [30].

$$\beta_{mod}(r) = \frac{\sum_{k=0}^{K-1} \rho_{mod} p_k G(r; \sigma_{rkmod}^2)}{\sum_{k=0}^{K-1} p_k G(r; \sigma_{rkmod}^2)} \quad (28)$$

where $\rho_{mod} = \frac{\sigma_x^2}{\sigma_x^2 + R \cdot \sigma_k^2}$ and $\sigma_{rkmod}^2 = \sigma_x^2 + R \cdot \sigma_k^2$. Both the methods are adaptable to complex valued OFDM data signals as in LDACS. This can be stated mathematically as in (29)

$$\hat{x}_{|kOPL|}^*(r) = \begin{cases} |r|e^{jarg(r)} & \text{if } |r| \leq \alpha_p \\ \beta_{mod}(r) \cdot |r|e^{jarg(r)} & \text{otherwise.} \end{cases} \quad (29)$$

5. Results and Discussions

This section exposes the advantages of ROAD statistics-based sensing over amplitude-based sensing in LDACS FL communication. The results obtained from the detailed analysis of threshold ROAD value-based sensing for different threshold values under different SNR conditions are distinctly presented. The section also discusses the performance of the proposed ROAD statistics-based nonlinear device (ROAD PB) in reducing DME interference when employed in OFDM-based LDACS communication. The discussion is based on the results obtained from the MATLAB simulation of the LDACS forward link communication prototype. The performance of the proposed method is compared with the conventional pulse blanking method which uses the amplitude of the received signal for the detection of DME interference. The nonlinear devices such as GAEPPA, GAEPPL, OBEPPA and OBEPPL are also included to compare the performance of ROAD PB. The mathematical model of the LDACS FL GS transmitter (Figure 3) and LDACS FL AS receiver (Figure 3) are developed as per the standards of the LDACS system for all the inner building blocks.

At the transmitter side, random data of 91 bytes are generated by the data source and given as the input of RS coder (91,101) for external coding. Once external encoding is performed by the RS encoder, 6-bit zero padding is performed before passing through internal encoding by the convolutional encoder (171,133). The encoded bits from the output of convolutional coder, with native coding rate half, are further interleaved (using permutation interleaver) and then mapped to symbols (using symbol mapper). The mapped symbols form complex values when they pass through the QPSK modulation block. The frame composer block forms the LDACS FL data/CC frame with proper insertion of pilot values (158), null values (728) and complex data values (2442) over a total of 3328 subcarriers. Further, the time domain composite waveform of this OFDM frame is generated by passing the frame through the IFFT block of length 64. The effect of the introduction of IFFT (windowing) is canceled by adding 16 cyclic prefix bits. Table 1 holds the OFDM system parameters used in this simulation study.

To analyze the performance degradation of the LDACS FL AS receiver due to DME interference, the AWGN channel is considered. The BER variation of the received signal when passed through the AWGN channel without the influence of DME interference is obtained as shown in Figure 7. For the study of interference on LDACS, DME signals are generated by (3) for a duration Δt of 12 μ s as shown in Figure 8. The baseband DME pulse pairs are modulated to the relative carrier frequency of the channel to 0.5 MHz left and to 0.5 MHz right of the LDACS1 system bandwidth. A reduction in performance of the LDACS FL AS receiver can be observed when DME interference is allowed to affect the transmitted data. Figure 7 also shows how the existing simple noise reduction method (pulse blanking) improved the performance of the receiver. The threshold value used for the pulse blanking method is 0.3. Careful analysis of Figure 7 reveals the fact that the pulse blanking technique showed a significant improvement in the performance of the receiver at high SNR powers and a slight decrease at low SNR values. The reason for the reduction in

performance of pulse blanking (at low SNR values) is the false detection caused due to the amplitude-based sensing and the resulting extra loss of data. It is possible to reduce this number of false detections by increasing the threshold value of detection. However, high threshold value can lead to an increase in missed detection and more interference power at the output of the receiver. The false detection caused due to the amplitude-based sensing for an SNR value of 15 dB is visible (sample values between 2500 and 3000) in Figure 9.

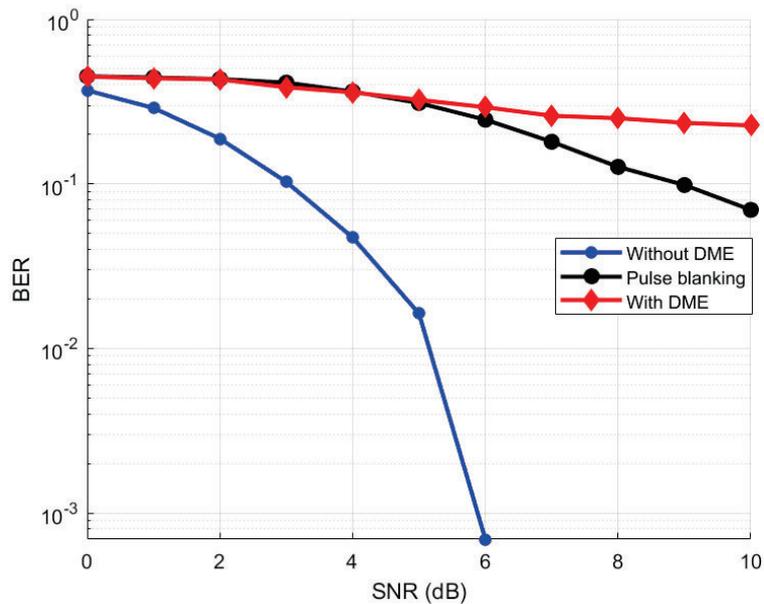


Figure 7. Performance of conventional pulse blanking technique vs. without DME interference and with DME interference.

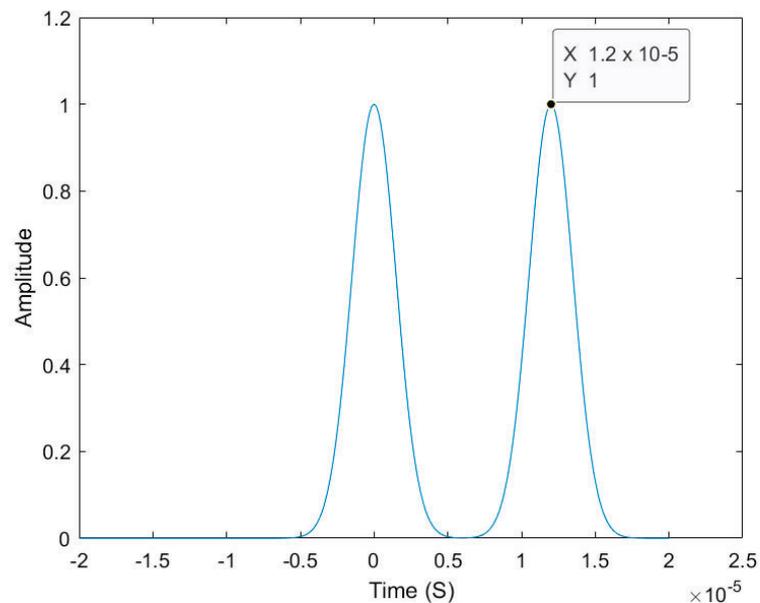


Figure 8. Standard DME pulse pair.

The amplitude of the DME interference signal, amplitude of the received signal along with the threshold value for sensing and the calculated ROAD values of the received signal are plotted in Figure 9. The signals are plotted together to figure out how both amplitude and ROAD value-based sensing accomplishes the detection of DME interference pulses. It can be observed from Figures 9 and 10 that the ROAD values of the received signal is

a magnified (though not exactly) version of absolute difference of each sample from the neighboring sample. Figure 11 depicts how well the ROAD values of the received data could identify the exact location and shape of the DME pulses than amplitude sensing. The performance of both amplitude-based sensing and ROAD statistics-based sensing for a low SNR value (0 dB) is shown in Figure 10. Comparison of Figures 9 and 10 shows that the amplitude-based sensing has less performance for low SNR value 0 dB due to the increased number of false detections. ROAD value-based sensing not only showed improved performance than the other but also preserved it (though not fully) irrespective of the SNR values as the rank-ordered difference value is considered for sensing. It has been observed that the performance of both amplitude-based sensing or ROAD value-based sensing may vary with both the SNR levels and threshold values. Hence, a detailed study of amplitude-based sensing and ROAD value-based sensing have been performed to get more insight of the process.

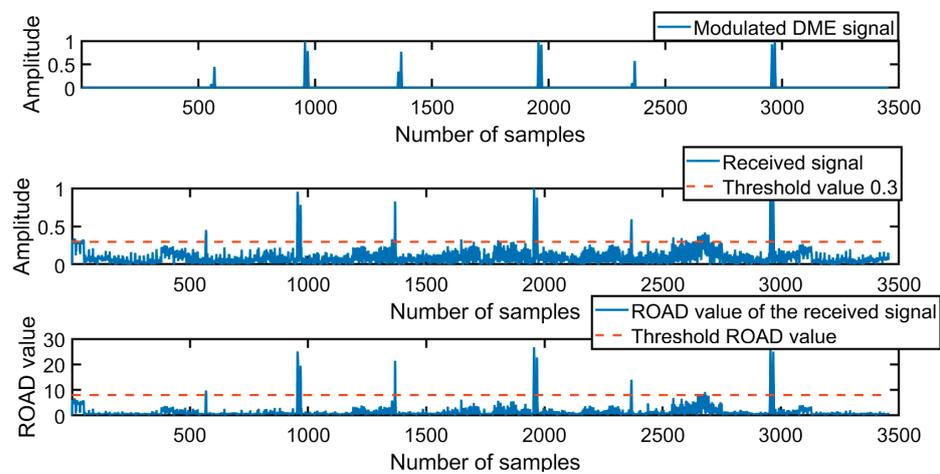


Figure 9. DME interference (**Top**), amplitude of the received signal without interference mitigation (**Middle**) and ROAD value of the received signal (**Bottom**) for an SNR of 15 dB.

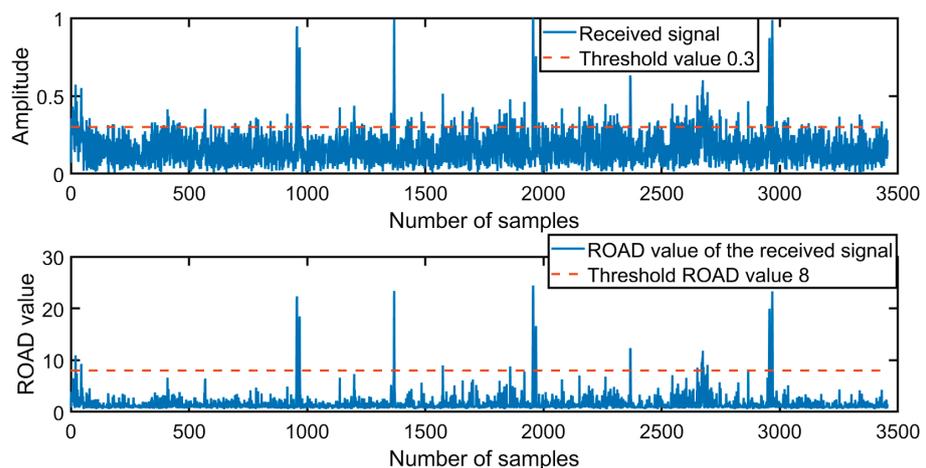


Figure 10. Amplitude of the received signal without interference mitigation and ROAD value of the received signal for an SNR of 0 dB.

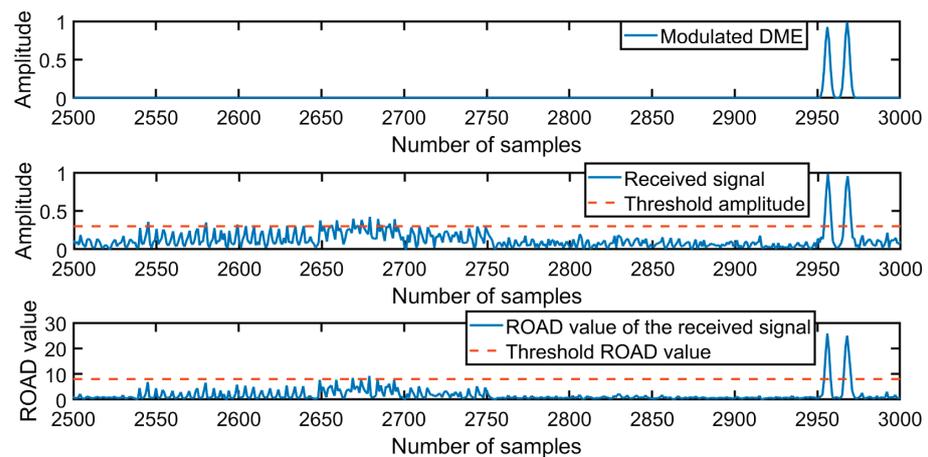


Figure 11. DME interference (**Top**), amplitude of the received signal without interference mitigation (**Middle**) and ROAD value of the received signal (**Bottom**) for an SNR of 15 dB.

An inclusive analysis of amplitude-based sensing and ROAD value-based sensing at low and high SNR levels has been portrayed in Figures 12 and 13, respectively. The characterization of different probability measures (accuracy, false detection and missed detection) are also analyzed in Figures 14–16. The findings from the comparative study of amplitude-based sensing and ROAD value-based sensing are as follows:

1. The number of false detections in amplitude-based sensing is more at a low SNR level of operation than high SNR values.
2. The number of false detections due to ROAD value-based sensing is always less (whether in high SNR level or in low SNR level) than amplitude-based sensing.
3. The performance of ROAD value-based sensing is almost the same in both low and high SNR levels.
4. ROAD value-based sensing shows some amount of missed detection as in between the samples 1000 and 1500 (Figure 13). The same sample is not lost in amplitude-based detection as it crosses the threshold value.
5. The probability of false detection that occurs in ROAD value-based sensing is observed to be always less than from amplitude-based sensing for all the SNR values under consideration. The probability of false detection is found to be decreased with an increase in SNR values due to the reduction in noise level (Figure 15).
6. The probability of missed detection is observed to increase with an increase in SNR levels in the case of amplitude-based sensing (Figure 16).
7. The probability of missed detection is found to be slightly decreasing with an increase in SNR values for ROAD value-based detection (Figure 16).
8. The probability of missed detection is found to be more for ROAD value-based detection than amplitude-based sensing.
9. Regardless of the elevated number of missed detections that occurred in ROAD value-based sensing, the method showed high accuracy in detection compared to conventional amplitude-based sensing (Figure 14).

The above mentioned (2, 3, 5 and 9) statements, which are realized from the obtained results, affirm that ROAD value-based sensing is admirable in comparison to amplitude value-based sensing in detecting DME interference.

Furthermore, the characterization of the ROAD statistic-based detection for different threshold ROAD values and SNRs has been performed. The results are as displayed in Figures 17–19. It has been observed that the probability of false detection decreases with an increase in threshold values (Figure 17) and the probability of missed detection increases with an increase in threshold value, Figure 18. Hence, there is a trade-off between false detection and missed detection for different values of threshold. Hence, to identify the optimum threshold, we have analyzed the variation in probability of correct detection

(accuracy) for different threshold values. From Figure 19, it is noted that the accuracy in ROAD value-based sensing increases from threshold ROAD value 5 to 8. The reason for this nature is the considerable reduction in false detection occurring in ROAD value-based sensing. Further, the performance starts diminishing from 8 to 12. This can be due to the increase in missed detection that occurs for high threshold value.

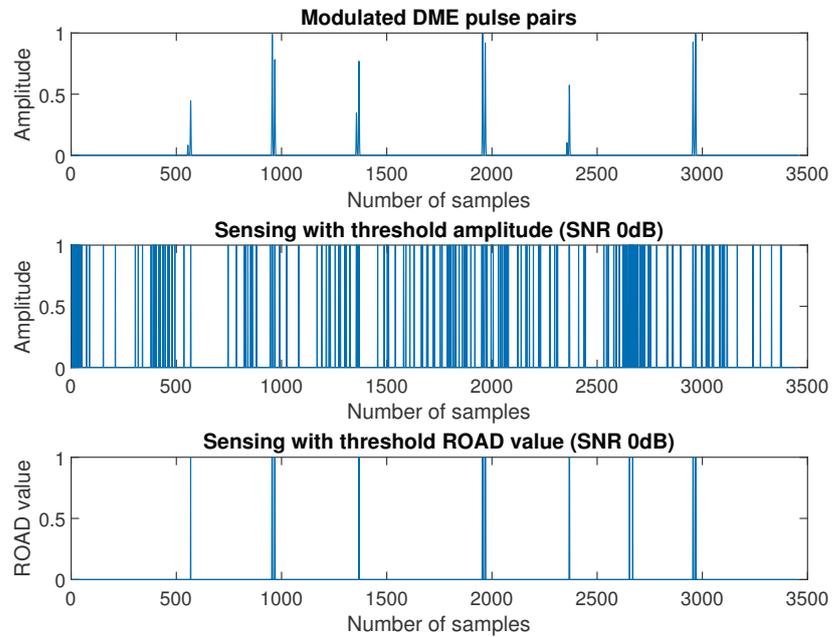


Figure 12. DME interference (**Top**), threshold amplitude-based sensing (**Middle**) and threshold ROAD value-based sensing (**Bottom**) for an SNR of 0 dB.

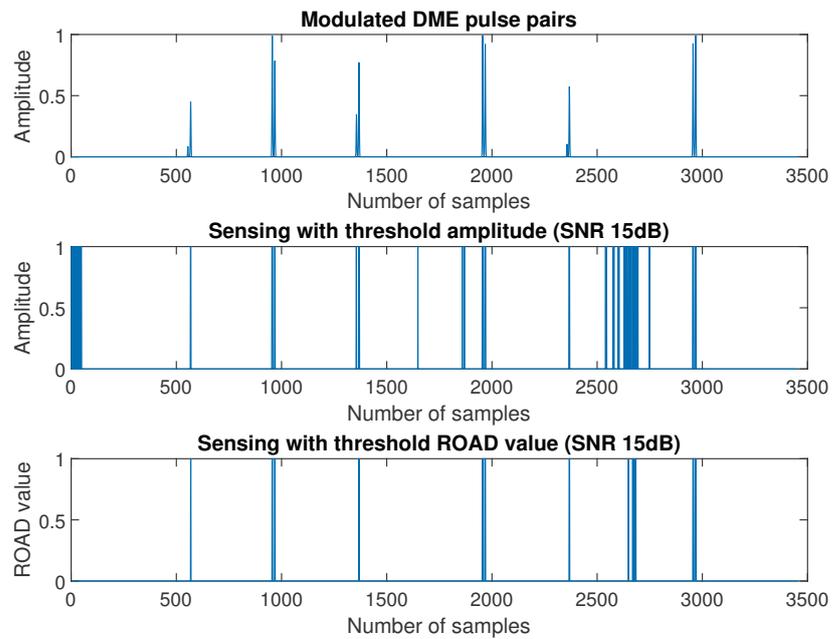


Figure 13. DME interference (**Top**), threshold amplitude-based sensing (**Middle**) and threshold ROAD value-based sensing (**Bottom**) for an SNR of 15 dB.

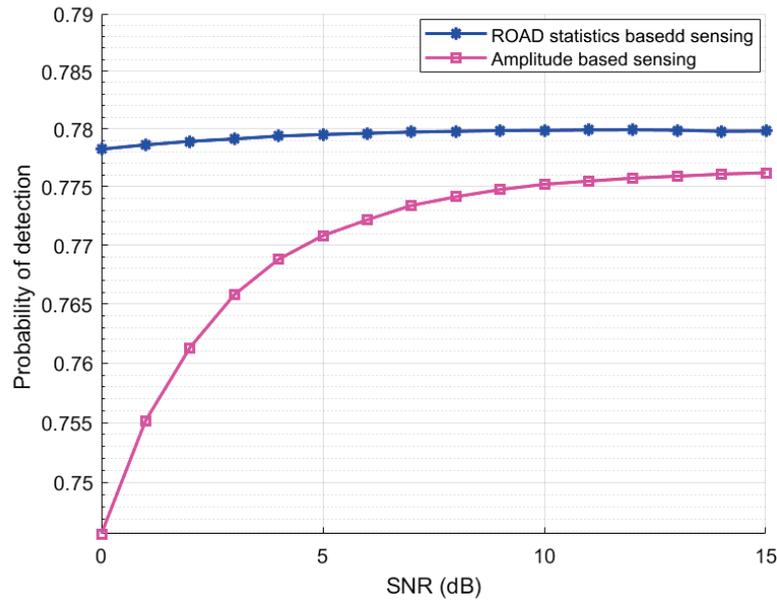


Figure 14. Probability of detection vs. SNR.

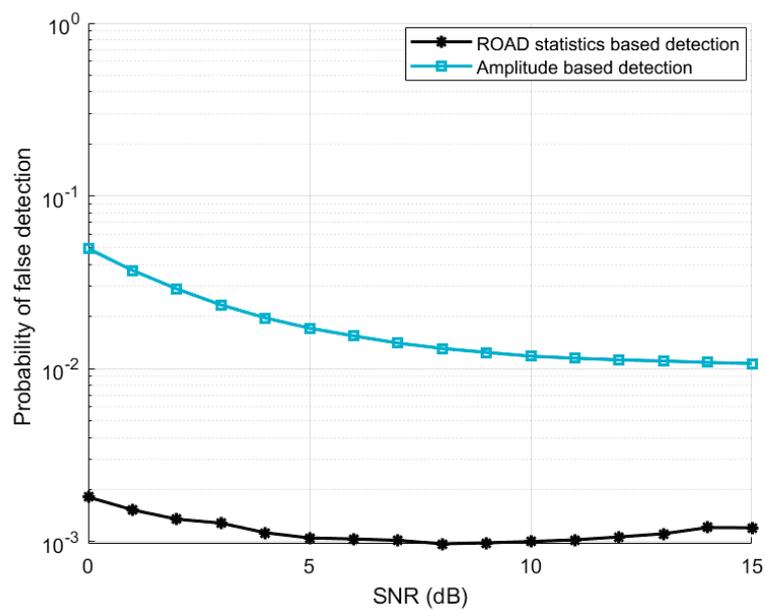


Figure 15. Probability of false detection vs. SNR.

For better clarity of the results, the variation in probability of false detection and missed detection has been plotted for a constant SNR value. The rate of decrease in false detection with the increase in threshold value is clearly visible in Figure 20. The variation in accuracy and missed detection are separately plotted to verify the optimum threshold value of detection and the reason behind it. From Figure 21, it is evident that optimum threshold value of detection is occurring for a value of 8. The reduction in detection after threshold value 8 is due to the increase in missed detection as in Figure 22.

Once the significance of ROAD value-based sensing and its optimum threshold ROAD value are identified (based on experimental results), the method is incorporated with an existing pulse blanking method to propose a new DME mitigation scheme.

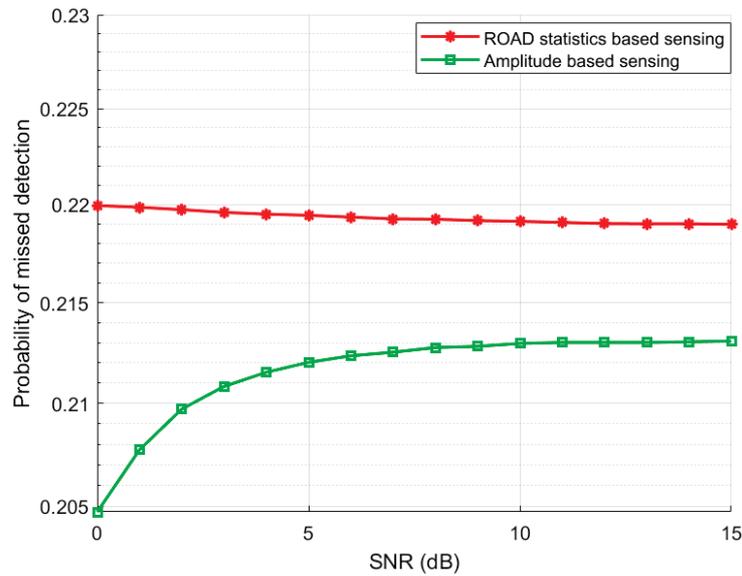


Figure 16. Probability of missed detection vs. SNR.

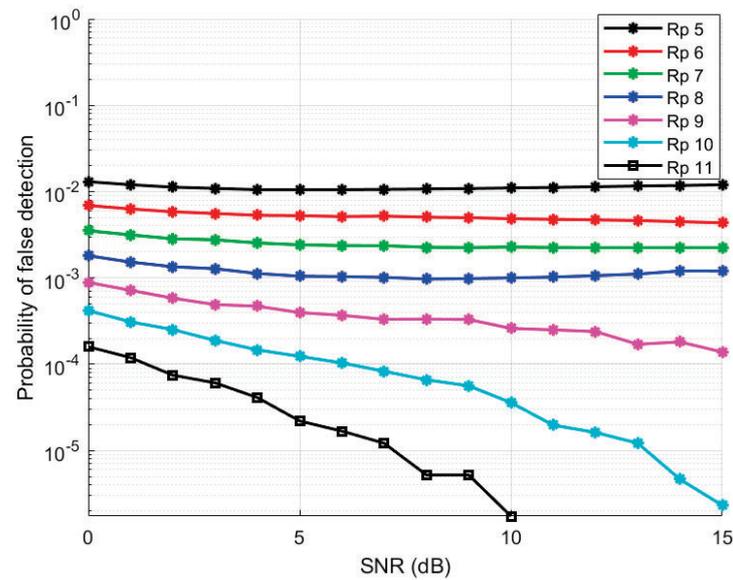


Figure 17. Probability of false detection in ROAD statistics-based detection for different threshold value vs. SNR.

The block diagram of the proposed LDACS receiver is shown in Figure 4. Initially, the cyclic prefix bits are removed from the received data. The resulting data are then converted into frequency domain using fast Fourier transform. Further, the frame decomposer separates the pilot symbols and complex data from the corresponding subcarriers. The segregated complex data values further undergo QPSK demodulation and symbol demapping to obtain the bitstreams. The bitstreams are de-interleaved and decoded using de-interleaver and vitterbi decoder, respectively. From the output of the vitterbi decoder, redundant bits are removed and decoded using the RS decoder to obtain the original data.

Figure 23 shows the variation in BER with different transmit SNR values. It has been observed that the proposed ROAD PB exhibits a much improved performance for a threshold value of 8. Figure 24 shows the variation in performance of ROAD PB for different threshold values ranging from 7 to 11. The performance of ROAD PB initially improved with a rise in threshold values and then started diminishing. The reason for this nature

is very obvious; ROAD detectors with low threshold values perceive the small variations from the neighboring carrier as DME interference. The actual data can cause a variation in ROAD values which can lead to false detection of DME interference. Moreover, once DME interference is detected in a subcarrier with a low threshold value, blanking the subcarrier causes the loss of more data. As OFDM systems have a self-removal noise mechanism due to the principle of orthogonality, the focus of detection is for large variation. Hence, there is an optimum high threshold value for which leaving the data is better than maintaining or estimating. From the results shown in Figure 24, the optimum threshold value is noted as 8. The accuracy of interference detection starts diminishing for a threshold value greater than an optimum threshold value. In this situation, the ROAD interference detector will only sense a very large variation from the neighboring carrier as DME interference leading to missed detection.

With the introduction of pulse blanking, a possibility of change in optimum threshold exists, if one considers the trade-off between interference power and signal distortion. Excess interference power may exist in mitigated data if the threshold value is high. On the other hand, reducing the threshold value can cause more distortion and loss of data due to blanking. In our work, the optimum threshold value of ROAD value-based sensing is recognized as the optimum value of operation to obtain data with minimum BER (Figures 21 and 24). As ROAD value-based noise detection is more accurate than the conventional amplitude-based method, it introduces less distortion in the mitigated data and prevents the extra loss of falsely detected data.

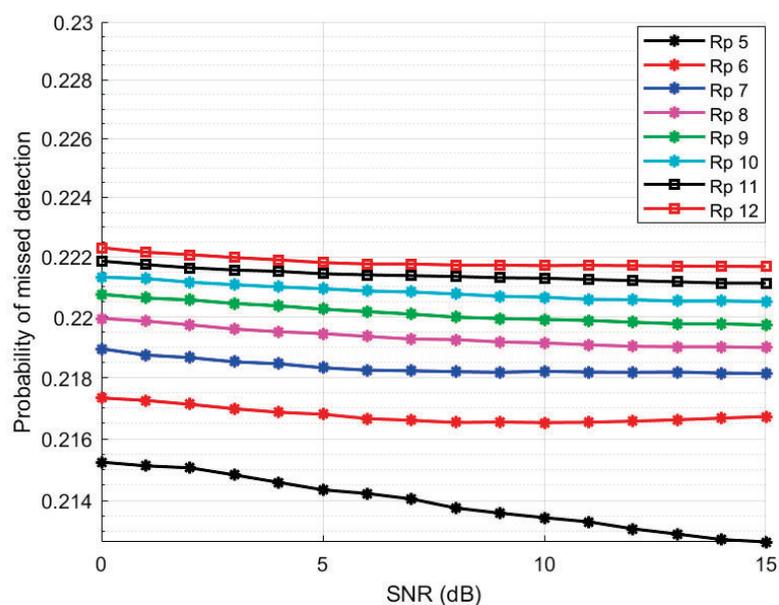


Figure 18. Probability of missed detection in ROAD statistics-based detection for different threshold value vs. SNR.

The performance of the proposed ROAD PB is further compared with GAE enhanced pulse peak processors and the results are as shown in Figure 25. The proposed ROAD PB outperformed the GAE enhanced pulse peak attenuators and limiters (Type I and II). Similarly, Figure 26 depicts the performance comparison of the proposed ROAD PB with OBE enhanced pulse peak processors. It has been observed that ROAD PB could outperform OBE PPA. Moreover, ROAD PB has similar or slightly better performance than OBE PPLs for low SNR values. For SNR values 8 dB and above, OBE enhanced pulse peak limiters performed better than ROAD PB. Figure 27 compares the performance of ROAD PB with pulse blanking, GAE PPL (Type 2) and OBE PPL (Type 2). It has been observed that GAE PPL and OBE PPL have an improved performance compared to pulse blanking as data estimation has been performed instead of blanking the noise affected

signal. The threshold value used for all types of pulse peak processors is 0.3 [30]. When ROAD statistics is incorporated with pulse blanking, the performance could be improved better than GAE PPL and comparable performance with OBE PPL.

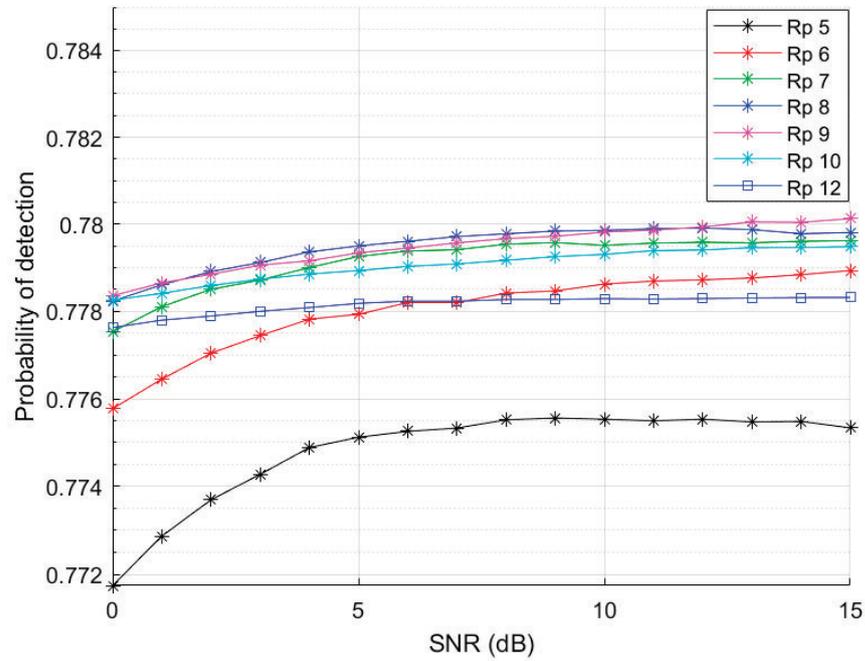


Figure 19. Accuracy in ROAD statistics-based detection vs. SNR.

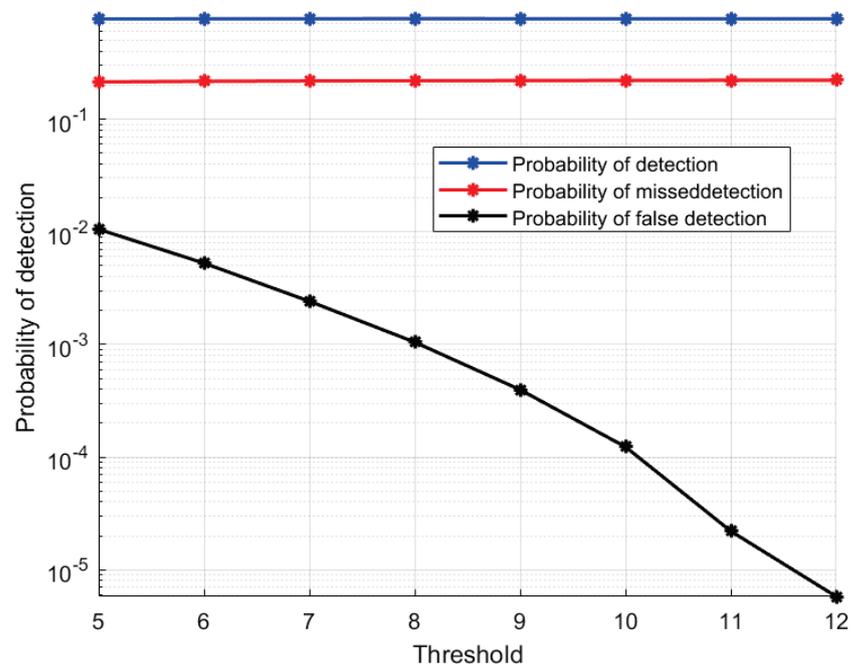


Figure 20. Probability of detection vs. threshold for SNR = 5 dB.

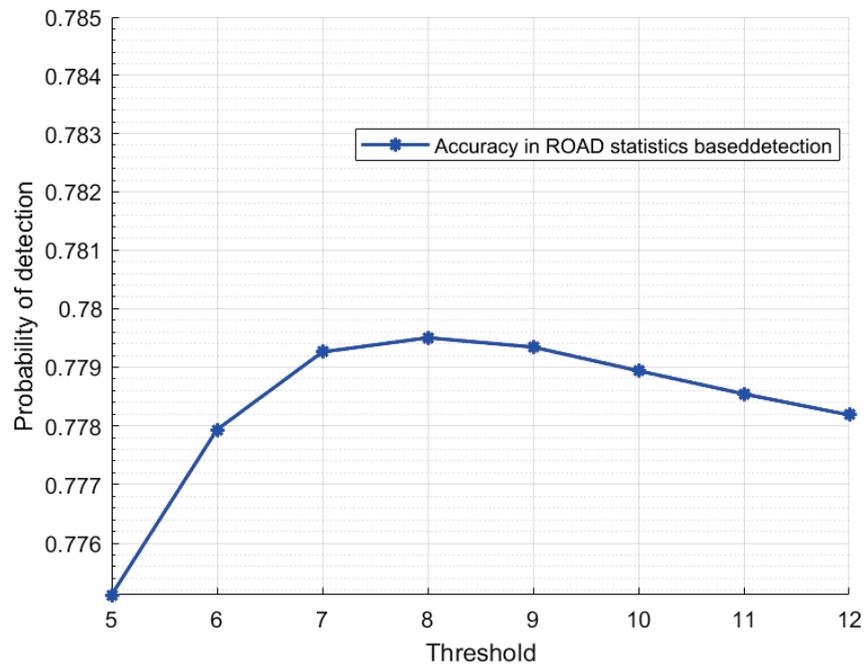


Figure 21. Accuracy in ROAD statistics-based detection (SNR = 5 dB) vs. threshold.

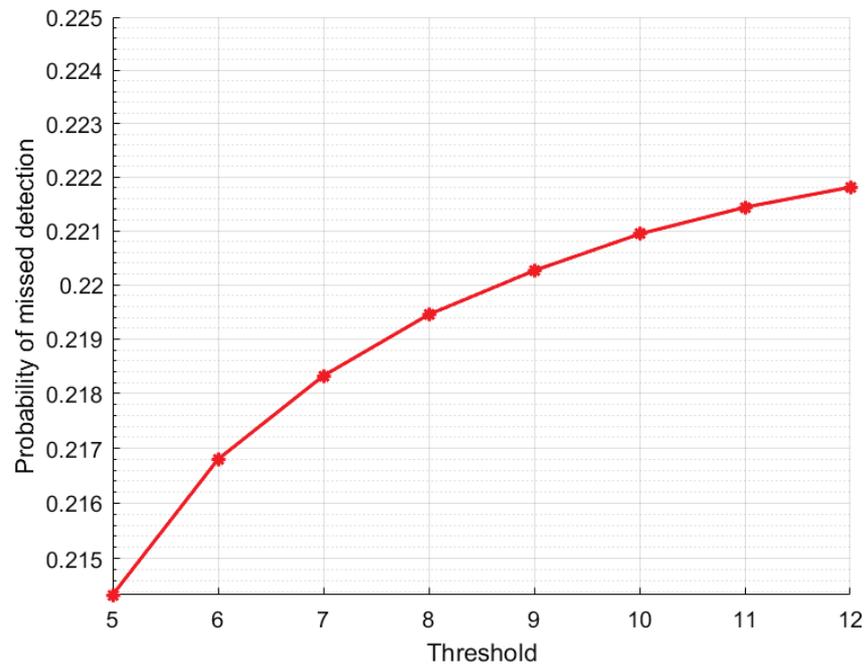


Figure 22. Probability of missed detection in ROAD statistics-based sensing (SNR = 5 dB) vs. threshold.

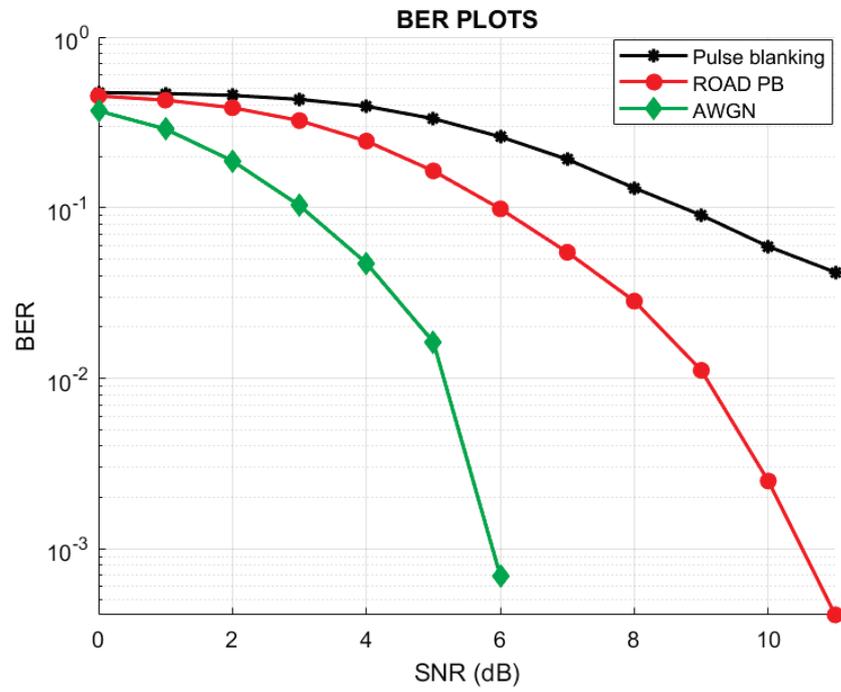


Figure 23. Performance of ROAD PB vs. pulse blanking.

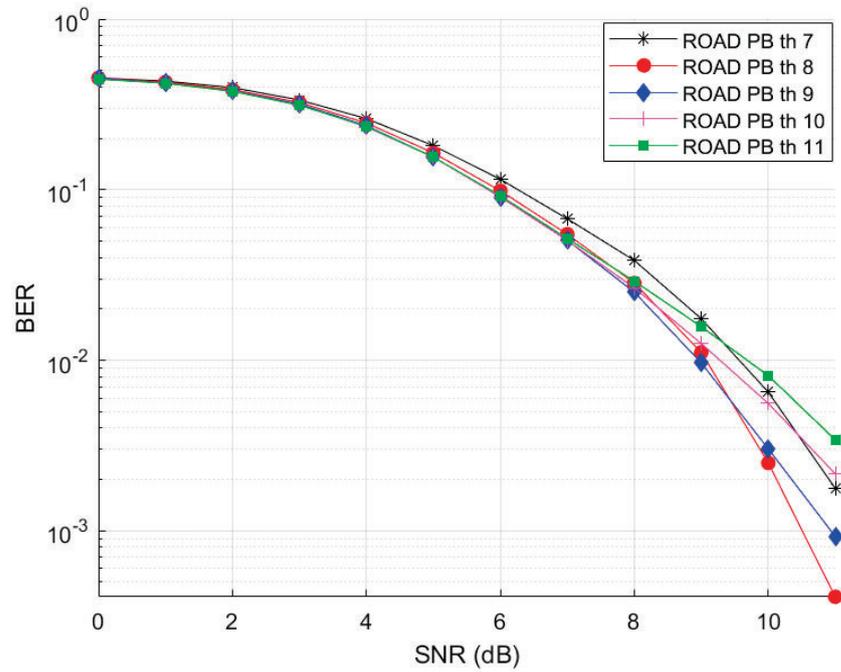


Figure 24. Variation of ROAD PB with different threshold value.

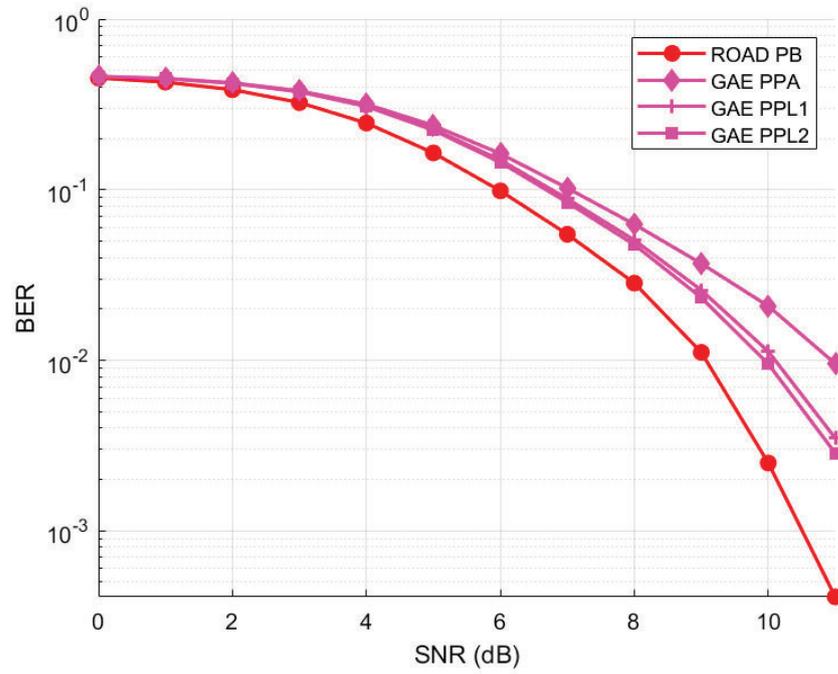


Figure 25. Performance of ROADPB vs. GAE enhanced pulse peak processors.

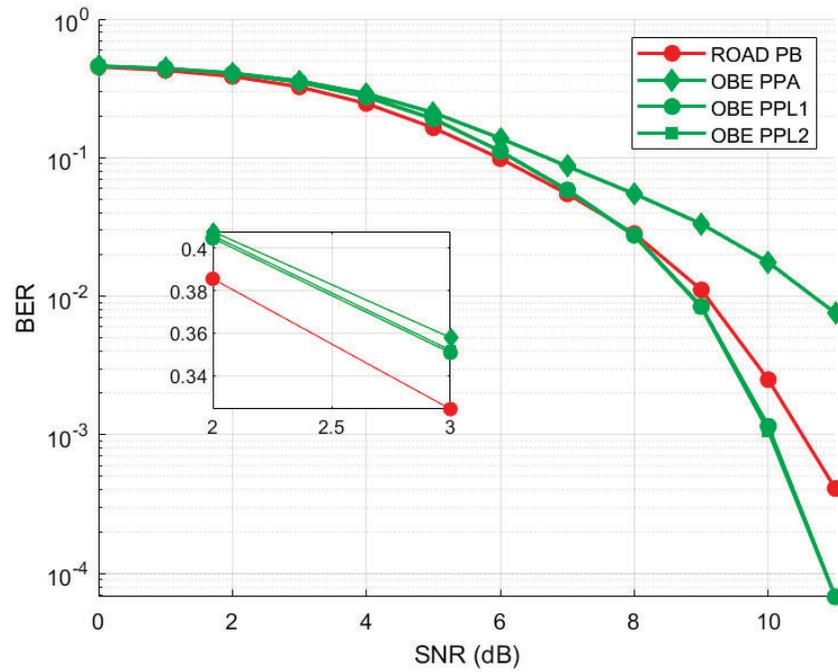


Figure 26. Performance of ROADPB vs. OBE enhanced pulse peak processors.

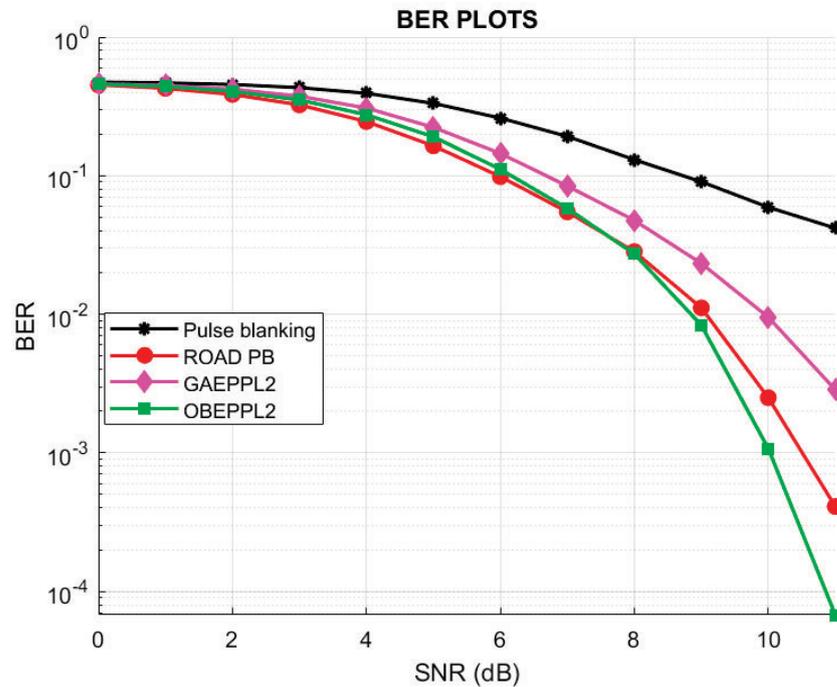


Figure 27. Performance of ROADPB vs. pulse blanking, GAE and OBE enhanced pulse processors.

Table 1. OFDM Parameters for LDACS1 [9].

OFDM Parameters	Values
Effective RF BW (FL or RL)	498.05 KHz
FFT size N_{FFT}	64
Sampling time T_{sa}	1.6 μ s
Subcarrier spacing f	9.765625 KHz
Used subcarriers N_u	50
Useful symbol time N_u	102.4 μ s
Cyclic prefix time T_{cp}	17.6 μ s
Total OFDM symbol time T_s	120 μ s
Guard time T_g	4.8 μ s
Windowing time T_w	12.2 μ s
Number of lower frequency guard subcarriers $N_{g,left}$	7
Number of higher frequency guard subcarriers	
Lower frequency guard subcarriers $N_{g,right}$	6
Total FFT BW B_g	625 KHz

6. Conclusions

In this paper, a new DME mitigation scheme named ROAD PB is proposed to mitigate the DME interference using pulse blanking. The ROAD PB detects DME interference with a method named ROAD statistics-based detection. The method detects the interference from the ROAD values of the received signal. The performance of the new detection method is compared with the conventional amplitude-based sensing method. The results guided to the following conclusions:

1. ROAD statistics detection method outperformed conventional amplitude-based sensing in identifying the location of the DME interference (Figure 11).
2. The probability of detection of ROAD value-based sensing remains the same for both low and high SNR values, whereas the same found be varying for amplitude-based sensing. The accuracy of detection of the proposed method always excelled over conventional amplitude-based sensing (Figure 14).

3. Though ROAD value-based sensing showed an increase in missed detection in comparison to amplitude-based sensing, the method could always achieve enlarged accuracy due to the substantial decrease in false detection. For an SNR value of 0 dB, regardless of the 2% increase in probability of missed detection compared to amplitude-based sensing, the proposed method achieved a 3% increase in probability of detection (accuracy) with the help of a 4% reduction in the probability of false detection (Figures 14–16).
4. The ROAD statistic-based interference sensing always shows a considerable reduction in false detection of the DME signal.
5. The optimum threshold value of detection is observed to be 8. The reduction in performance for a threshold value lower than 8 is due to the presence of more numbers of false detection than threshold 8. The reduction in performance for a threshold value higher than 8 is due to the increase in missed detection at higher threshold values.

From the results obtained with the comparative study of the proposed method (ROADPB) with amplitude-based detection methods such as pulse blanking and pulse peak processors, we observed the following:

1. ROAD PB exhibited improved performance than the pulse blanking method which uses the amplitude of the received signal for the detection of DME interference.
2. ROAD PB always outperformed the three types of GAE enhanced pulse peak processors in its optimum threshold value of operation.
3. ROAD PB showed better performance than OBE enhanced pulse peak attenuator and comparable performance with OBE enhanced pulse peak limiters.

In comparison to pulse blanking, ROAD PB could achieve the SNR saving of 2.7 dB at a BER of 10^{-1} by introducing some amount of complexity in the receiver. Moreover, at a BER of 10^{-1} , ROAD PB could accomplish SNR savings of 2.7, 1.1, 0.7, 0.25 and 0.2 dBs compared to GAEPPA, GAEPLL, OBEPPA and OBEPLL, respectively. The proposed ROAD PB is significant due to its improved performance at low SNR regions in comparison to pulse blanking. Moreover, ROAD value-based detection can be used to sense impulse noise in any type of OFDM-based communication systems where threshold-based detection can be used. In the future, ROAD value-based detection can be incorporated with any other threshold-based DME mitigation scheme such as GAE enhanced methods. ROAD PB-based LDACS receiver can be extended for the en-route channel. The performance of this method in LDACS RL can also be analyzed. Though this method is investigated on the LDACS background, the method is compatible in cutting down impulse noise in any OFDM-based communication systems.

Author Contributions: Conceptualization, M.K. and R.M.; methodology, M.K. and R.M.; software, M.K.; validation, A.R. and R.M.; formal analysis, M.K.; writing—original draft preparation, writing—review and editing, M.K.; supervision, R.M. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Exclude this statement.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

GAE	Genie-aided estimator
OBE	Optimal Bayesian estimator
LDACS	L-band Digital Aeronautic Communication Systems
DME	Distance measuring equipment

PPA	Pulse peak attenuator
PPL	Pulse peak limiter
GAEPPA	Genie-aided estimator enhanced pulse peak attenuator
GAEPPL	Genie-aided estimator enhanced pulse peak limiter
OBEPPA	Optimal Bayesian estimator enhanced pulse peak attenuator
OBEPPL	Optimal Bayesian estimator enhanced pulse peak limiter
WAIC	Wireless avionic intra communication
UAV	Unmanned aerial vehicle
A/G	Air-ground communications
DSB-AM	Double sideband amplitude modulation
ATM	Air traffic management
TBO	Trajectory-based operations
ICAO	International civil aviation organization
GSM	Global system for mobile communication
TACAN	Tactical air navigation
JTIDS	Joint tactical information distribution system
UAT	Universal access transceiver
SSR	secondary surveillance radar
ACAS	airborne collision avoidance system
FL	Forward link
GS	Ground station
RL	Reverse link
AS	Air station
FDD	Frequency division duplexing
CCK	Complementary code keying
DWT	Discrete wavelet transform
PLC	Power-line communication
ADSL	Asymmetric digital subscriber lines
DVBT	Digital video broadcasting—terrestrial
QPSK	Quadrature Phase Shift Keying
AWGN	Additive white Gaussian noise
K-GMM	K-component Gaussian mixture mode
F	LDACS forward link frame
F_t	t th symbol of LDACS forward link frame (F)
σ_F^2	variance of F_t
N	Number of orthogonal subcarriers
S_t	The OFDM symbol
N_{CP}	number of cyclic prefix bits
X'_t	Transmitted vector
X'	Transmitted signal
r'	Received signal
i_t	Noise
A_t	AWGN noise
p_t	Impulse noise
r'_t	Received signal at an instant ' t '
r_t	Received signal after CP removal ' t '
$P_d(t)$	A pair of DME pulses
$I_{pd}(f)$	Modulated DME spectrum
$I_{pd}(t)$	DME interfering signal to LDACS system
t	Output signal at each time epoch
\hat{x}_t	Estimated output signal at each time epoch
$\sigma^2 X$	Signal power of transmitted signal X'
$\sigma^2 i$	Noise power added to transmitted signal
$\sigma^2 X_t$	Transmitted signal power at an instant
$\sigma^2 i_t$	Noise power at an instant
P_k	Probability of occurrence of each Gaussian component k
Γ	The ratio of thermal noise to impulse noise
pI	Probability of impulse noise

$\sigma^2 I$	Impulse noise power
$\sigma^2 0$	Thermal noise power
$\hat{x}_{ kPA }(y)$	Output of GAEPPA
$\hat{x}_{ kPL }(y)$	Output of GAEPPL Type 1
$\hat{x}_{ kPLs }(y)$	Output of GAEPPL Type 2
ρ_k	Attenuation factor for GAEPPA
ρ_{mod}	Attenuation factor for GAEPPL Type 1
$\hat{x}_{ kOPA }(y)$	Output of OBEPPA
$\hat{x}_{ kOPL }(y)$	Output of OBEPPL (Types 1 and 2)
$\beta_o(y)$	Attenuation factor for OBEPPA and JOBEPPA
$\beta_{mod}(y)$	Attenuation factor for OBEPPPL
μ	General scaling factor

Appendix A

Following, we provide the derivation of extra computational complexity introduced by ROAD PB, Section 4.1.

Appendix A.1

The extra computational complexity put forward by ROAD PB compared to conventional pulse blanking is the sum of the computational complexity put forward by the steps to calculate ROAD value. The following steps are used to calculate ROAD value of a center sample in one OFDM symbol:

1. The received OFDM symbol for each time epoch is considered as a one-dimensional vector. For a one-dimensional vector of size $(1 \times N)$, the absolute difference between a center sample and a receiving sample for each time epoch $f_d(k)$ are calculated as in (A1)

$$f_d(k) = |r_k - [r_{k-\text{frac}N-1/2}, \dots, r_{k+\text{frac}(N-1)/2}]| \tag{A1}$$

No complex multiplications are introduced in this step.

Here, the number of complex additions put forward is $N - 1$ for center sample or for one subcarrier. It is the same as that of $(2 \cdot N - 1)$ real multiplications.

Thus, for N number of subcarriers, the number of complex additions involved in the calculation of absolute difference from a center sample C_a^d is as follows,

$$C_a^d = N \cdot (N - 1). \tag{A2}$$

It is the same as $2N \cdot (2 \cdot N - 1)$ real additions and can be expressed as in (A3)

$$R_a^d = 2N \cdot (2N - 1). \tag{A3}$$

2. The difference vector $(f_d(k))$ is sorted in increasing order.

$$Q(k) = \text{sort}(f_d(k)). \tag{A4}$$

No complex multiplications or additions are introduced in this step.

For a one-dimensional vector of size N , the number of real-time additions put forward by sorting depends on the type of sorting that is used. For instance, if selective sorting is used, it introduces $\frac{N \cdot (N-1)}{2}$ number of real-time additions for a single center sample. Thus, for N number of samples, the number of real-time additions involved in selective sorting is as in (A5)

$$R_a^s = \frac{N^2 \cdot (N - 1)}{2}. \tag{A5}$$

3. The ROAD value is calculated as the sum of first $(N - 1)/2$ values of $Q(k)$

$$ROAD = \sum_{k=1}^{(N-1)/2} Q(k). \quad (A6)$$

No complex multiplications or additions are introduced in this step.

Finally, the number of real-time additions put forward by adding the first half values of the sorted output is $\frac{(N-1)}{2}$.

For N number of samples, the number of real-time additions involved is as in (A7)

$$R_a^a = \frac{N \cdot (N - 1)}{2}. \quad (A7)$$

References

1. EUROCONTROL. EUROCONTROL's Challenges of Growth 2018 Study Report. Available online: <https://www.eurocontrol.int/sites/default/files/2019-07/challenges-of-growth-2018-annex1-0.pdf> (accessed on 29 September 2019).
2. Available online: <https://www.ldacs.com/wp-content/uploads/2013/12/ACP-DCIWG-IP01-LDACS-White-Paper.pdf> (accessed on 11 November 2021).
3. L-DACS1 System Definition Proposal: Deliverable D2. 2009. Available online: <https://www.ldacs.com/wp-content/uploads/2014/02/LDACS1-Specification-Proposal-D2-Deliverable.pdf> (accessed on 11 November 2021).
4. L-DACS2 System Definition Proposal: Deliverable D2, (EUROCONTROL, Brussels, Belgium, 2009) Tech. Rep. 1.0, Version 1.0. Available online: <https://www.eurocontrol.int/sites/default/files/2019-05/11052009-ldacs2-d2-deliverable-v1.0.pdf> (accessed on 11 November 2021).
5. Epple, U.; Schnell, M. Mitigation of impulsive interference for LDACS1—Potentials of blanking nonlinearity. In Proceedings of the 2015 Integrated Communication, Navigation and Surveillance Conference (ICNS), Herdon, VA, USA, 21–23 April 2015; pp. 1–16.
6. Mostafa, M.; Schnell, M. DME-compliant LDACS1 cell planning: Initial steps. In Proceedings of the 2015 Integrated Communication, Navigation and Surveillance Conference (ICNS), Herdon, VA, USA, 21–23 April 2015; pp. 1–42.
7. Schnell, M.; Epple, U.; Shutin, D.; Schneckenburger, N.T. LDACS: Future aeronautical communications for air-traffic management. *IEEE Commun. Mag.* **2014**, *52*, 104–110. [\[CrossRef\]](#)
8. Thiasiriphet, T.; Schneckenburger, N.; Schnell, M. Impact of the DME interference on the LDACS1 ranging performance. In Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2015), Tampa, FL, USA, 18 September 2015; pp. 1446–1467.
9. Muthalagu, R.T. Mitigation of DME interference in LDACS1-based future air-to-ground (A/G) communications. *Cogent Eng.* **2018**, *5*, 1472199. [\[CrossRef\]](#)
10. Agrawal, N.; Darak, S.J. Performance analysis of reconfigurable filtered OFDM for LDACS. In Proceedings of the 2019 11th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, India, 7 January 2019; pp. 500–503.
11. Agrawal, N.; Darak, S.J.; Bader, F. Spectral Coexistence of LDACS and DME: Analysis via Hardware software co-design in presence of real channels and RF impairments. *IEEE Trans. Veh. Technol.* **2020**, *69*, 9837–9848. [\[CrossRef\]](#)
12. Agrawal, N.; Darak, S.J.; Bader, F. New Spectrum Efficient Reconfigurable Filtered-OFDM Based L-Band Digital Aeronautical Communication System. *IEEE Trans. Aerosp. Electron. Syst.* **2019**, *55*, 1108–1122. [\[CrossRef\]](#)
13. Mathew, L.K.; Vinod, A.P. An energy-difference detection based spectrum sensing technique for improving the spectral efficiency of LDACS1 in aeronautical communications. In Proceedings of the 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016; pp. 1–5.
14. Mathew, L.K.; Vinod, A.P.; Madhukumar, A.S. A cyclic prefix assisted spectrum sensing method for aeronautical communication systems. In Proceedings of the 2019 IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 26–29 May 2019; pp. 1–5.
15. Raja, M. Application of cognitive radio and interference cancellation in the L-Band based on future air-to-ground communication systems. *Digit. Commun. Netw.* **2019**, *5*, 111–120. [\[CrossRef\]](#)
16. Epple, U.; Schnell, M. Overview of interference situation and mitigation techniques for LDACS1. In Proceedings of the 2011 IEEE/AIAA 30th Digital Avionics Systems Conference, Seattle, WA, USA, 16–20 October 2011; p. 4C5-1.
17. Wang, H.; Liu, R. DME/TACAN interference mitigated by mixed blanking in B2 band. In Proceedings of the 2011 IEEE 13th International Conference on Communication Technology, Jinan, China, 25–28 September 2011; pp. 897–900.
18. Bai, Y.; Zhang, X.; Zhao, L. Co-channel CCK transmission overlapped with DME in aeronautical communication. In Proceedings of the 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP), Xi'an, China, 9–13 July 2014; pp. 403–406.
19. Li, Q.; Zhang, J.; Xie, J.; Mostafa, M.; Schnell, M. Iterative interference mitigation and channel estimation for LDACS1. In Proceedings of the 2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC), Colorado Springs, CO, USA, 5–9 October 2014; p. 3B2-1.

20. Hirschbeck, M.; Huber, J.B. Block based frequency selective mitigation of pulsed interference in the L-band. In Proceedings of the 10th International ITG Conference on Systems, Communications and Coding, Hamburg, Germany, 2–5 February 2015; pp. 1–6.
21. Li, D.; Wu, Z.T. DME Interference mitigation for L-DACS1 based on system identification and sparse representation. *Chin. J. Aeronaut.* **2016**, *6*, 1762–1773. [[CrossRef](#)]
22. Saaifan, K.A.; Henkel, W. Lattice signal sets to combat pulsed interference from aeronautical signals. In Proceedings of the 2011 IEEE International Conference on Communications (ICC), Kyoto, Japan, 5–9 June 2011; pp. 1–5.
23. Abd-Elaty, E.; El-Agoz, S.; Zekry, A.E. Compact Spread Spectrum LDACS Wavelet Based for DME Interference Mitigation. In Proceedings of the 2019 36th National Radio Science Conference (NRSC), Port Said, Egypt, 16–18 April 2019; pp. 259–265.
24. Abd-Elaty, E.; Zekry, A.; El-Agoz, S.; Helaly, A.M. Nested code division multiple access for distance measuring equipment interference mitigation in L-band digital aeronautical communication system. *Int. J. Commun. Syst.* **2021**, *34*, e4804. [[CrossRef](#)]
25. Muthalagu, M.K.R.; Mathew, L.K.; Rajak, A. Deep Clipping Based Interference Mitigation Technique for LDACS. In Proceedings of the 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 3–7 October 2021; pp. 1–6. [[CrossRef](#)]
26. Eriksson, H.B.; Ödling, P.; Koski, T.; Börjesson, P.O. A genie-aided detector with a probabilistic description of the side information. In Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, BC, Canada, 17–22 September 1995; p. 332.
27. Rožić, N.; Banelli, P.; Begušić, D.; Radić, J.T. Multiple-Threshold Estimators for Impulsive Noise Suppression in Multicarrier Communications. *IEEE Trans. Signal Process.* **2018**, *10*, 1619–1633. [[CrossRef](#)]
28. Banelli, P. Bayesian estimation of a Gaussian source in Middleton's class—A impulsive noise. *IEEE Signal Process. Lett.* **2013**, *20*, 956–959. [[CrossRef](#)]
29. Banelli, P.; Rugini, L. Impulsive noise mitigation for wireless OFDM. In Proceedings of the 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Stockholm, Sweden, 28 June–1 July 2015; pp. 346–350.
30. Keshkar, M.; Muthalagu, R.; Rajak, A.; Mathew, L.K. GAE and OBE Enhanced Interference Mitigation Techniques in LDACS. *Aerospace* **2022**, *9*, 45. [[CrossRef](#)]
31. Garnett, R.; Huegerich, T.; Chui, C.; He, W. A universal noise removal algorithm with an impulse detector. *IEEE Trans. Image Process.* **2005**, *14*, 1747–1754. [[CrossRef](#)] [[PubMed](#)]
32. Barazideh, R.; Niknam, S.; Natarajan, B. Impulsive Noise Detection in OFDM-based Systems: A Deep Learning Perspective. In Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 7–9 January 2019; pp. 0937–0942. [[CrossRef](#)]
33. Gao, G.X. DME/TACAN Interference and its Mitigation in L5/E5 Bands. In Proceedings of the 20th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2007), Fort Worth, TX, USA, 28 September 2007; pp. 1191–1200.
34. Epple, U.; Hoffmann, F.; Schnell, M. Modeling DME interference impact on LDACS1. In Proceedings of the 2012 Integrated Communications, Navigation and Surveillance Conference, G7-1-G7-13, Herndon, VA, USA, 24–26 April 2012.
35. Brandes, S.; Schnell, M. Interference Mitigation for the Future Aeronautical Communication System in the L-Band. In *Multi-Carrier Systems & Solutions*; Plass, S., Dammann, A., Kaiser, S., Fazel, K., Eds.; Lecture Notes in Electrical Engineering; Springer: Dordrecht, The Netherlands, 2009; Volume 41.
36. Oppenheim, A.V.; Schaffer, R.W. *Discrete-Time Signal Processing*; Pearson Education: London, UK, 2014.
37. Kassam, S.A. *Signal Detection in Non-Gaussian Noise*; Springer Science & Business Media: New York, NY, USA, 1988.
38. Barazideh, R.; Natarajan, B.; Nikitin, A.V.; Niknam, S.T. Performance Analysis of Analog Intermittently Nonlinear Filter in the Presence of Impulsive Noise. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3565–3573. [[CrossRef](#)]
39. Blackard, K.L.; Rappaport, T.S.; Bostian, C.W.T. Measurements and models of radio frequency impulsive noise for indoor wireless communications. *IEEE J. Sel. Areas Commun.* **1993**, *10*, 991–1001. [[CrossRef](#)]
40. Sanchez, M.G.; De Haro, L.; Ramón, M.C.; Mansilla, A.; Ortega, C.M.; Oliver, D.T. Impulsive noise measurements and characterization in a UHF digital TV channel. *IEEE Trans. Electromagn. Compat.* **1999**, *41*, 124–136. [[CrossRef](#)]
41. Nakai, T.; Kawasaki, Z.T.T. On impulsive noise from Shinkansen. *IEEE Trans. Electromagn. Compat.* **1983**, *4*, 396–404. [[CrossRef](#)]
42. Dempster, A.P.; Laird, N.M.; Rubin, D.B.T. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Methodological* **1977**, *39*, 1–22.
43. Al-Naffouri, T.Y.; Quadeer, A.A.; Caire, G.T. Impulse noise estimation and removal for OFDM systems. *IEEE Trans. Commun.* **2014**, *62*, 976–989. [[CrossRef](#)]

Article

IMU: A Content Replacement Policy for CCN, Based on Immature Content Selection

Salman Rashid *, Shukor Abd Razak * and Fuad A. Ghaleb *

School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

* Correspondence: rashid.salman@graduate.utm.my (S.R.); shukorar@utm.my (S.A.R.);
abdulgaleel@utm.my (F.A.G.)

Abstract: In-network caching is the essential part of Content-Centric Networking (CCN). The main aim of a CCN caching module is data distribution within the network. Each CCN node can cache content according to its placement policy. Therefore, it is fully equipped to meet the requirements of future networks demands. The placement strategy decides to cache the content at the optimized location and minimize content redundancy within the network. When cache capacity is full, the content eviction policy decides which content should stay in the cache and which content should be evicted. Hence, network performance and cache hit ratio almost equally depend on the content placement and replacement policies. Content eviction policies have diverse requirements due to limited cache capacity, higher request rates, and the rapid change of cache states. Many replacement policies follow the concept of low or high popularity and data freshness for content eviction. However, when content loses its popularity after becoming very popular in a certain period, it remains in the cache space. Moreover, content is evicted from the cache space before it becomes popular. To handle the above-mentioned issue, we introduced the concept of maturity/immaturity of the content. The proposed policy, named Immature Used (IMU), finds the content maturity index by using the content arrival time and its frequency within a specific time frame. Also, it determines the maturity level through a maturity classifier. In the case of a full cache, the least immature content is evicted from the cache space. We performed extensive simulations in the simulator (Icarus) to evaluate the performance (cache hit ratio, path stretch, latency, and link load) of the proposed policy with different well-known cache replacement policies in CCN. The obtained results, with varying popularity and cache sizes, indicate that our proposed policy can achieve up to 14.31% more cache hits, 5.91% reduced latency, 3.82% improved path stretch, and 9.53% decreased link load, compared to the recently proposed technique. Moreover, the proposed policy performed significantly better compared to other baseline approaches.

Citation: Rashid, S.; Razak, S.A.; Ghaleb, F.A. IMU: A Content Replacement Policy for CCN, Based on Immature Content Selection. *Appl. Sci.* **2022**, *12*, 344. <https://doi.org/10.3390/app12010344>

Academic Editor: Yangquan Chen

Received: 18 November 2021

Accepted: 27 December 2021

Published: 30 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: content replacement; content placement; content-centric networking; cache networks; immaturity; stretch reduction

1. Introduction

Due to advancements in technology, things are becoming more integrated and intelligent, leading to a rapid increase in Internet usability. Internet usage patterns demonstrate that new era applications are becoming more sensitive in bandwidth and latency. IP video traffic is expected to dominate overall IP traffic by 82% by 2022 [1], up from 74% in 2017 [2]. Internet users are not interested in the location of the storage server. Their primary interest is having Internet connectivity that assures fast and reliable retrieval of desired information. Content-centric networking (CCN) has proven to be a promising solution to meet the needs of future networks [3]. CCN naturally supports in-network caching and it attempts to respond to the requested data when a user request contains the name or identity of the desired data. CCN assigns each piece of data a unique identity and addresses data objects at the network level, in contrast to the Internet's host-centric architecture. CCN

naturally supports in-network caching and many-to-many communication [4]. When the user request contains the name or identification of the desired data object, the network attempts to respond to the request with the data object. The name can also belong to a location or a host machine. This mechanism makes CCN more general than the host-to-host communication model [5].

In-network caching provides a solution for traditional Internet architecture that works in the application layer [3]. The content of the CCN cache changes rapidly due to enormous data demands. Furthermore, CCN is a solution that works on the network layer level [6]. It allows the CCN node to cache a temporary copy of the requested content. CCN can minimize network traffic and maximize content availability by providing the desired content closer to the consumer [7]; it is difficult to decide the cache location of the content to satisfy consumer requests and improve network performance [8]. In addition, it is also important to determine which content should be removed from the cache space to accommodate new content in the cache. Improper content selection causes the degradation of network performance [9]. In-network caching faces several challenges, including limited cache storage, caching replacement and placement, caching traffic, and complex network topology [10,11].

The performance of the CCN depends on the content placement and replacement policies. The content placement policy decides the appropriate cache location of each content [1]. Hence, the node selection for content caching should be optimized to satisfy consumer requests, with minimum overhead. Due to the limited cache capacity in the node, any cached content in the cache needs to be removed to accommodate new content [12]. Content replacement policy is responsible for choosing the right content against defined criteria [13]. The network performance and cache hit ratio decreases if popular content is removed from the cache or unpopular content remains in cache for a long time [14–17].

Although cached contents at all nodes, along with the routing path, increase the network performance and cache hit ratio, it is not a practical approach due to the finite cache space. That is, if the cache space is full and new content arrives, one of the cached content needs to be removed from the cache space. However, most existing replacement policies follow the concept of the Least Frequently Used (LFU) or Least Recently Used (LRU) policy to replace the content which is not effective for CNN [18]. The newly arrived content may become popular over time due to high demand. When popular content loses its popularity, it stays in the cache due to previous popularity. Therefore, the network performance may decrease due to the overstay of previous popular content that is currently unpopular or the eviction of currently popular content. To solve this issue and improve network performance, we introduced a new concept of content maturity and immaturity to deal with the aforementioned issues. The content that loses its popularity over a specific time frame and stays in the cache for a long time is called immature content. In contrast, the content will be considered mature if it has high popularity and is also recently requested in the network within a specific time frame. Every new content is neither popular nor mature. Content should stay in the cache for some time to know its maturity level. Hence, such content is not evicted from the cache, which is yet to become popular. In addition, this concept removes content from the cache that loses its popularity after being highly popular for some time.

A content replacement policy is proposed in this work called IMU (Immaturity used). This policy removes content from the cache that is immature within a limited time frame. Therefore, most of the contents in the cache are recently used and highly popular, leading to a better cache hit ratio and network performance. The key contributions are summarized below:

- A new concept of content maturity/immaturity has been introduced to design and develop an effective content eviction policy. The proposed content eviction policy evicts the content from cache through the immature content selection to improve the cache hit ratio, latency, path stretch, and link load.

- A mechanism to calculate the maturity level of the content has been designed and developed using the content frequency and arrival time of the content.
- Icarus [19] has been used to verify the performance of the proposed policy with existing state-of-the-art content replacement policies. Subsequently, we gained substantial improvement in cache hit ratio, latency, path stretch, and link load.

The rest of this paper is organized as follows: We discuss the related work in Section 2. The proposed policy is described in Section 3, which highlights its contribution. Section 4 describes the simulation environment and parameters as well as the result analysis and discussion. Finally, the conclusion and future work are in Section 5.

2. Related Work

Content eviction policy works when the cache space is filled with content. The eviction policy provides a mechanism to replace existing contents with requested contents in the cache. The eviction policy must keep popular contents in the cache with the least processing complexity. In general, an eviction policy should have two properties. First, the eviction policy should not remove popular content from the cache. Second, it also keeps the most frequently used contents in the cache by applying some sort of priority. Several eviction policies were proposed in the past [9,19–25]. Some of the most popular eviction policies include First in First out (FIFO), Random Replacement (RR), Least Recently Used (LRU), Least Frequently Used (LFU), Window-LFU (W-LFU), Least Frequent Recently Used (LFRU), Popularity Prediction Caching (PPC), Network-oriented Information-centric Centrality for Efficiency (NICE), NC-SDN, and Least Fresh First (LFF). A brief description of each cache eviction policy is summarized below.

As the name suggests, FIFO replaces the content from the cache based on a first-come, first-serve basis. The content item that comes first in the cache is evicted first when there is a need for replacement [20]. It does not deal with the importance or priority of the content being replaced by the new content. RR policy randomly selects existing content from the cache to replace it with new content [21]. However, it has no particular criteria for content selection from the cache. LRU is a typical policy that has extensive usage in cache eviction [22]. LRU keeps track of the usage of each content in cache. When the replacement request is received, LRU checks the requested content in the cache. If this requested content is not already in the cache, it evicts the least recently used content to accommodate the requested content. Therefore, LRU is simple to implement and has less computational delay. But on the other side, LRU does not consider the content frequency (dynamic changes of popularity over time), which plays a significant role in network performance and the cache hit ratio.

LFU keeps track of the frequency of each content in the cache [23]. LFU serves to store the most popular content in the cache statically. It keeps a counter of how many times the content is requested. Whenever a request is received for content, the counter value is incremented by one. When the cache space is full and there is a requirement to replace content, the content with the least counter value is selected to evict. LFU keeps popular content in the cache, but it requires a very high processing time that leads to performance degradation in CCN. Further, when content that has been popular for some time loses its popularity, it stays in the cache, causing severe performance losses. W-LFU is an eviction policy that uses a limited number of access requests over a time window [24]. This technique tries to solve the LFU problem by keeping the history of the requested contents. This record of history is referred to as a window. The size of this window is directly proportional to the total number of contents and the cache size in the network. This policy demonstrates considerable improvements, but it fails to evict suitable content in the case of bursty requests. Moreover, this policy only observes a small portion of the cache, making it impractical for full cache capacity.

LFRU is the combination of LRU and LFU [25]. According to the LFRU eviction policy, a cache divides into unprivileged and privileged partitions. The privileged partition is known as a protected partition. The popular content is pushed into the privileged partition.

If the privileged partition is fully occupied and there is no more space available to store content, the LFRU ensures that the content is evicted from the unprivileged partition and that content is transferred from the privileged to the unprivileged partition. Filtering out the locally popular contents and placing the popular contents in the privileged partition are the key features of the LFRU eviction policy. This policy demonstrates considerable improvements, but it fails to evict suitable content in the case of bursty requests. Moreover, this technique also requires a large processing time to manage partitions.

PPC is a chunk-level in-network caching eviction policy [26]. It is capable of predicting the popularity of the video chunks. PPC stores content based on the popularity that it predicts. On the other side, the contents that have the least popular prediction are evicted out. This eviction scheme is also termed the Assist-predict model. It is based on the request information of the neighboring chunks. It also predicts future popularity by using past experience with the popularity of the content. If the popularity of the new content is less than the former popularity, the newly incoming chunk does not cache. Otherwise, it evicts the future content based on popular prediction. This model-based prediction technique works well but fails to predict properly against frequently changing requests. Moreover, this policy leads to high network load due to control signaling overhead and high computational workload. The NC-SDN eviction model was introduced as a cache eviction algorithm that relies on SDN (software-defined networking) [16]. NC-SDN model uses three arguments. First, it calculates data popularity; second, it comes to know the location of cache management switches; third, it facilitates cooperation among different nodes in the network. When the cache is fully occupied, it checks the popularity of each content and replaces the least popular content with new content. Although the replacement technique is straightforward, the control traffic and exchange of information between the switches are very high, leading to performance losses.

LFF is a content replacement policy that predicts the time of the next event [27]. Based on the prediction, it controls the residual life of retrieved content. When the cache capacity is full, this policy measures the time for which the content is considered invalid. In addition, this policy checks whether the source has been updated after retrieving the content to check the validity of each content. This policy ignores the high replacement rate in the central node and does excessive computing, making it impractical for large CCN. NICE has been introduced as a new metric for cache management in ICN [28]. This policy uses a method that computes the centrality. Centralization is used in the replacement phase to manage cache contents. This method is based on the number of caches instead of the number of contents. Content is replaced when the NICE value is high, as the contents move from one cache to the other due to the centrality of the content. However, it causes high network load and computational complexity.

Most of the replacement strategies [27–34] on CCN focus on content frequency, popularity, and time freshness. These policies ignore the concept of content immaturity in content eviction; it is neither popular nor mature when new content is cached in the cache. We need some time to evaluate whether this content has become popular or not. If that content is removed from the cache, the consumer has to retrieve it from the publisher, which affects network performance. Therefore, content may become popular for a certain period, and then its popularity starts decreasing [29]; if that content is not removed from the cache, network performance and the cache hit ratio also degrade. When the cache space is low and the popularity of the content changes frequently, it becomes challenging for the content eviction policy to decide which content should be evicted from the cache space. A content eviction policy should be able to provide equal opportunities for each content to become mature. Therefore, we introduced a concept of maturity and immaturity of the content, and our proposed cache replacement policy uses this concept to accommodate the request of new content. The proposed policy evicts immature content to solve content popularity issues.

3. Proposed Content Replacement Policy

Content replacement policy is an integral part of CCN cache management. The nodes in CCN need to free up space over time, due to limited cache space, so that new contents are cached. It is a crucial decision to evict content from the cache, which, in turn, increases or decreases network performance and the cache hit ratio. Numerous content replacement policies decide to evict content from the cache using various criteria, such as time in the cache, frequency, popularity, and node centralization. These policies do not use content immaturity for eviction. The proposed policy selects immature content from the cache that stays for a long time in the cache and has a lower frequency in a particular time window. Thus, the proposed policy avoids unnecessary content occupation in the cache space. Due to immature content eviction, network nodes contain more requested content within the cache space. Therefore, more customer interests are satisfied within the network.

The proposed technique determines the mature/immature contents. Algorithm 1 elaborates the procedure to label a content, s_i is mature or immature. The proposed policy keeps track of each content's arrival time and frequency at each node. The current time and the frequency of the node s_i is denoted by T_{cf,s_i} and $F_{\check{c},s_i}$ respectively. The proposed technique calculates the content period, T_{p,s_i} , with the help of content frequency, $F_{\check{c},s_i}$, and content arrival time, T_{cf,s_i} . Therefore, it determines the duration of the content s_i in the cache space. Then, the proposed policy calculates the maturity index $M_{c\check{I},s_i}$ by dividing the frequency of the content $F_{\check{c},s_i}$ and content period T_{p,s_i} . The maturity classifier M_L is calculated using the median of maturity indexes $M_L \leftarrow Median(M_{c\check{I},s_n})$. Content s_i whose maturity index $M_{c\check{I},s_i}$ exceeds the value of M_L is classified as mature content; otherwise, it is immature content. The median is used for finding the relevant mean value of the maturity index $M_{c\check{I},s_i}$, because it is not affected by lower or extreme high set values. Thus, this provides a fair value to the maturity classifier M_L .

Algorithm 1: Determine the mature and immature content.

Input: Suppose $S \in \{s_1, s_2, s_3, \dots, s_n\}$ is set of contents.

Output: Categorization of contents.

T_{cf,s_i} is the arrival time of i th content.

$F_{\check{c},s_i}$ is the frequency of i th content.

W_T is the size of the time window.

T_{p,s_i} is the time period of i th content.

$M_{c\check{I},s_i}$ is the maturity index of i th content.

M_L is the maturity classifier.

1. for $i = 1 : n$
 - $T_{p,s_i} \leftarrow W_T - T_{cf,s_i}$
 - $M_{c\check{I},s_i} \leftarrow F_{\check{c},s_i} / T_{p,s_i}$
 2. $M_L \leftarrow Median(M_{c\check{I},s_n})$
 3. for $i = 1 : n$
 - if $M_{c\check{I},s_i} \geq M_L$
 s_i is mature
 - else
 s_i is immature
-

Algorithm 2 describes the next part of our proposed policy. When a node v receives an interest packet for content s_i , and the time window has not expired, then the proposed policy finds the requested content s_i in the local cache. In the case of a cache hit, the proposed policy increments the frequency of content s_i by one and associates a new arrival time T_{cf,s_i} . Moreover, node v discards the interest packet from PIT and replies through the

data packet to the requested consumer. Otherwise, a cache miss means that the requested content s_i is being cached for the first time in CS. Thus, its frequency $F_{\check{c},s_i}$ is one and it is associated with the current timestamp T_{cf,s_i} . When the cache is full, it selects content s_k with a minimum value of the maturity index $M_{c\check{i},s_i}$ and evicts it from the cache space. Then, the proposed technique checks the time window W_T ; if W_T is expired, then the frequency of all content $F_{\check{c},s_n}$ is set to one, and the previously associated timestamp T_{cf,s_n} remains the same.

Algorithm 2: IMU Replacement Policy.

Input: Request for a content s_i at node v
Output: Content selection for replacement of newly arrived content

1. if W_T is not expired
 - check local cache
 - if cache hit
 - $F_{\check{c},s_i} \leftarrow F_{\check{c},s_i} + 1$
 - $T_{cf,s_i} \leftarrow$ current time
 - else if cahe_size == full
 - $s_k \leftarrow$ select the content with min $M_{c\check{i},s_i}$
 - evict s_k
 - place s_i in cache
 - $F_{\check{c},s_i} \leftarrow 1$
 - $T_{cf,s_i} \leftarrow$ current time
 - else
 - place s_i in cache
 - $F_{\check{c},s_i} \leftarrow 1$
 - $T_{cf,s_i} \leftarrow$ current time
2. else
 - for each s_i
 - $F_{\check{c},s_i} \leftarrow 1$
 - Update W_T
 - go to step 1

For simplicity, we assume that all the CCN based routers (node) have the same cache sizes, cached content, and discrete instants of time for interests to arrive. CS is the local cache size, and the window size is denoted by W_T . There are some events related to content s_i , including received interest packet, received data packet, reply data packet, forward interest packet, cached content, eviction from the cache, and look-up content in local CS. The received interest packet (RIP), received data packet (RDP), reply data packet (REDP), forward interest packet (FIP), cached content (CC), eviction from the cache (EC), and look-up content (LU) are denoted by $\check{R}_{s_{ip}}, \check{R}_{s_{dp}}, \check{R}_{s_{edp}}, \check{F}_{s_{ip}}, \check{C}_s, \check{E}_s$ and \check{L}_s , respectively. These notations are helpful to understand the whole process of the proposed policy. For example, initially, we assumed the value of cache space (CS) = 6, $W_T = 4$ s, $t = [1, 2, 3, 4, \dots, 13]$, $S \in \{A, B, C, D, E, F, G, H, I\}$, as presented in Figure 1.

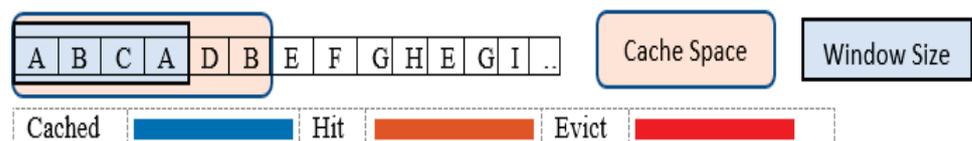


Figure 1. Node cache space, window size, and caching events.

The consumer’s requested content (RC) is in the same sequence, window size, and cache space illustrated in Figure 1. The colors indicate three caching processes: cached, hit, and evicted content from the cache. With the help of these colors, we can easily understand the new entry in our tables and variations in the values.

We assume that the cache of a node is empty. The detailed caching process at $t = 1$ to $t = 4$ is expressed in Table 1, and Table 2 also maps the IMU process with values. Moreover, we see the effect of values after the window W_T is not expired. The router receives an interest packet for content A $\check{R}_{A_{ip}}$ but does not find that content in its CS after look-up \check{L}_A , then routers the same interest packet forwards $\check{F}_{A_{ip}}$ to the next router. The next router has A content and responds through the data packet $\check{R}_{A_{edp}}$, from which it receives the packet of interest. Furthermore, the router received A data packet $\check{R}_{A_{dp}}$ is then cached \check{C}_A in CS, along with values $T_{cf,s_A} = 1, F_{\check{c},s_A} = 1, T_{p,s_A} = 4,$ and $M_{\check{c},s_A} = 0.25$. This process will be the same for content B and C. When a hit occurs at $t = 4$, then the values of $T_{cf,s_A} = 4, F_{\check{c},s_A} = 2, T_{p,s_A} = 1,$ and $M_{\check{c},s_A} = 2$. Furthermore, window W_T has expired at the same $t = 4$, but the cache space is not yet full. Then, the router receives an interest packet for content A $\check{R}_{A_{ip}}$. The hit occurred at $t = 4$ and changed the values of values $T_{cf,s_A} = 4, F_{\check{c},s_A} = 2, T_{p,s_A} = 1,$ and $M_{\check{c},s_A} = 2.00$.

Table 1. Caching process at $t = 1$ to $t = 4$.

Time t	RC s_i	RIP $\check{R}_{s_{ip}}$	LU \check{L}_s	FIP $\check{F}_{s_{ip}}$	REDP $\check{R}_{s_{edp}}$	RDP $\check{R}_{s_{dp}}$	CC \check{C}_s	EC \check{E}_s
1	A	$\check{R}_{A_{ip}}$ (1)	\check{L}_A (2)	$\check{F}_{A_{ip}}$ (3)	$\check{R}_{A_{edp}}$ (4)	$\check{R}_{A_{dp}}$ (5)	\check{C}_A (6)	
2	B	$\check{R}_{B_{ip}}$ (1)	\check{L}_B (2)	$\check{F}_{B_{ip}}$ (3)	$\check{R}_{B_{edp}}$ (4)	$\check{R}_{B_{dp}}$ (5)	\check{C}_B (6)	
3	C	$\check{R}_{C_{ip}}$ (1)	\check{L}_C (2)	$\check{F}_{C_{ip}}$ (3)	$\check{R}_{C_{edp}}$ (4)	$\check{R}_{C_{dp}}$ (5)	\check{C}_C (6)	
4	A	$\check{R}_{A_{ip}}$ (1)	\check{L}_A (2)		$\check{R}_{A_{edp}}$ (3)			

Table 2. IMU process at $t = 1$ to $t = 4$.

Time	s_i	T_{cf,s_i}	$F_{\check{c},s_i}$	T_{p,s_i}	$M_{\check{c},s_i}$
t = 1	A	1	1	4	0.25
t = 2	A	1	1	4	0.25
	B	2	1	3	0.33
t = 3	A	1	1	4	0.25
	B	2	1	3	0.33
	C	3	1	2	0.50
t = 4	A	4	2	1	2.00
	B	2	1	3	0.33
	C	3	1	2	0.50

Table 3 describes the detailed caching process at $t = 5$ to $t = 8$, and Table 4 also maps the IMU process with values. Table 4 demonstrates that the new window W_T starts at $t = 5$, and that all the values of $F_{\check{c},s_i}$ become 1 and retain all the values of T_{cf,s_i} . Content D has cached \check{C}_D at $t = 5$ and displays the values of $F_{\check{c},s_D} = 1, T_{cf,s_D} = 5, T_{p,s_D} = 4,$ and $M_{\check{c},s_D} = 0.25$ in Table 4. Furthermore, at this stage, cache space $CS = 4$. After the hit occurs

at $t = 6$, the values of $T_{cf,s_D} = 6$, $F_{\check{c},s_D} = 2$, $T_{p,s_D} = 3$, and $M_{c\check{i},s_D} = 0.67$ have changed. New contents are cached at $t = 7$ and $t = 8$, \hat{C}_E and \hat{C}_F , respectively.

Table 3. Caching process at $t = 5$ to $t = 8$.

Time t	RC s_i	RIP $\check{R}_{s_{ip}}$	LU \check{L}_s	FIP $\check{F}_{s_{ip}}$	REDP $\check{R}_{s_{edp}}$	RDP $\check{R}_{s_{dp}}$	CC \check{C}_s	EC \check{E}_s
5	D	$\check{R}_{D_{ip}}$ (1)	\check{L}_D (2)	$\check{F}_{D_{ip}}$ (3)	$\check{R}_{D_{edp}}$ (4)	$\check{R}_{D_{dp}}$ (5)	\check{C}_D (6)	
6	B	$\check{R}_{B_{ip}}$ (1)	\check{L}_B (2)		$\check{R}_{B_{edp}}$ (3)			
7	E	$\check{R}_{E_{ip}}$ (1)	\check{L}_E (2)	$\check{F}_{E_{ip}}$ (3)	$\check{R}_{E_{edp}}$ (4)	$\check{R}_{E_{dp}}$ (5)	\hat{C}_E (6)	
8	F	$\check{R}_{F_{ip}}$ (1)	\check{L}_F (2)	$\check{F}_{F_{ip}}$ (3)	$\check{R}_{F_{edp}}$ (4)	$\check{R}_{F_{dp}}$ (5)	\hat{C}_F (6)	

Table 4. IMU process at $t = 5$ to $t = 8$.

Time	s_i	T_{cf,s_i}	$F_{\check{c},s_i}$	T_{p,s_i}	$M_{c\check{i},s_i}$	Time	s_i	T_{cf,s_i}	$F_{\check{c},s_i}$	T_{p,s_i}	$M_{c\check{i},s_i}$
t = 5	A	4	1	5	0.20	t = 7	B	6	2	3	0.67
	B	2	1	7	0.14		A	4	1	5	0.20
	C	3	1	6	0.17		C	3	1	6	0.17
	D	5	1	4	0.25		D	5	1	4	0.25
t = 6	B	6	2	3	0.67	t = 8	E	7	1	2	0.50
	A	4	1	5	0.20		B	6	2	3	0.67
	C	3	1	6	0.17		A	4	1	5	0.20
	D	5	1	4	0.25		C	3	1	6	0.17
						D	5	1	4	0.25	
						E	7	1	2	0.50	
						F	8	1	1	1.00	

Content E and F have cached \hat{C}_E and \hat{C}_F , respectively, at $t = 6$ and $t = 7$, and the new values are presented in Table 4. We see that the caching process is displayed step by step in Table 3, and the numbers are associated with each process to illustrate the sequence of this process.

Table 5 reflects caching events from $t = 9$ to $t = 12$. At $t = 9$, the router receives an interest packet of G $\check{R}_{G_{ip}}$. After the look-up \check{L}_G content is not found in CS, the interest packet is forwarded $\check{F}_{G_{ip}}$ to the next router. This time, CS is full when it receives the $\check{R}_{G_{dp}}$ data packet. Now, we find the lowest $M_{c\check{i},s_C} = 0.10$ value and remove that content \check{E}_C from CS. Therefore, it caches the new content \hat{C}_G with the associated values of $T_{cf,s_G} = 9$, $F_{\check{c},s_G} = 1$, $T_{p,s_G} = 4$, and $M_{c\check{i},s_G} = 0.25$ in the CS. We can observe in Table 6 how the IMU works when the memory is full and new content arrives simultaneously.

We repeat the same process at $t = 10$ for content H. The hit occurs at $t = 11$, and $t = 12$ updates the values of T_{cf,s_E} , $F_{\check{c},s_E}$, T_{p,s_E} , and $M_{c\check{i},s_E}$ as illustrated in Table 6. The hit occurred at $t = 11$ and $t = 12$ for requested contents E and G, respectively. Table 5 illustrates that the minimum caching process and forwarding operations have been minimized when the hit occurs.

Table 5. Caching process at t = 9 to t = 12.

Time t	RC s_i	RIP $\check{R}_{s_{ip}}$	LU \check{L}_s	FIP $\check{F}_{s_{ip}}$	REDP $\check{R}_{s_{edp}}$	RDP $\check{R}_{s_{dp}}$	CC \check{C}_s	EC \check{E}_s
9	G	$\check{R}_{G_{ip}}$ (1)	\check{L}_G (2)	$\check{F}_{G_{ip}}$ (3)	$\check{R}_{G_{edp}}$ (4)	$\check{R}_{G_{dp}}$ (5)	\check{C}_G (7)	\check{E}_C (6)
10	H	$\check{R}_{H_{ip}}$ (1)	\check{L}_H (2)	$\check{F}_{H_{ip}}$ (3)	$\check{R}_{H_{edp}}$ (4)	$\check{R}_{H_{dp}}$ (5)	\check{C}_H (7)	\check{E}_A (6)
11	E	$\check{R}_{E_{ip}}$ (1)	\check{L}_E (2)		$\check{R}_{E_{edp}}$ (3)			
12	G	$\check{R}_{G_{ip}}$ (1)	\check{L}_G (2)		$\check{R}_{G_{edp}}$ (3)			

Table 6. IMU process at t = 9 to t = 12.

Time	s_i	\mathcal{T}_{cf,s_i}	$\mathcal{F}_{\check{c},s_i}$	\mathcal{T}_{p,s_i}	$M_{\check{c}\check{l},s_i}$	Time	s_i	\mathcal{T}_{cf,s_i}	$\mathcal{F}_{\check{c},s_i}$	\mathcal{T}_{p,s_i}	$M_{\check{c}\check{l},s_i}$
t = 9	B	6	1	7	0.14	t = 11	E	11	2	2	1.00
	A	4	1	9	0.11		B	6	1	7	0.14
	C	3	1	10	0.10		D	5	1	8	0.13
	D	5	1	8	0.13		F	8	1	5	0.20
	E	7	1	6	0.17		G	9	1	4	0.25
	F	8	1	5	0.20		H	10	1	3	0.33
	G	9	1	4	0.25		G	12	2	1	2.00
t = 10	B	6	1	7	0.14	t = 12	E	11	2	2	1.00
	A	4	1	9	0.11		B	6	1	7	0.14
	D	5	1	8	0.13		D	5	1	8	0.13
	E	7	1	6	0.17		F	8	1	5	0.20
	F	8	1	5	0.20		H	10	1	3	0.33
	G	9	1	4	0.25						
	H	10	1	3	0.33						

Table 7 describes the detailed process of caching at t = 13. The cache space CS is full, and the time window W_T has expired; Table 8 demonstrates that when the new time window starts, all the values of $\mathcal{F}_{\check{c},s_i}$ become one (1) and retain the values of \mathcal{T}_{cf,s_i} . The exact process that was performed at t = 9 and t = 10 is repeated at t = 13. The IMU used the \mathcal{T}_{cf,s_i} and $\mathcal{F}_{\check{c},s_i}$ for calculating the maturity index $M_{\check{c}\check{l},s_i}$ of the content s_i . This value indicates the maturity of the content with the specific time window W_T .

The tables demonstrate that the lower value of a content maturity index $M_{\check{c}\check{l},s_i}$ represents a longer stay in the cache space, with a lower frequency (popularity) over a particular time frame W_T . Therefore, this content is evicted from the cache when the cache space is full. It takes some time to define the maturity/immaturity of new cached content. Therefore, the content should not be evicted without checking the level of a content maturity index; the tables indicate that the maturity index value of new cached content is greater than others. Content that has become popular over time, but loses its popularity, has a higher frequency than other content. Therefore, this kind of content stays in CS for a long time and wastes cache space. However, the window W_T is used to equalize the frequency of all contents after a specific time, and immature content is selected from the maturity index $M_{\check{c}\check{l},s_i}$ to evict content from the CS. The proposed policy has significantly improved the cache hit ratio, bandwidth usage, latency, and path stretch.

Table 7. Caching process at t = 13.

Time t	RC s_i	RIP $\check{R}_{s_{ip}}$	LU \check{L}_s	FIP $\check{F}_{s_{ip}}$	REDP $\check{R}_{s_{edp}}$	RDP $\check{R}_{s_{dp}}$	CC \check{C}_s	EC \check{E}_s
13	I	$\check{R}_{I_{ip}}$ (1)	\check{L}_I (2)	$\check{F}_{I_{ip}}$ (3)	$\check{R}_{I_{edp}}$ (4)	$\check{R}_{I_{dp}}$ (5)	\check{C}_I (7)	\check{E}_D (6)

Table 8. IMU process at t = 13.

Time	s_i	T_{cf,s_i}	$F_{\check{c},s_i}$	T_{p,s_i}	M_{cl,s_i}	Time	s_i	T_{cf,s_i}	$F_{\check{c},s_i}$	T_{p,s_i}	M_{cl,s_i}
t = 13	G	12	1	5	0.20		F	8	1	9	0.11
	E	11	1	6	0.17		H	10	1	7	0.14
	B	6	1	11	0.10		I	13	1	4	0.25
	D	5	1	12	0.08						

4. Performance Evaluation

We performed a simulation in the GEANT network topology using the Icarus [13] simulator, to evaluate the performance of our policy. The GEANT topology consists of 40 nodes and 60 edges. The cache capacity of each node in the network is the same and ranges between 4% to 20% of the total content population. We used warm-up requests to settle caches before running the actual experiment, to minimize experimental errors. The cache warm-up requests are 40,000 and measured requests are also 40,000. We also used measured requests for performance evaluation. Zipf’s law is used to distribute the popularity of the content and popularity distribution of the exponent alpha ($\alpha \in [0.6, 0.8, 1.0]$) used in our simulation. For fair comparison with state-of-the-art replacement policies, the popularity of requested contents follows a Zipf distribution with a parameter ranging from 0.6 to 1.0, as presented in [10]. The lower and higher values indicate a low and high correlation between content requests [30]. The parameters of our simulation setup are mentioned in Table 9.

Table 9. Simulation Parameters.

Parameters	Value
Warm-up Requests	40,000
Measured Requests	40,000
Model of Popularity	0.6, 0.8, 1.0
Total Contents	100,000
Cache Capacity	4–20%
Consumer Request Rate	1.0 request/s
Placement Policy	LCE, CL4M, ProbCache, LCD, opt-Cache
Topology	GEANT

The obtained results have been compared with state-of-the-art content replacement policies, including LRU, LFU, FIFO, and LFRU. To check the effectiveness of our approach, we compared popular cache placement policies, including Leave Copy Everywhere (LCE) [27], Cache Less for More (CL4M) [31], ProbCache [32], Leave Copy Down (LCD) [33], and opt-Cache [10], with our proposed replacement policy (IMU). These placement policies indicate the more redundant data to less redundant data in the network, respectively [10]. These placement policies indicate the more redundant data to less redundant data in the network. These results prove the effectiveness of our proposed technique with different cache sizes and populations, using various performance metrics such as cache hit ratio, latency, link load, and path stretch. These performance metrics are compared one by one, as explained below.

4.1. Cache Hit Ratio

The cache hit ratio is an essential metric for evaluating the performance of CCN cache. It identifies the response to network cache storage, in which content is cached locally within a specific time frame. Two terms are important in the cache hit ratio. The first is the cache hit (requested content is found from the cache), and the second is the cache miss (unlike cache hit). When content is available in the cache, the content request does not forward to the publisher. Therefore, a higher hit ratio indicates good cache performance and represents low bandwidth utilization, reduction in latency, and low server load. The cache hit ratio is defined as follows:

$$Hit\ Ratio = \frac{Cache_{hits}}{Cache_{hits} + Server_{hits}} \tag{1}$$

Our proposed strategy, IMU, compared to existing well-known replacement strategies in terms of the cache hit ratio. We have extracted the results from low to high popularity and different cache sizes. We first comment that content eviction policies behave the same under different caching strategies. Regardless of the content eviction policy, we observe in Figure 2 that the opt_cache performs best and the LCE performs the worst in terms of the cache hit ratio. Moreover, different eviction policies affect the performance of the cache hit ratio.

Figure 2 illustrates that the IMU’s performance is better than the existing replacement strategies; this is because the IMU not only considers the time $T_{c\bar{f},s_i}$ but also the frequency $F_{\bar{c},s_i}$ of the requested content within the specific period W_T . When the W_T is expired, then all the $F_{\bar{c},s_i}$ initialize to their starting frequency ($F_{\bar{c},s_i} = 1$). Moreover, it helps to evict content from the cache space whose popularity increases for a while and decreases shortly. When the cache is full, it is evicted from the cache after selecting the least value of the maturity index $M_{c\bar{f},s_i}$. The advantage of immature content eviction from the cache is that most of the content is mature, which leads to a higher cache hit ratio.

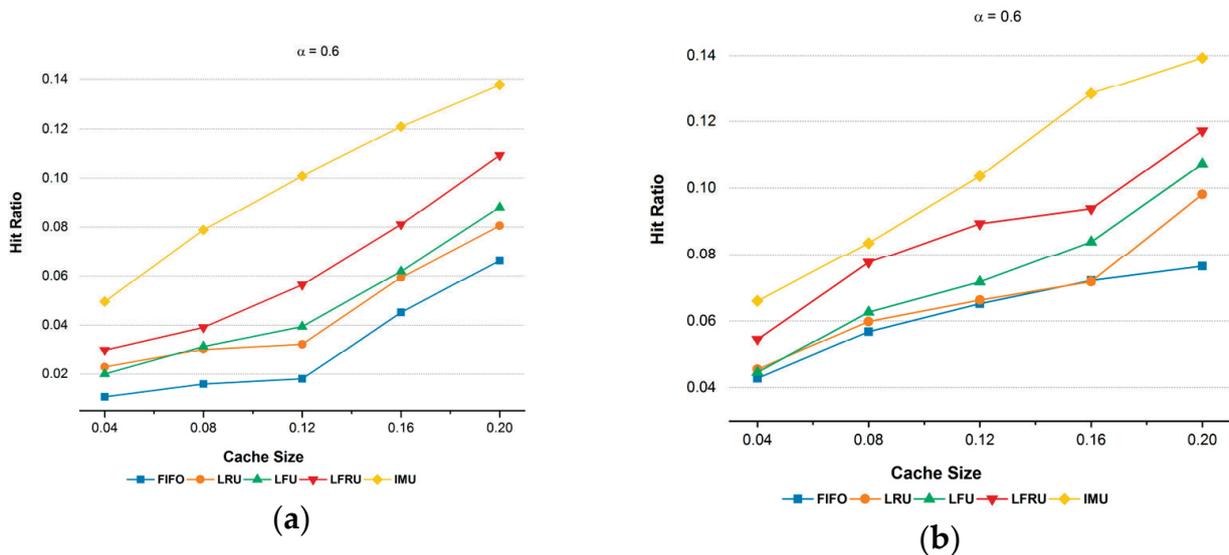


Figure 2. Cont.

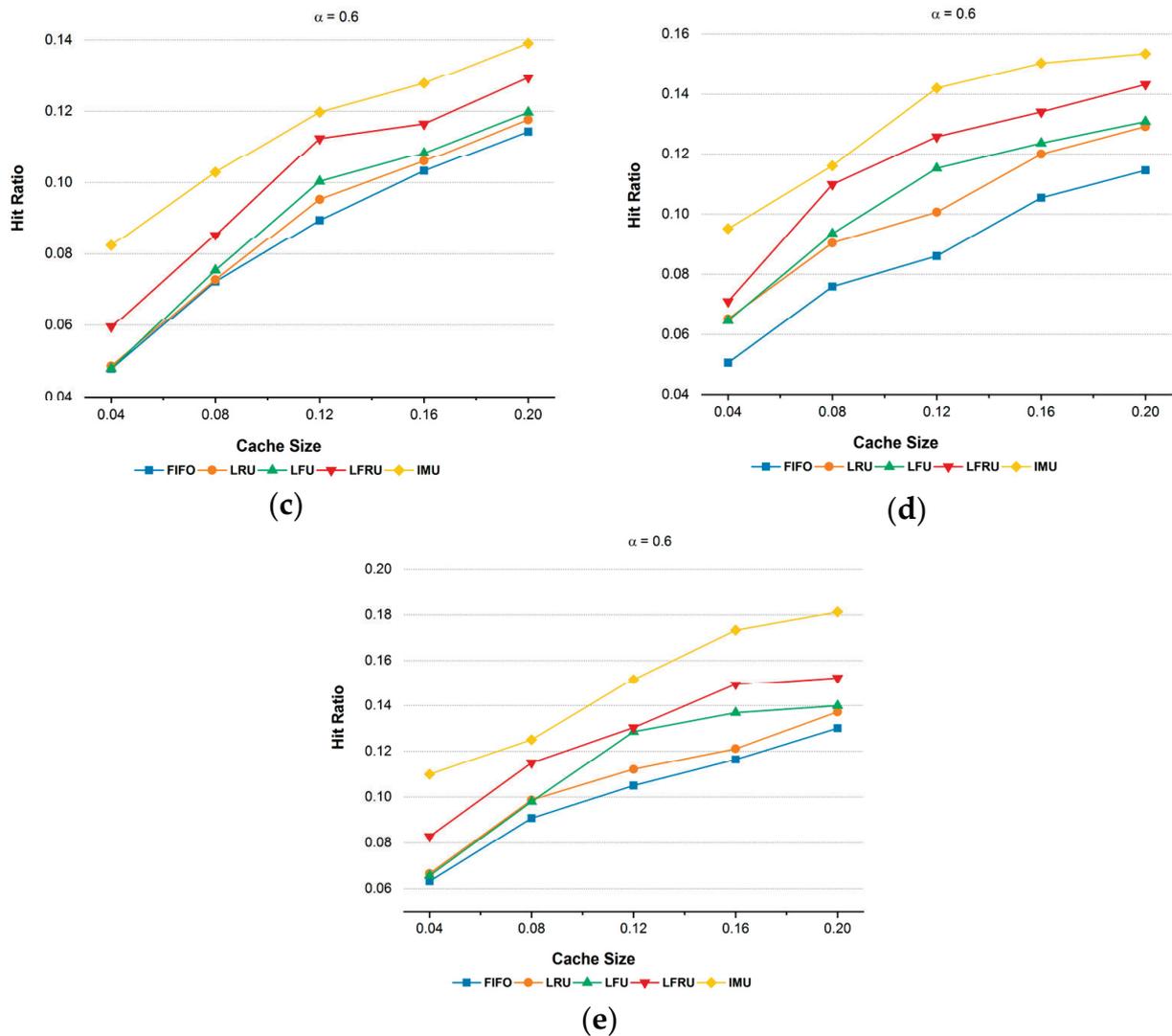


Figure 2. Caching hit ratio with different cache sizes and α by using different placement policies. (a) Cache hit ratio with LCE. (b) Cache hit ratio with CL4M. (c) Cache hit ratio with ProbCache. (d) Cache hit ratio with LCD. (e) Cache hit ratio with opt-Cache.

We observed that FIFO underperformed because contents are removed from the cache in the same order in which they were cached, regardless of how many times they were previously accessed. Besides, increasing cache space and similar content requests improve FIFO’s performance because the content stays in the cache for a longer period, which increases the chances of increasing the cache hit ratio. LFU performs better than LRU when the cache size is large and the content is repeatedly requested because LFU considers the frequency of the requested content, while LRU does not. Moreover, LFU caches popular content and evicts unpopular content from the cache. Besides, contents are often evicted from the cache when the cache size is small. However, LFU displays low performance in small cache sizes. LFRU has better performance due to the coupling of LRU and LFU; however, when the content request rate is minimum from the maximum normalized request rate, the content is evicted from the unprivileged partition. Therefore, the new content is cached in the unprivileged partition. Besides, if the content request rate is higher than the maximum normalized request rate, it chooses the least recent content from the privileged partition and pushes that content into the unprivileged partition. Hence, new content is cached in the privileged partition and hit counter associated with each partition. However, content that loses popularity stays in the unprivileged partition for a long time due to its

high frequency. IMU outperformed FIFO, LRU, LFU, and LFRU in terms of the cache hit ratio by 48.33%, 30.07%, 26.34%, and 14.31%, respectively.

The percentage (%) of IMU performance in different popularities, and low to high cache sizes with different content placement strategies, is presented in Table 10. We observed that IMU is outperformed with low popularity because, if such content is popular for some time but its popularity decreases with time and its frequency is high, then IMU evicts this content from the cache space. When the cache space is low and the popularity of the content changes frequently, it becomes very difficult for the content eviction policy to decide which content should be removed from the cache space. Hence, the IMU policy evicts immature content from the cache space and gives each content an equal opportunity to define its maturity/immaturity level. Such content is not removed from the cache space that is gaining popularity.

4.2. Path Stretch (Hop Count)

Path stretch indicates the distance traveled to the content provider by the consumer’s interest. The value of the path stretch is low when the consumer’s interest packet is found from the routing path. Therefore, the better content replacement policy identifies content that users are interested in and that is mature. Such content should not be evicted from the cache. If such content is evicted from the cache, the publisher’s load and bandwidth utilization will be high. Therefore, a better content replacement strategy should be to minimize the hops between the consumer and the publisher. Path stretch is defined as follows:

$$Path\ Stretch = \frac{\sum_{i=1}^n Hop - Traveled}{\sum_{i=1}^n THop - Hop} \tag{2}$$

where $\sum_{i=1}^n Hop - Traveled$ is the number of hops between the consumer and publisher nodes covered by consumer interest. The value $\sum_{i=1}^n THop - Hop$ denotes the total number of hops between the consumer and the provider. n represent the total number of generated interests for specific content.

Table 10. IMU cache hit ratio percentage improvement.

α	Cache Size	Placement	FIFO (%)	LRU (%)	LFU (%)	LFRU (%)
0.6	4% to 20%	LCE	320.35	133.19	121.34	64.43
0.6		CL4M	64.17	52.41	41.74	20.14
0.6		ProbCache	38.71	35.01	32.35	16.53
0.6		LCD	56.49	31.90	26.71	14.37
0.6		opt-Cache	48.84	40.46	33.82	18.57
0.8	4% to 20%	LCE	47.34	36.94	27.08	17.86
0.8		CL4M	27.84	25.22	18.72	12.93
0.8		ProbCache	17.64	14.11	11.70	6.38
0.8		LCD	20.15	14.08	13.66	6.53
0.8		opt-Cache	22.30	17.98	17.12	9.48
1.0	4% to 20%	LCE	25.60	22.04	21.29	14.05
1.0		CL4M	11.47	9.76	9.24	5.35
1.0		ProbCache	7.56	6.21	6.38	1.98
1.0		LCD	7.49	4.96	6.82	2.59
1.0		opt-Cache	9.02	6.74	7.21	3.46

Figure 3 illustrates that the IMU’s performance in terms of path stretch is better than other existing replacement policies. The placement strategy chooses the location of the cache, which may reduce the number of hops. IMU removes content that has been in the cache for a long time but has not matured. Therefore, when immature content is removed from the cache and new content is cached so that the consumer’s requested content is available nearby, the request is not forward to the publisher. However, cached content on nearby routers is mostly popular or close to being popular.

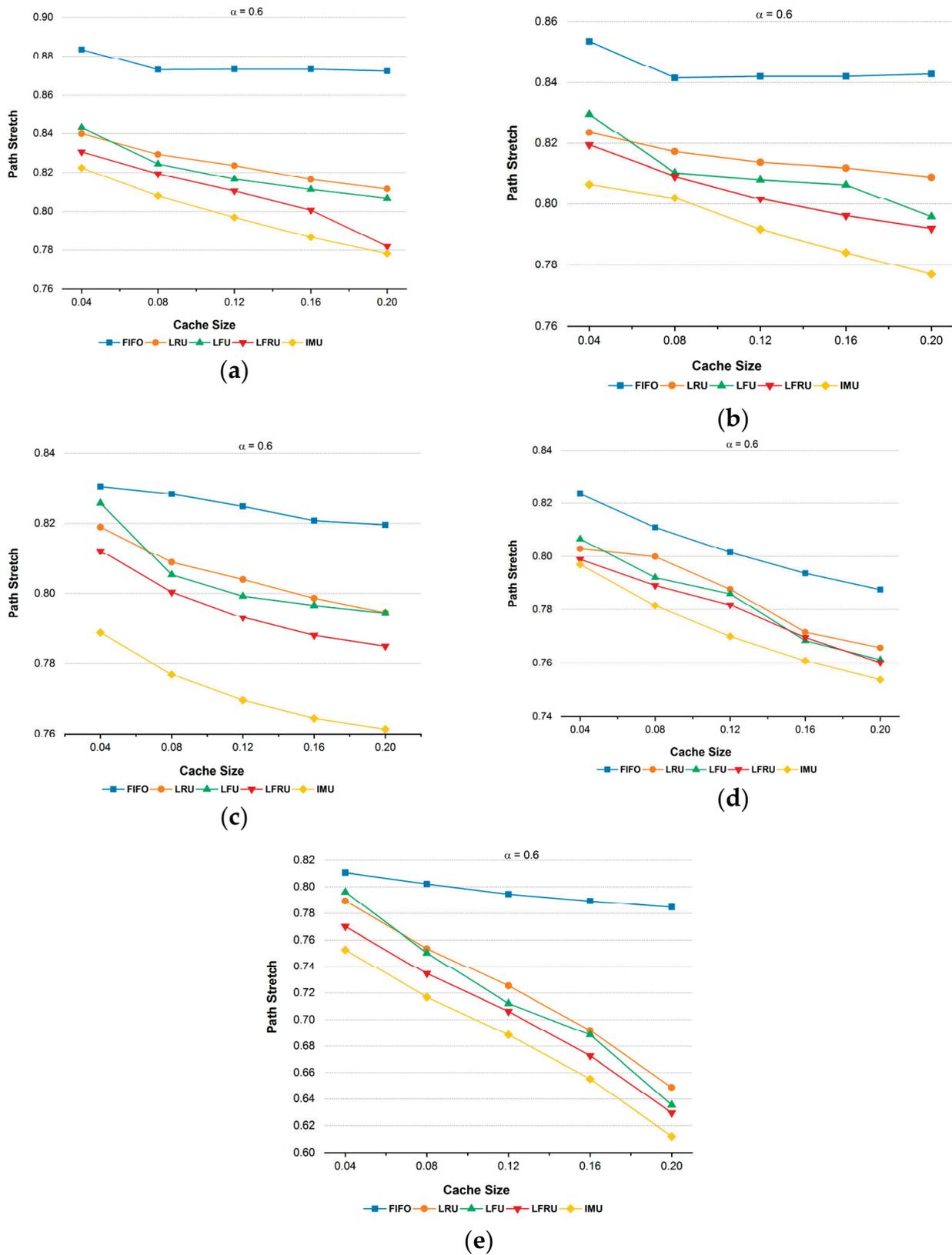


Figure 3. Path stretch with different cache sizes and α , using different placement policies. (a) Path stretch with LCE. (b) Path stretch with CL4M. (c) Path stretch with ProbCache. (d) Path stretch with LCD. (e) Path stretch with opt-Cache.

FIFO, LRU, and LFU represent the high path stretch due to content selection for eviction based on a single factor. FIFO content is evicted in the order in which it was cached. However, no matter how many times the content has been accessed, the timeline of popular and unpopular contents in FIFO will be the same, which increases the path stretch value. Figure 4 indicates that LRU is better than LFU when the cache size is smaller; however, as the cache size increases, the performance of LFU improves because LFU considers the popularity of content. Therefore, as cache size increases, popular content stays longer in the cache. LRU ignores the popularity of the content and the least recently used content evicts from the cached. However, content that is not popular, but, over time, their request keeps coming, are present in the cache space, making the path stretch higher. LFRU divides the cache space into two parts: LRU used privilege partition and LFU used unprivileged partition. With the higher request rate, the least recently used content has been evicted from the privilege partition and that content pushes it to the unprivileged partition. When unpopular content is pushed into the unprivileged partition, the content stays in the cache space for a long time. Further, these techniques are not focused on the maturity of the content. IMU outperformed FIFO, LRU, LFU, and LFRU in terms of path stretch by 11.33%, 6.16%, 5.77%, and 3.82%, respectively.

Table 11 illustrates the improvement of IMU in terms of path stretch using different content placement strategies with content eviction policies. We have observed IMU perform better in low to high cache space. In addition, IMU is better in high popularity. When the cache space is full, IMU selects immature content and evicts it from the cache. Therefore, popular content and content that may be popular remain in the cache. However, the consumer’s request for specific content is fulfilled from the nearest node.

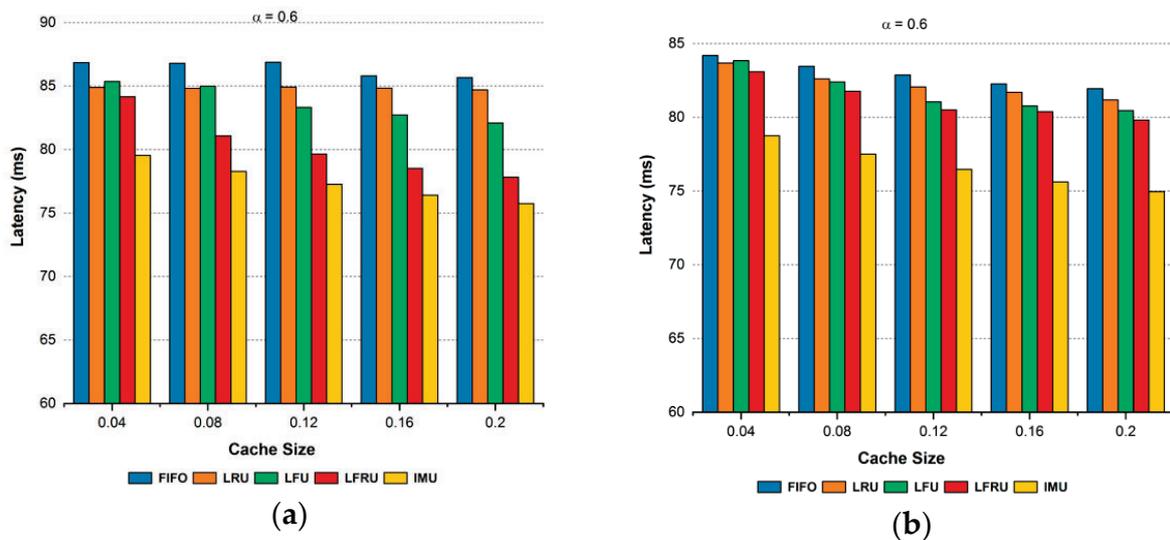


Figure 4. Cont.

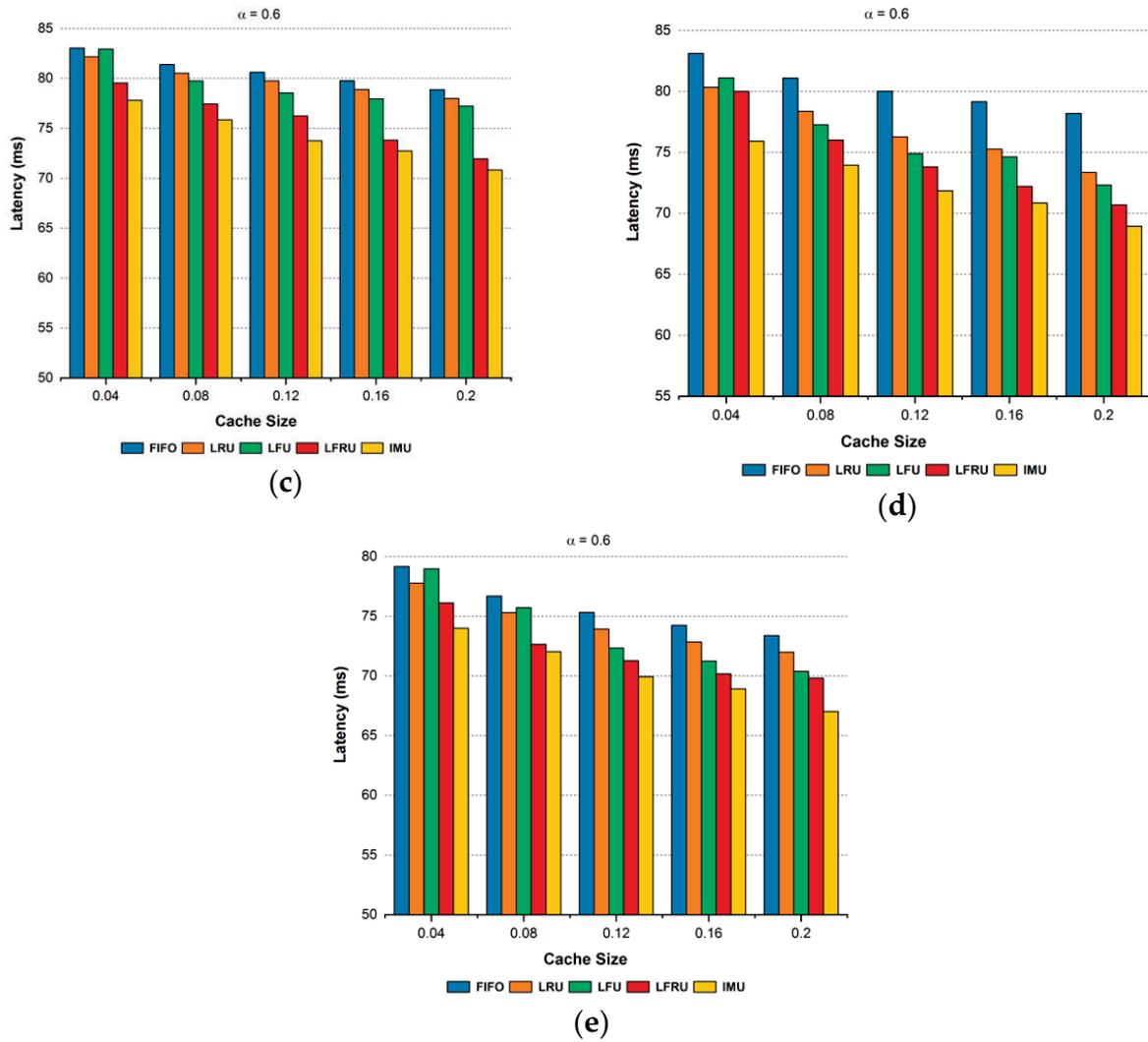


Figure 4. Latency with different cache sizes and α , using different placement policies. (a) Latency with LCE. (b) Latency with CL4M. (c) Latency with ProbCache. (d) Latency with LCD. (e) Latency with opt-Cache.

Table 11. IMU path stretch percentage improvement.

α	Cache Size	Placement	FIFO (%)	LRU (%)	LFU (%)	LFRU (%)
0.6		LCE	8.76	3.97	2.68	1.26
0.6		CL4M	6.17	2.80	2.20	1.42
0.6	4% to 20%	ProbCache	6.38	4.07	3.98	2.95
0.6		LCD	3.84	1.65	1.30	0.94
0.6		opt-Cache	14.02	5.10	4.35	2.51
0.8		LCE	10.32	4.88	4.30	2.44
0.8		CL4M	8.56	5.40	5.34	3.14
0.8	4% to 20%	ProbCache	9.43	4.20	3.62	3.05
0.8		LCD	7.42	4.96	4.20	3.16
0.8		opt-Cache	10.14	6.10	5.56	3.18
1.0		LCE	21.51	14.47	14.54	10.00
1.0		CL4M	20.35	9.04	9.27	6.25
1.0	4% to 20%	ProbCache	17.03	8.16	8.30	5.25
1.0		LCD	10.73	8.49	8.69	7.13
1.0		opt-Cache	15.30	9.06	8.17	4.68

4.3. Latency

Latency indicates the delay in the delivery of requests and content from consumers. It is a vital metric for evaluating the performance of the CCN cache, and it is defined as follows:

$$\text{Latency} = \text{Request Travel Delay} + \text{Content Travel Delay} \quad (3)$$

The IMU provides low latency because it evicts the most suitable content from the cache, based on immaturity. If the cache is full, the IMU jointly considers the frequency and time and selects the content for potential eviction from the cache. Hence, more popular and mature content will be in the cache, and content that may be popular. However, most consumer requests are satisfied along the routing path, which reduces latency. Figure 5 illustrates that IMU's performance is better than other content replacement policies, regarding latency with different cache sizes and popularity.

Figure 5 illustrates that FIFO represents a high latency because the duration of popular and unpopular content is the same. However, latency increases when popular content is evicted from the cache. LRU ignores the popularity of the content. Therefore, requests for less popular content come before eviction, and the content will remain in the cache, which causes high latency. LFU considers the frequency of the content, and contents that increase in frequency over a short period of time but are no longer popular; such contents use cache space due to their high frequency.

Therefore, fresh contents are reduced in the cache, which increases the latency. LFRU performs better than the previous two discussed replacement techniques because LRU and LFU are used together. When the request rate is high, then the required processing should be high, because the least recently used content is evicted from the privileged partition and pushed to the unprivileged partition and associated with the access history of content. In addition, low-frequency content is evicted from unprivileged partition. However, content with a high access history that is no longer popular will spend more time in the cache space, reducing the freshness of the content. IMU outperformed FIFO, LRU, LFU, and LFRU in terms of latency by 12.32%, 9.97%, 9.08%, and 5.91%, respectively.

When the alpha equals 0.8 with cache size is 0.04, IMU is 64.44 ms, which is 9.45% lower than LFRU (71.16 ms), 13.90% lower than LFU (74.84 ms), 13.34% lower than LRU (74.35 ms), and 15.60% lower than FIFO (76.34 ms). When the alpha equals 0.8, the cache size is 0.12, IMU is 61.03 ms, which is 5.40% lower than LFRU (64.51 ms), 9.65% lower than LFU (67.55 ms), 12.05% lower than LRU (69.39 ms), and 14.50% lower than FIFO (71.38 ms). When the alpha equals 0.8 with cache size is 0.2, IMU is 60.05 ms, which is 3.71% lower than LFRU (62.37 ms), 5.58% lower than LFU (63.60 ms), 8.93% lower than LRU (65.94 ms), and 11.59% lower than FIFO (67.92 ms). Therefore, as the cache size increases, the latency is reduced because more content in the network can be cached.

The latency improvement performance from IMU is illustrated in Table 12, using different content placement strategies with low to high popularities and cache sizes. We have observed that, as the popularity of content increases, so performs IMU. The IMU evicts content from the cache that has been in the cache for a long time and has few requests. Furthermore, content that has been in high demand for some time but declined over time has also been evicted from the cache. Therefore, the cache contains mostly mature content. However, when a consumer requests specific content, the consumer's request does not reach the publisher because the consumer is satisfied along the routing path.

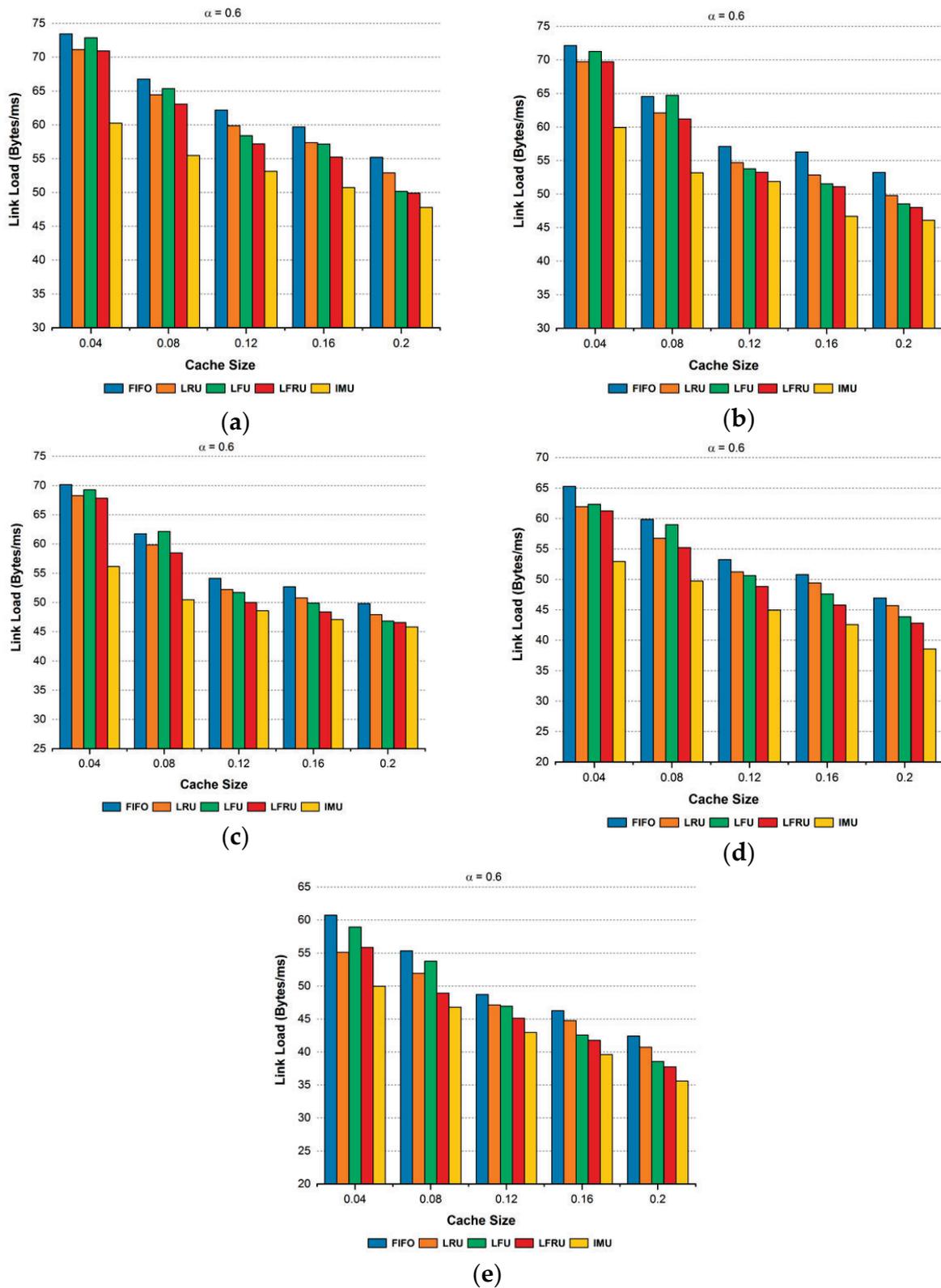


Figure 5. Link load with different cache sizes and α , using different placement policies. (a) Link load with LCE. (b) Link load with CL4M. (c) Link load with ProbCache. (d) Link load with LCD. (e) Link load with opt-Cache.

Table 12. IMU latency percentage improvement.

α	Cache Size	Placement	FIFO (%)	LRU (%)	LFU (%)	LFRU (%)
0.6	4% to 20%	LCE	10.37	8.71	7.48	3.45
0.6		CL4M	7.58	6.79	6.17	5.49
0.6		ProbCache	8.13	7.12	6.43	2.10
0.6		LCD	10.00	5.76	4.90	6.55
0.6		opt-Cache	7.12	5.37	4.51	2.26
0.8	4% to 20%	LCE	14.01	11.55	10.08	6.08
0.8		CL4M	14.32	9.65	8.71	6.30
0.8		ProbCache	7.39	6.12	5.44	1.95
0.8		LCD	10.12	7.01	6.36	4.07
0.8		opt-Cache	9.26	7.18	5.97	3.38
1.0	4% to 20%	LCE	21.37	19.60	18.87	9.98
1.0		CL4M	18.36	15.96	15.24	13.53
1.0		ProbCache	14.55	12.39	12.24	7.51
1.0		LCD	14.61	11.90	11.08	6.89
1.0		opt-Cache	17.67	14.39	12.74	9.05

4.4. Link Load

Link load indicates the total number of bytes (consumer's request size and content size) traversed for retrieving the interesting content at the specific time limit. It measures bandwidth usage in the network and is defined as follows:

$$\text{Link Load} = \frac{(\text{request}_{\text{size}} \times \text{request}_{\text{link_count}}) + (\text{content}_{\text{size}} \times \text{content}_{\text{link_count}})}{\text{Duration}} \quad (4)$$

$$\text{Duration} = \text{Content Retrieval Time} - \text{Content Request Time} \quad (5)$$

where, $\text{request}_{\text{size}}$ denotes the request's size in bytes, $\text{request}_{\text{link_count}}$ designates the number of the links traversed that reach the source, $\text{content}_{\text{size}}$ is the content size to retrieve, $\text{content}_{\text{link_count}}$ is the number of links where the content reaches the request's originator.

Figure 5 illustrates that IMU performs better than other existing strategies in terms of link load. IMU does not replace such content from the cache, which has a frequency over a certain period of time. Therefore, consumer request is mostly satisfied with the routing path or close to the consumer. Therefore, most of the content in the cache is of interest to the user. In addition, content that increases in frequency for some time but does not become popular later also removes such content from the cache. However, IMU maintains the freshness of the content as well as the mature content in the cache.

FIFO does not compete with the popularity of content because this technique only considers the order in which the content is cached and evicts the content from the cache in that order. Therefore, popular content is evicted from the cache. However, most consumer requests are satisfied with the publisher. Figure 5 demonstrates that LRU is better than LFU when the cache size is smaller. LFU performance improves as the cache size increases, as LFU takes into account the popularity of the content. Therefore, popular content stays in the cache for a long time, and the consumer's request is found in the cache space and not forwarded to the publisher. In addition, content that increases in frequency stays in the cache space, even if it is not popular. However, this is a misuse of cache space and leads to a higher link load. LRU ignores the popularity of the content as well as the maturity of the content. Therefore, popular content requested in the past is likely to be used in the future, but recently requested content may be replaced with less popularity; thus, it does not adapt to changing workloads. When the request rate is high in LFRU, the least recently used content is evicted from the privileged partition and pushed towards the unprivileged partition, with complete access history. However, this content is no longer popular but has a high access history; this content spends more time in cache space, which causes high

link load. IMU outperformed FIFO, LRU, LFU, and LFRU in terms of link load by 18.04%, 13.61%, 12.49%, and 9.53%, respectively.

When the alpha equals 0.8, with a cache size of 0.04, IMU is 55.48 bytes/ms, which is 16.41% lower than LFRU (66.37 ms), 19.60% lower than LFU (69.01 bytes/ms), 17.85% lower than LRU (67.53 bytes/ms), and 20.57% lower than FIFO (69.85 bytes/ms). When the alpha equals 0.8, with a cache size is 0.12, IMU is 48.99 bytes/ms, which is 15.25% lower than LFRU (57.81 bytes/ms), 18.57% lower than LFU (60.16 bytes/ms), 19.21% lower than LRU (60.64 bytes/ms), and 22.19% lower than FIFO (62.96 bytes/ms). When the alpha equals 0.8, with cache size is 0.2, IMU is 44.93 bytes/ms, which is 10.40% lower than LFRU (50.15 bytes/ms), 10.68% lower than LFU (50.30 bytes/ms), 11.51% lower than LRU (50.78 bytes/ms), and 15.37% lower than FIFO (53.09 bytes/ms). As the cache size increases, we observed that the link load decreases, as the proposed scheme removes immature content from the cache. Therefore, IMU maintains the data freshness with popularity within the network. However, none of the previous eviction policies have adopted the concept of immaturity for content selection.

Table 13 describes the IMU's improvement in percentage (%) of the link load, which used different content placement strategies along with content eviction policies. We observed that IMU outperformed the other content eviction policies against low to high popularity and cache space. It performed better in a fully redundant and low redundancy environments. IMU contains the most popular and mature content in the cache and makes better use of cache space. Moreover, the consumer is mostly satisfied along the routing path when requesting content. Therefore, the link load value is low because the request is not sent to the publisher.

Table 13. IMU link load percentage improvement.

α	Cache Size	Placement	FIFO (%)	LRU (%)	LFU (%)	LFRU (%)
0.6	4% to 20%	LCE	15.58	12.34	11.49	9.30
0.6		CL4M	14.83	10.52	10.34	8.46
0.6		ProbCache	13.45	10.40	10.30	7.61
0.6		LCD	17.08	13.71	12.93	9.68
0.6		opt-Cache	15.07	10.43	10.27	6.10
0.8	4% to 20%	LCE	20.32	17.19	17.19	14.40
0.8		CL4M	21.54	18.04	17.45	15.28
0.8		ProbCache	19.18	16.17	14.95	12.04
0.8		LCD	23.44	18.48	16.51	14.64
0.8		opt-Cache	22.03	11.74	9.19	6.97
1.0	4% to 20%	LCE	20.89	17.40	15.27	13.09
1.0		CL4M	16.10	11.15	9.99	5.95
1.0		ProbCache	16.23	12.78	11.73	7.43
1.0		LCD	17.56	11.62	9.37	6.40
1.0		opt-Cache	17.30	12.14	10.35	5.55

5. Conclusions and Future Work

In-network caching is one of the essential features in the CCN architecture network, allowing content items to be cached in the router nodes for some time, to meet subsequent consumer requests. Due to the limited cache capacity in the node, any cached content in the cache needs to be evicted to accommodate new content. Content replacement policy is responsible for choosing the right content against defined criteria. Existing cache replacement policies use the concept of popularity or time for content eviction. However, when content loses its popularity after becoming very popular in a certain period, it remains in the cache space. Moreover, content is evicted from the cache space before it becomes popular. Therefore, the proposed policy handles cached items that lose their popularity over a specific time frame and remain in the cache for a long time. We introduced the new concept of content maturity and immaturity for content eviction in CCN. The

proposed content replacement policy (IMU) uses the concept of maturity/immaturity of the content. This policy finds the content maturity index by using the content arrival time and its frequency. Also, it determines the maturity level through a maturity classifier. We have performed extensive simulations to evaluate the proposed content replacement policy, using the Icarus simulator under different cache sizes and content popularity. The simulation results indicate that the proposed policy outperformed recent and baseline content replacement policies (FIFO, LRU, LFU, and LFRU). The results demonstrate that the proposed policy is better in terms of the cache hit ratio, latency, path stretch, and link load. In the future, this work can be extended to use the content replacement policy (IMU) with different constraints in different use cases. Another potential future work is in-depth investigation of content diversity for the nodes with very high content popularity.

Author Contributions: Conceptualization, S.R., S.A.R. and F.A.G.; methodology, S.R.; software, S.R.; validation, S.R., S.A.R. and F.A.G.; formal analysis, S.R. and F.A.G.; investigation, S.R., S.A.R. and F.A.G.; writing—original draft preparation, S.R.; writing—review and editing, S.R., S.A.R. and F.A.G.; supervision, S.A.R. and F.A.G.; funding acquisition, S.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Universiti Teknologi Malaysia (UTM) under UTM Research University Grant Scheme: (VOT Q.J130000.3613.03M41), Universiti Teknologi Malaysia, Johor Bahru, Malaysia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gui, Y.; Chen, Y. A Cache Placement Strategy Based on Compound Popularity in Named Data Networking. *IEEE Access* **2020**, *8*, 196002–196012. [[CrossRef](#)]
2. Khandaker, F.; Oteafy, S.; Hassanein, H.S.; Farahat, H. A functional taxonomy of caching schemes: Towards guided designs in information-centric networks. *Comput. Netw.* **2019**, *165*, 106937. [[CrossRef](#)]
3. Noh, H.; Song, H. Progressive Caching System for Video Streaming Services Over Content Centric Network. *IEEE Access* **2019**, *7*, 47079–47089. [[CrossRef](#)]
4. Zheng, X.; Wang, G.; Zhao, Q. A Cache Placement Strategy with Energy Consumption Optimization in Information-Centric Networking. *Futur. Internet* **2019**, *11*, 64. [[CrossRef](#)]
5. Ioannou, A.; Weber, S. A survey of caching policies and forwarding mechanisms in information-centric networking. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2847–2886. [[CrossRef](#)]
6. Ullah, R.; Rehman, M.A.U.; Naeem, M.A.; Kim, B.-S.; Mastorakis, S. ICN with edge for 5G: Exploiting in-network caching in ICN-based edge computing for 5G networks. *Futur. Gener. Comput. Syst.* **2020**, *111*, 159–174. [[CrossRef](#)]
7. Naeem, M.A.; Ali, R.; Alazab, M.; Yhui, M.; Zikria, Y. Bin Enabling the content dissemination through caching in the state-of-the-art sustainable information and communication technologies. *Sustain. Cities Soc.* **2020**, *61*, 102291. [[CrossRef](#)]
8. Naeem, M.A.; Rehman, M.A.U.; Ullah, R.; Kim, B.-S. A Comparative Performance Analysis of Popularity-Based Caching Strategies in Named Data Networking. *IEEE Access* **2020**, *8*, 50057–50077. [[CrossRef](#)]
9. Ji, Y.; Zhang, X.; Liu, W.; Zhang, G. Replacement based content popularity and cache gain for 6G content-centric network. *Phys. Commun.* **2021**, *44*, 101238. [[CrossRef](#)]
10. Qazi, F.; Khalid, O.; Rais, R.N.B.; Khan, I.A. Optimal Content Caching in Content-Centric Networks. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 6373960. [[CrossRef](#)]
11. Khattak, H.; Ul Amin, N.; Din, I.U.; Insafulah; Iqbal, J. LeafPopDown: Leaf Popular Down Caching Strategy for Information-Centric Networking. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 148–151. [[CrossRef](#)]
12. Kumar, S.; Tiwari, R. Dynamic popularity window and distance-based efficient caching for fast content delivery applications in CCN. *Eng. Sci. Technol. Int. J.* **2021**, *24*, 829–837. [[CrossRef](#)]
13. Liu, H.; Han, R. A Hierarchical Cache Size Allocation Scheme Based on Content Dissemination in Information-Centric Networks. *Futur. Internet* **2021**, *13*, 131. [[CrossRef](#)]
14. Amadeo, M.; Ruggeri, G.; Campolo, C.; Molinaro, A. Diversity-improved caching of popular transient contents in vehicular named data networking. *Comput. Netw.* **2021**, *184*, 107625. [[CrossRef](#)]

15. Gui, Y.; Chen, Y. A Cache Placement Strategy Based on Entropy Weighting Method and TOPSIS in Named Data Networking. *IEEE Access* **2021**, *9*, 56240–56252. [[CrossRef](#)]
16. Kalghoum, A.; Gammar, S.M.; Saidane, L.A. Towards a novel cache replacement strategy for Named Data Networking based on Software Defined Networking. *Comput. Electr. Eng.* **2018**, *66*, 98–113. [[CrossRef](#)]
17. Seyyed Hashemi, S.N.; Bohlooli, A. Analytical characterization of cache replacement policy impact on content delivery time in information-centric networks. *Int. J. Commun. Syst.* **2019**, *32*, e4154. [[CrossRef](#)]
18. Ai, L.; Deng, Y.; Zhou, Y.; Feng, H. RUE: A caching method for identifying and managing hot data by leveraging resource utilization efficiency. *Softw. Pract. Exp.* **2021**. [[CrossRef](#)]
19. Saino, L.; Psaras, I.; Pavlou, G. Icarus: A caching simulator for information centric networking (icn). In Proceedings of the SimuTools, Lisbon, Portugal, 17–19 March 2014; ICST: Ghent, Belgium, 2014; Volume 7, pp. 66–75.
20. Goian, H.S.; Al-Jarrah, O.Y.; Muhaidat, S.; Al-Hammadi, Y.; Yoo, P.; Dianati, M. Popularity-based Video Caching Techniques for Cache-enabled Networks: A survey. *IEEE Access* **2019**, *7*, 27699–27719. [[CrossRef](#)]
21. Pfender, J.; Valera, A.; Seah, W.K.G. Performance comparison of caching strategies for information-centric IoT. In Proceedings of the Proceedings of the 5th ACM Conference on Information-Centric Networking, Boston, MA, USA, 21–23 September 2018; ACM: New York, NY, USA, 2018; pp. 43–53.
22. Paschos, G.S.; Iosifidis, G.; Tao, M.; Towsley, D.; Caire, G. The role of caching in future communication systems and networks. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 1111–1125. [[CrossRef](#)]
23. Chen, L.; Song, L.; Chakareski, J.; Xu, J. Collaborative Content Placement among Wireless Edge Caching Stations with Time-to-Live Cache. *IEEE Trans. Multimed.* **2019**, *22*, 432–444. [[CrossRef](#)]
24. Karakostas, G.; Serpanos, D.N. Exploitation of different types of locality for web caches. In Proceedings of the ISCC 2002 Seventh International Symposium on Computers and Communications, Taormina-Giardini Naxos, Italy, 1–4 July 2002; IEEE: Piscataway, NJ, USA, 2002; pp. 207–212.
25. Bilal, M.; Kang, S.-G. A cache management scheme for efficient content eviction and replication in cache networks. *IEEE Access* **2017**, *5*, 1692–1701. [[CrossRef](#)]
26. Zhang, Y.; Tan, X.; Li, W. PPC: Popularity prediction caching in ICN. *IEEE Commun. Lett.* **2017**, *22*, 5–8. [[CrossRef](#)]
27. Meddeb, M.; Dhraief, A.; Belghith, A.; Monteil, T.; Drira, K.; Mathkour, H. Least fresh first cache replacement policy for NDN-based IoT networks. *Pervasive Mob. Comput.* **2019**, *52*, 60–70. [[CrossRef](#)]
28. Khan, J.; Westphal, C.; Garcia-Luna-Aceves, J.; Ghamri-Doudane, Y. Nice: Network-oriented information-centric centrality for efficiency in cache management. In Proceedings of the 5th ACM Conference on Information-Centric Networking, Boston, MA, USA, 21–23 September 2018; ACM: New York, NY, USA, 2018; pp. 31–42.
29. Lal, K.N.; Kumar, A. A popularity based content eviction scheme via betweenness-centrality caching approach for content-centric networking (CCN). *Wirel. Netw.* **2019**, *25*, 585–596. [[CrossRef](#)]
30. Kumar, S.; Tiwari, R. An efficient content placement scheme based on normalized node degree in content centric networking. *Cluster Comput.* **2021**, *24*, 1277–1291. [[CrossRef](#)]
31. Chai, W.K.; He, D.; Psaras, I.; Pavlou, G. Cache “less for more” in information-centric networks. In Proceedings of the International Conference on Research in Networking, Prague, Czech Republic, 21–25 May 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 27–40.
32. Naeem, M.A.; Nor, S.A.; Hassan, S.; Kim, B.-S. Compound Popular Content Caching Strategy in Named Data Networking. *Electronics* **2019**, *8*, 771. [[CrossRef](#)]
33. Seetharam, A. On caching and routing in information-centric networks. *IEEE Commun. Mag.* **2017**, *56*, 204–209. [[CrossRef](#)]
34. Ren, J.; Qi, W.; Westphal, C.; Wang, J.; Lu, K.; Liu, S.; Wang, S. MAGIC: A distributed MAX-gain in-network caching strategy in information-centric networks. In Proceedings of the 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 27 April–2 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 470–475.

Article

Attitudes toward Applying Facial Recognition Technology for Red-Light Running by E-Bikers: A Case Study in Fuzhou, China

Yanqun Yang¹, Danni Yin¹, Said M. Easa² and Jiang Liu^{3,*}

¹ College of Civil Engineering, Fuzhou University, Fuzhou 350008, China; yangyanqun@fzu.edu.cn (Y.Y.); yin-dn@foxmail.com (D.Y.)

² Department of Civil Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada; seasa@ryerson.ca

³ School of Architecture and Urban-Rural Planning, Fuzhou University, Fuzhou 350008, China

* Correspondence: jiang.liu@fzu.edu.cn

Abstract: The application of facial recognition technology (FRT) can effectively reduce the red-light running behavior of e-bikers. However, the privacy issues involved in FRT have also attracted widespread attention from society. This research aims to explore the public and traffic police's attitudes toward FRT to optimize the use and implementation of FRT. A structured questionnaire survey of 270 people and 94 traffic police in Fuzhou, China, was used. In the research, we use several methods to analyze the investigation data, including Mann–Whitney U test, Kruskal–Wallis test, and multiple correspondence analysis. The survey results indicate that the application of FRT has a significant effect on reducing red-light running behavior. The public's educational level and driving license status are the most influential factors related to their attitudes to FRT ($p < 0.001$). Public members with these attributes show more supportive attitudes to FRT and more concerns about privacy invasion. There are significant differences between the public and traffic police in attitudes toward FRT ($p < 0.001$). Compared with the public, traffic police officers showed more supportive attitudes to FRT. This research contributes to promoting the application of FRT legitimately and alleviating people's concerns about the technology.

Keywords: facial recognition technology; e-biker; red-light running behavior; privacy invasion

Citation: Yang, Y.; Yin, D.; Easa, S.M.; Liu, J. Attitudes toward Applying Facial Recognition Technology for Red-Light Running by E-Bikers: A Case Study in Fuzhou, China. *Appl. Sci.* **2022**, *12*, 211. <https://doi.org/10.3390/app12010211>

Academic Editors: Giovanni Randazzo, Anselme Muzirafuti and Dimitrios S. Parafaros

Received: 14 December 2021

Accepted: 22 December 2021

Published: 26 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The e-bike is a vital means of transportation in many Chinese cities [1], given its convenience and fast characteristics. As of 2021, the number of e-bikes in China has reached nearly 300 million. The rapidly increasing number of e-bikes has resulted in increased accidents. In 2019, there were approximately 8639 deaths and 44,677 injuries caused by e-bike accidents, which is close to 70% of non-motorized vehicle casualties [2]. In China, e-bikes are categorized as non-motorized vehicles, and riders must drive on non-motorized lanes and comply with the same regulations as bicycles [3]. However, red-light running, illegal use of motor vehicle lanes, and over-speed cycling are the main reasons for accidents involving e-bikes [4]. These violations are often caused by low traffic safety awareness [5], among which running the red light is the leading cause of e-bike accidents [6,7]. Previous research points out that e-bikers run a red light more frequently than traditional bicycle riders [8], and e-bikes are faster than bicycles before collisions, with a higher risk ratio at intersections [9].

To reduce the red-light running behavior of e-bikers, many cities in China, such as Shenzhen, Shanghai, Jinan, and Fuzhou, have launched the Red-light Record System to regulate traffic violations. The system can capture and recognize the red-light running behavior of pedestrians and e-bikers and expose the screen's on-site violation images. The application of this system has achieved satisfying results. Since the Red-light Record System trial in Jiangbei, Chongqing, the violation rate of pedestrians and e-bikes has dropped from

40% to less than 3%. With facial recognition technology (FRT), traffic police need not face the violators, and the difficulty of enforcement is reduced with the evidence provided from FRT. However, there is no specific law related to applying FRT in the traffic area. Thus, different cities have different standards for FRT. China has not yet established a unified standard for the application of FRT in transportation. The application of FRT has aroused public concerns about privacy invasion. Controversial opinions exist regard the extent to which violators' information is exposed and the suspicion around releasing personal privacy. Furthermore, whether it is a punishment beyond the law.

Thus, to understand the application effects of FRT, we investigate the attitudes of two significant stakeholders (the public and traffic police) on applying FRT in Fuzhou, China. The study aims to determine: (1) The public's opinion on the privacy violation of exposing personal information of red-light running behavior, (2) how personal characteristics of the public affect their attitudes toward FRT, and (3) the attitudes of traffic police toward FRT. Based on the above analysis, we propose several practical suggestions to improve the efficiency and rationality of FRT.

The methodology of the study is shown in Figure 1. The methodology consists of a literature review, experimental design, questionnaire design, data collection, and statistical analysis. In the analysis, all statistical calculations and plots were performed using SPSS 22.0.

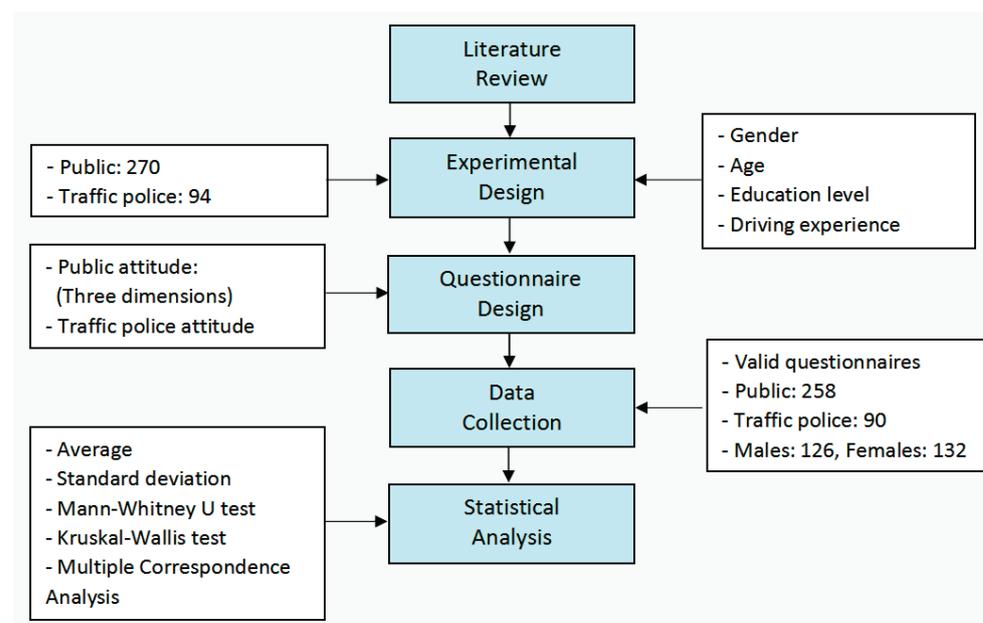


Figure 1. Study methodology.

2. Literature Review

2.1. Red-Light Running Behavior of E-Bikers

Many studies have been conducted to determine the factors that affect the red-light running behavior of e-bikers, mainly from external and internal perspectives. In terms of external factors, the higher acceleration rate and weight of an e-bike enables bikers to reach a higher speed than bicycle riding. Thus, e-bikers are more likely to run a red light [10,11]. Traffic conditions and situation factors have been verified to impact red-light running behavior [12]. They are also prone to accidents when the speed of an e-bike is underestimated by other road users [13]. As to internal factors, the attitudes of e-bikers are in close relationships with red-light running behavior. Red-light running intention and willingness could be predicted by the attitudes and past behaviors of e-bikers [14]. Self-discipline to follow traffic regulations, herd tendency, and past behaviors of e-bikers are crucial factors that affect the likelihood of accidents [15]. Of course, higher safety awareness

and more concern about their traffic risk could reduce dangerous riding behaviors [16]. An acceptable waiting time for e-bikers at signalized intersections is shorter than that of bicycle riders, which may also be one reason for the higher probability of red-light running behavior [17]. Some scholars have found that gender and age may affect red-light running behavior. In terms of gender, males are more likely to run against a red light than females [18]. Although the effect of age on red-light running behavior is still not clear, young and middle-aged people are more likely to run a red light [19–21]. Whether holding a driving license or not could also affect the red-light-running violation rate [22]. To sum up, the complexity of these influential factors poses a significant challenge to red-light running behavior.

2.2. Preventive Measures

To prevent red-light running behavior, different intervention measures have been taken, such as educational programs, enforcement activities, and social marketing [23]. These interventions could use the positive influence of e-biker groups to promote law-obeying behavior [21]. Educating and training e-bikers is fundamental to reducing red-light running behavior [24]. E-bikers were recommended to participate in training programs to provide relevant skills [25]. Education and training programs for e-bikers with different characteristics reduce their unsafe behavior [26]. Besides, a comprehensive e-bike treatment needs enforcement [27]. Some scholars have recommended launching an e-bike license system with point-based penalties by factoring in China's unique regional and political characteristics [28]. Police enforcement of traffic regulations could effectively curb the red-light running behavior of e-bikers [24].

Technical equipment is widely used as an essential supplementary measure to monitor red-light running behavior. The equipment includes red-light cameras for motor vehicle drivers [29] and red-light running detectors performed by a system that consists of a camera and computer embedded in a motor vehicle [30]. Recognition systems using different technologies are used to monitor the red-light running behavior of cyclists and pedestrians. These technologies include video sequences, adaptive mapping techniques, and trained classifiers. Most of these technologies are related to image recognition. The video sequence is applied to detect red-light running behavior [31]. A real-time pedestrian recognition system that ensures high accuracy using a deep learning classifier and zebra-crossing recognition techniques is proposed using an adaptive mapping technique and a dual camera mechanism [32]. Finally, a recognition system for recognizing people at a pedestrian crossing is developed, which includes a trained classifier and two sets of images taken from an open database containing images of city streets from outdoor cameras [33]. These technologies can be used for image recognition of pedestrians and cyclists.

Among this technical equipment, FRT could be the most advanced one to monitor the red-light running behavior of pedestrians and non-motorized vehicles. These systems use FRT, including the red-light automatic early warning system and the red-light snapping system. The former is used for cyclists and pedestrians with automatic crossing reminders, red-light recording, exposure, and information inquiry [34,35]. In order to address the issue that the targeted face is subject to varying conditions, particularly of illumination, a novel pedestrian detection algorithm with multi-source face images is proposed [36]. With the red-light snapping system, the tracking success rate is increased to 85%, and the number of simultaneous tracking reaches 25 people [37].

However, due to the sensitivity of biometric data and the heterogeneity and openness of the network environment, the privacy leakage of biometric data is difficult to avoid [38]. Therefore, how to improve face recognition accuracy while ensuring high security of private data has provoked fierce public discussion.

2.3. Regulations and Privacy Concerns about FRT

Although advanced technologies could improve traffic safety, there are drawbacks at the same time. The main problem is the risk of privacy invasion since these technologies

can collect, store, and share personal information [39]. For example, privacy and safety are the main concerns expressed concerning traffic enforcement drones, and the citizens once opposed this technology in Los Angeles. They felt the department would use drones to track and observe them [40]. Privacy concerns are also reflected in in-vehicle data recorders. This concern tends to hinder the acceptance of innovations [41].

There are limited studies on the application feedback of FRT in recognition of red-light running behavior. However, numerous studies have conducted public surveys about FRT application, indicating their concerns about privacy invasion. In many cases, their facial information is collected involuntarily [42], which may lead to undesirable results of intrusions of privacy [43]. The privacy concerns are affected by privacy control, which means giving users the autonomy to control their private information [44]. The legitimacy of FRT contributes to allay, deaden, or possibly circumvent privacy concerns. In other words, FRT with less legitimacy could heighten people's concerns about privacy [45]. FRT also raises concerns about control over personal information, where it is used, and the potential for misrecognition [46]. These concerns about privacy invasion that FRT may cause have attracted worldwide attention.

The application of FRT for legal regulation has become the focus of legislative protection in various countries. Many states in the US have issued several bills about FRT. Government agencies in the US are cautious about using FRT and focus on prohibitive regulations. For example, the Body Camera Accountability Act states that the operation of FRT with a camera is an invasion of personal privacy [47]. Non-governmental organizations in the US are more open to using FRT, and they allow the restricted use of FRT to a certain extent. For example, Illinois proposed the Biometric Information Privacy Act (BIPA) to regulate the collection, storage, use, retention, and destruction of biometric information, including facial feature information, through individual empowerment and enhanced obligations [48]. The EU also restricts the application of FRT strictly. The General Data Protection Regulation (GDPR) incorporates different types and properties of personal information and protects personal information through civil, administrative, and criminal measures. In exceptional circumstances, the processing must meet the requirements of legal, legitimate, consent, and voluntary [49].

In China, the protection of facial features about FRT is distributed in laws and regulations. The Civil Code became effective on 1 January 2021, stipulating a natural person's personal information is protected, and the personal information mainly includes a name, birthday, and ID number. However, the Civil Code does not stipulate the contents and methods of protection expressly. China has announced more detailed regulations on the facial feature information involved in applying FRT in administrative regulations, rules, and other normative documents. Information Security Technology-Personal Information Security Specification revised in March 2020 explicitly regulates that personal biometric information is sensitive personal information. Sensitive personal information needs special protection. For example, before collecting personal biometric information, the subject should be informed of the purpose, method, and scope of personal information, storage time, and other rules, and the subject's consent should be obtained. Personal biometric information should be stored separately from personally identifiable information. In principle, original personal biometric information should not be stored. However, these regulations are only recommended and not mandatory [50]. The regulations about the application of FRT in China need to be further improved.

3. Questionnaire Design and Data Collection

3.1. Research Background

The research was conducted in Fuzhou, China. The application of FRT in Fuzhou, dates back to 2016 when the Fuzhou Traffic Police Department launched the first Red-light Record System at the intersection of Yangqiao Road and Daming Road. The system automatically can capture the images of the violators when they run a red light and recognize their personal information, and Figure 2 is the screen part of this system. Figure 3 is the red-light

running behavior of e-bikers at the intersection. Then, the violators' mobile phones will receive a message from the system, including the time and place of the violations. When the violators pay the fine, their images will disappear from the screen.



Figure 2. Part of the red-light monitoring system.



Figure 3. Red-light running behavior of e-bikers.

At the end of 2019, about 2.09 million e-bikes registered in the five districts of Fuzhou, China, resulting in increased regulatory difficulty. However, the application of FRT is facing contrary opinions. On the one hand, the effectiveness of FRT is recognized by part of the public who believe that FRT is more a deterrent than just a fine and by the traffic police for whom the technology substantially reduces the need for on-site supervision and provides reasonable evidence for punishment. On the other hand, some members of the public hesitate about accepting FRT as they are unsure whether their privacy is infringed and whether the collected information can be effectively protected.

3.2. Measures

3.2.1. Public Investigation

Referring to the Motor Vehicle Risky Driver Behavior Scale [51], and according to the characteristics of e-bikers and behavior, we designed a public investigation questionnaire. The questionnaire consists of two parts. The first part consists of basic personal information, including gender, age, education level, and driving license (Table 1). The second part is the

public’s attitudes toward FRT, which includes three variables: Attitudes toward red-light running behavior, the application effect of FRT, and whether FRT violates privacy (Table 2). The first part of the questionnaire uses a single-choice form, and the second part uses a Likert five-level scale (from “strongly disagree = 1” to “strongly agree = 5”).

Table 1. Items of basic personal information.

Demographic Variables	Category
Gender	Male
	Female
Age	18–36
	37–54
	>54
Education level	Junior high school and below
	Senior high school
	College and undergraduate
	Postgraduate and above
Driving license	Have
	Do not have

Table 2. Survey items of public and traffic police attitudes toward FRT.

Variables	Item
A: Attitudes toward red-light running behavior of e-bikers	A1: Red-light running behavior of e-bikers has a negative impact on traffic.
	A2: Even if I have good riding skills, running a red light may be dangerous.
	A3: Although running a red light can shorten the travel time, it is prone to accidents and is unworthy.
	A4: Red-light running behavior is irresponsible to lives.
	A5: More e-bikers are running red-light in China cities, and management needs to be strengthened.
	A6: I am familiar with the traffic regulations related to e-bikes, and I ride per the regulations.
B: Application effect of FRT	B1: FRT can significantly reduce the red-light running behavior of e-bikers.
	B2: The application of FRT helps to strengthen personal traffic safety awareness.
	B3: Need to take specific measures to punish the identified behavior.
	B4: The information of violators shown on the screen has a deterrent effect on the public.
	B5: Most people will support and actively obey the application of FRT in traffic management.
	B6: FRT is progress of technology and is worth promoting.
C: Whether FRT violates privacy	C1: It is not a violation of personal privacy to show red-light running behavior on the screen.
	C2: The following information published on the screen is appropriate: The offender’s image running a red light (the face is covered), the middle part of the name is concealed, and the ID number is concealed in the middle digits.
	C3: Personal information identified by FRT will be strictly protected and will not be leaked.
	C4: The application of FRT also needs to be regulated by improving relevant laws.
	C5: When using FRT, the public’s right to know needs to be guaranteed.

Table 2. Cont.

Variables	Item
P: Traffic police's attitudes toward FRT	P1: I have a good understanding of applying FRT to manage the red-light running behavior of e-bikers.
	P2: FRT can significantly reduce the red-light running behavior of e-bikers.
	P3: FRT can reduce the management difficulty of traffic police.
	P4: Need to take specific measures to punish the identified behavior.
	P5: Require to give safety education to the identified violators.
	P6: Most people will support and actively obey the application of FRT in traffic management.
	P7: The application of FRT helps to strengthen personal traffic safety awareness.
	P8: It is not a privacy violation to show red-light running behavior on the screen.
	P9: Personal information identified by FRT will be strictly protected and will not be leaked.

3.2.2. Traffic Police Investigation

At the same time, we designed a questionnaire for traffic police from law enforcement officials' perspectives to understand their attitude towards FRT (Table 2). All questions use a Likert five-level scale (from "strongly disagree = 1" to "strongly agree = 5").

3.3. Participants

In July 2019, the questionnaires were distributed to the public and traffic police in Fuzhou, China. All ethical norms and standards were strictly followed during the survey. The survey randomly selected 270 people from the public. The requirements were: (1) They are between 18 and 70 years old and use e-bikes more than three times a week; (2) have lived in Fuzhou, China for more than 6 months, and (3) are able to understand and answer the questionnaire. Among the 270 public questionnaires, we excluded 12 partially unanswered questionnaires, and the remaining 258 questionnaires were valid. In addition, 94 traffic police officers in Fuzhou, China, were randomly selected. Four partially unanswered questionnaires were excluded, and the remaining 90 questionnaires were valid. Therefore, in the subsequent data analysis, only the valid questionnaires of public and traffic police are discussed.

3.4. Questionnaire Data Reliability

The test of the reliability and validity of the data set indicates that the Cronbach's α coefficient of the two questionnaires is greater than 0.7 [52], indicating good reliability of the questionnaires. Furthermore, the KMO and Bartlett spherical tests also meet the requirements of being greater than 0.6 with significance. Thus, the two questionnaire datasets used in this research are credible and compelling.

3.5. Demographic Data

Table 3 shows the demographic information of 258 interviewees of all valid questionnaires. The statistical results showed that the percentages of males to females surveyed are almost equal. Most of the people surveyed are in two age groups: 18–36 and 37–54. In terms of education level, most of them are with college and undergraduate degrees (43.4%) or high school degrees (33.7%), while other degrees account for a relatively low proportion. In addition, most of the interviewees have driving licenses (62.0%).

Table 3. Demographic information of the public.

Demographic Variable	Category	Number	Percentage
Gender	Male	126	48.8%
	Female	132	51.2%
Age	18–36	120	46.5%
	37–54	114	44.2%
	>54	24	9.3%
Education level	Junior high school and below	34	13.2%
	Senior high school	87	33.7%
	College and undergraduate	112	43.4%
	Postgraduate and above	25	9.7%
Driving license	Have	160	62.0%
	Do not have	98	38.0%

4. Analysis and Results

4.1. Statistical Analysis of Public Questionnaire Data

Table 4 shows the average and standard deviation of each item in the public questionnaire. From data statistics, the scores of the three variables are all between 3.2 and 3.3. Variable A has a score of 3.238, indicating the public generally regards running a red light to be dangerous behavior (A1 ~ A6), but two items of variable A (A5 and A6) have lower scores. The scores of variable A reflect that the public's awareness of observing traffic rules is relatively poor. The score of variable B is 3.278, indicating they are more supportive of the effect of using FRT (B1 ~ B6). Among items of variable B, only the scores of B2 are lower than the average scores, and the results reflect that FRT is less effective in improving the safety awareness of the public. Variable C has the highest score of 3.297, indicating they generally view that FRT does not violate their privacy (C1 ~ C5). However, C5 has the lowest score of 3.019, suggesting that the public is less concerned about the right to know the use of FRT. Above all, the public generally supports monitoring red-light running behavior by using FRT without worrying about privacy invasion too much.

Table 4. The average and standard deviation of each item in the public questionnaire, $n = 258$.

Variable	Item	M	S.D.	Variable Average
A: Attitudes toward red-light running behavior of e-bikers	A1: Red-light running behavior of e-bikers has a negative impact on traffic.	3.225	0.960	3.238
	A2: Even if I have good riding skills, running a red light may be dangerous.	3.256	0.782	
	A3: Although running a red light can shorten the travel time, it is prone to accidents and is unworthy.	3.302	0.865	
	A4: Red-light running behavior is irresponsible to lives.	3.318	0.784	
	A5: More e-bikers are running red-light in China cities, and management needs to be strengthened.	3.140	0.806	
	A6: I am familiar with the traffic regulations related to e-bikes, and I ride per the regulations.	3.186	0.849	

Table 4. Cont.

Variable	Item	M	S.D.	Variable Average
B: Application effect of FRT	B1: FRT can significantly reduce the red-light running behavior of e-bikers.	3.516	0.852	3.278
	B2: The application of FRT helps to strengthen personal traffic safety awareness.	3.012	0.766	
	B3: Need to take specific measures to punish the identified behavior.	3.287	0.848	
	B4: The information of violators shown on the screen has a deterrent effect on the public.	3.360	0.798	
	B5: Most people will support and actively obey the application of FRT in traffic management.	3.264	0.856	
	B6: FRT is progress of technology and is worth promoting.	3.229	0.836	
C: Whether FRT violates privacy	C1: It is not a violation of personal privacy to show red-light running behavior on the screen.	3.376	0.947	3.297
	C2: The following information published on the screen is appropriate: the offender's image running a red light (the face is covered), the middle part of the name is concealed, and the ID number is concealed in the middle digits.	3.384	0.853	
	C3: Personal information identified by FRT will be strictly protected and will not be leaked.	3.349	0.901	
	C4: The application of FRT also needs to be regulated by improving relevant laws.	3.357	0.907	
	C5: When using FRT, the public's right to know needs to be guaranteed.	3.019	0.996	

The research uses Mann–Whitney U and Kruskal–Wallis tests. The Mann–Whitney U test is used to explore: (a) Whether different genders and driving license statuses resulted in differences in the three variables regarding attitudes toward red-light running behavior of e-bikers, (b) determine the application effect of FRT, and (c) whether FRT violates privacy. There is no significant difference in terms of public's gender and three variables. However, there is a significant difference in terms of driving license status and the public's attitudes toward red-light running behavior ($U = 998.000, p < 0.001$) and the application effect of FRT ($U = 2865.5, p < 0.001$). However, there is no significant difference in the public's attitudes toward privacy invasion.

The Kruskal–Wallis test is used to explore whether the public's different ages and education levels resulted in differences in the three variables of public attitudes. There is no significant difference in terms of age and the three variables. There is a significant difference in terms of education level and the three variables, i.e., attitudes toward red-light running behavior of e-bikers ($\chi^2(3) = 114.730, p < 0.001$), application effect of FRT ($\chi^2(3) = 103.534, p < 0.001$), whether FRT violates privacy ($\chi^2(3) = 90.292, p < 0.001$).

Then, we use Multiple Correspondence Analysis (MCA) to study the correspondence between the public's characteristics and the three variables. In order to meet MCA's data requirements, the scope of variables (A, B, and C), values, and the classification values are shown in Table 5. Figure 4 is the joint plot of the category points. Table 6 and Figure 5 present the discrimination measures of the variables. The MCA transforms all variables of the original data through the optimal scale transformation to obtain two dimensions (Dimension 1 and Dimension 2).

Table 5. The scores of variables (A, B, and C) and the values after classification.

Variable	Scope of Variable Value	Classification Values
A, B, C	(0, 1]	1
	(1, 2]	2
	(2, 3]	3
	(3, 4]	4
	(4, 5]	5

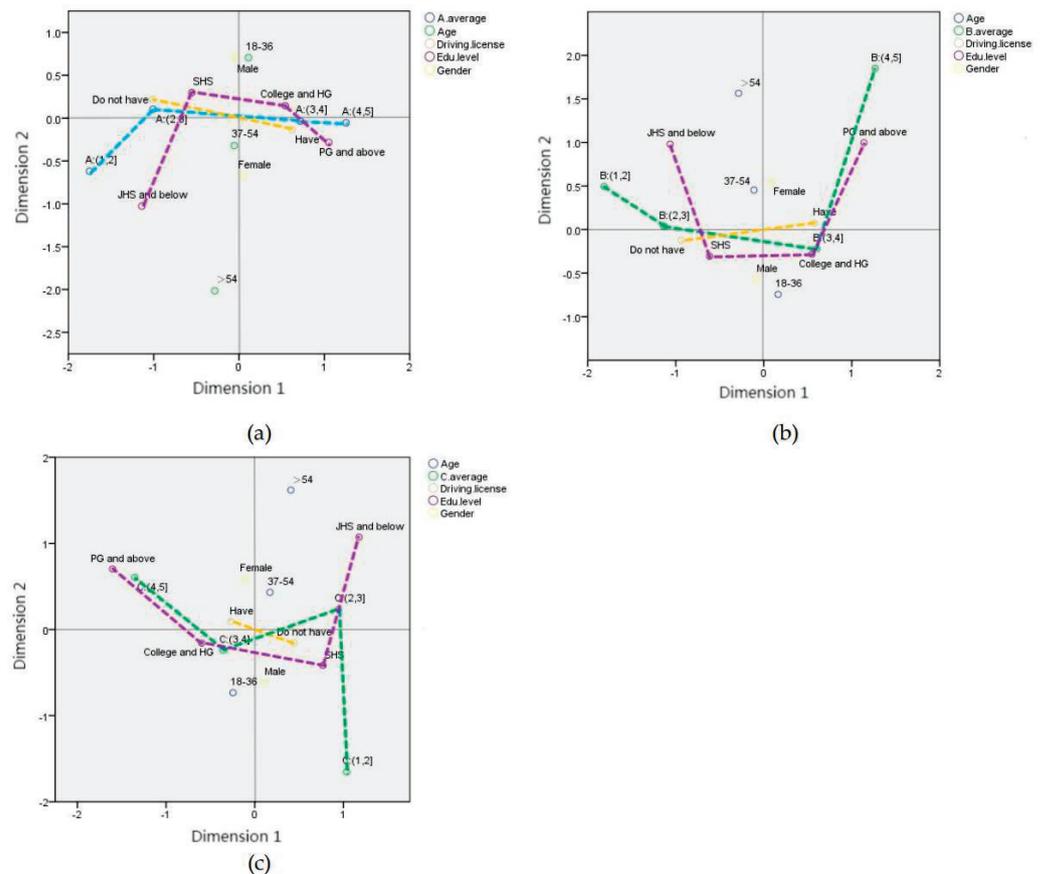


Figure 4. Joint plot of the category points. Correspondence between the variables: (a) The correspondence between the public’s characteristics and variable A; (b) the correspondence between the public’s characteristics and variable B; and (c) the correspondence between the public’s characteristics and variable C. The education level in the figures is abbreviated (JSH and below stands for junior high school and below; SHS stands for senior high school; college and HG stands for college and undergraduate; PG and above stands for postgraduate and above).

From Figures 4 and 5, and Table 6, we could observe the correspondence between the public’s characteristics and the three variables. In Figure 5 and Table 6, the public’s education level ($x = 0.509$, $y = 0.186$) and whether holding a driving license ($x = 0.630$, $y = 0.030$) are related to the value of variable A, and the two characteristics also possess greater explanatory power to variable B ($x = 0.533$, $y = 0.289$; $x = 0.541$, $y = 0.009$). However, variable C only related to the public’s education level ($x = 0.787$, $y = 0.269$).

Table 6. Discrimination measures of the variables.

Variable	Dimension		Variable	Dimension		Variable	Dimension	
	1	2		1	2		1	2
Variable A	0.889	0.017	Variable B	0.798	0.230	Variable C	0.679	0.196
Gender	0.003	0.483	Gender	0.007	0.310	Gender	0.011	0.362
Age	0.015	0.655	Age	0.025	0.577	Age	0.056	0.576
Educ. Level	0.509	0.186	Educ. Level	0.533	0.289	Educ. Level	0.787	0.269
Driving License	0.630	0.030	Driving License	0.541	0.009	Driving License	0.120	0.015

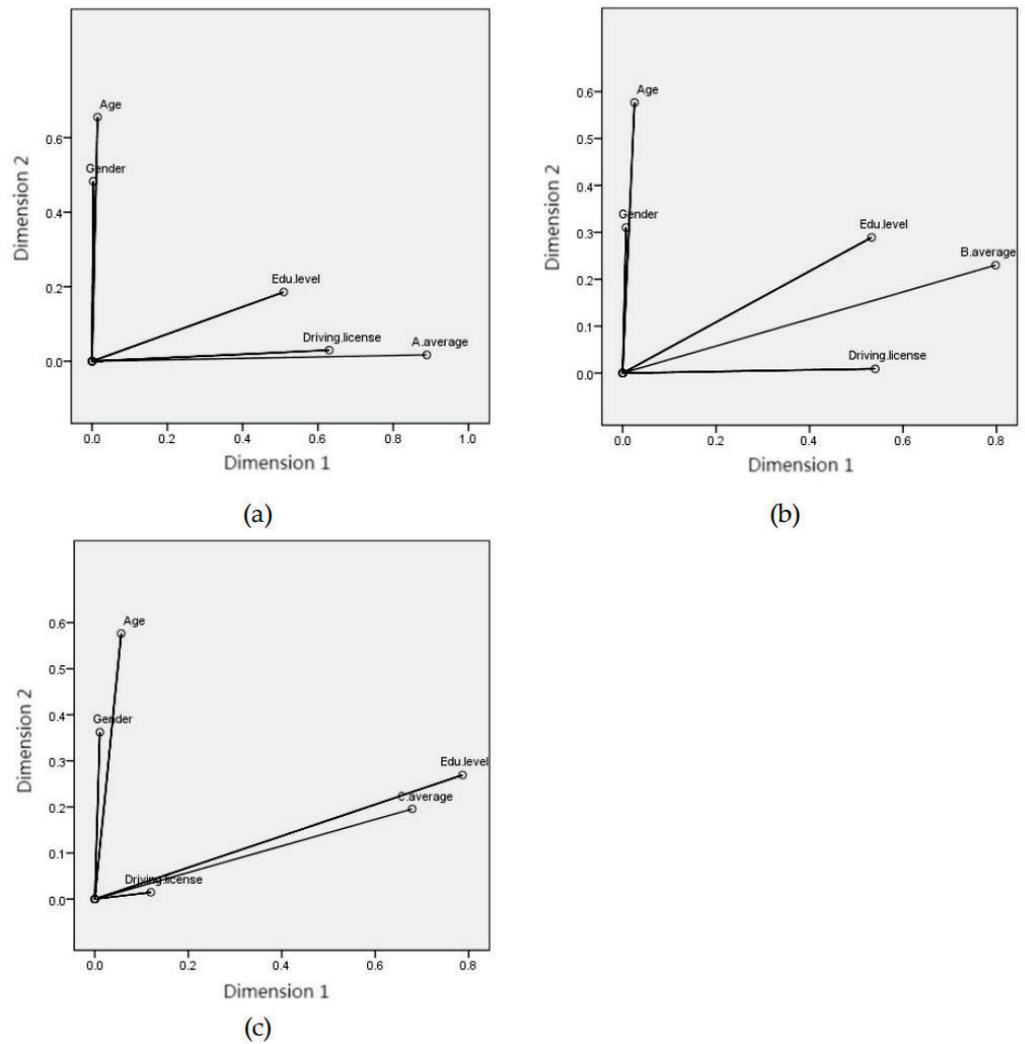


Figure 5. Discrimination measures of the variables: (a) Correspondence between the public’s characteristics and variable A. (b) Correspondence between the public’s characteristics and variable B. (c) Correspondence between the public’s characteristics and variable C.

Figure 4 presents the points from the various categories. Different variables that are close to the same direction and the area of the graph may be related. In Figure 4a,b, the category points of the education level and whether holding a driving license are close to the specific scores of the variables A, B, and C. Specifically, the points of higher education level are closer to the higher scores of variables A and B. For example, “PG and above” is close to “A: (4, 5]” and “B: (4, 5]” and “College and HG” is close to “A: (3, 4]” and “B: (3, 4]”.

In Figure 4c, “C: (1, 2]” and “JHS and below” have a long distance. The relationships between education level and variable C are similar to the situation in Figure 4a,b. In general, there are positive correlations between the driver’s education level and the three variables. Besides, the points of whether holding a driving license are close to the points of variables A and B. Specifically, “Do not have” is close to “A: (2, 3]” and “B: (2, 3]” and “Have” is close to “A: (3, 4]” and “B: (3, 4]”. However, whether holding a driving license does not have an obvious relationship with the points of variable C. The results indicate that people with a driving license get higher scores in variables A and B than those without a driving license. In short, whether holding a driving license positively affects variables A and B.

Table 7 and Figure 6 show the corresponding results between each of the three variables. In Figure 6b, the position of variable A ($x = 0.752, y = 0.464$) is close to that of variable B ($x = 0.724, y = 0.419$), and variable C ($x = 0.305, y = 0.300$) is farther than the two variables. In Figure 6a, the points position of the three variables with the same scores are also similar, except for the score (1, 2]. The results illustrate that the scores of the three variables have correspondence when the scores are higher. In other words, the scores of the three variables reach a higher level at the same time.

Table 7. Discrimination measures of the three variables.

Variable	Dimension	
	1	2
Variable A	0.752	0.464
Variable B	0.724	0.419
Variable C	0.305	0.300

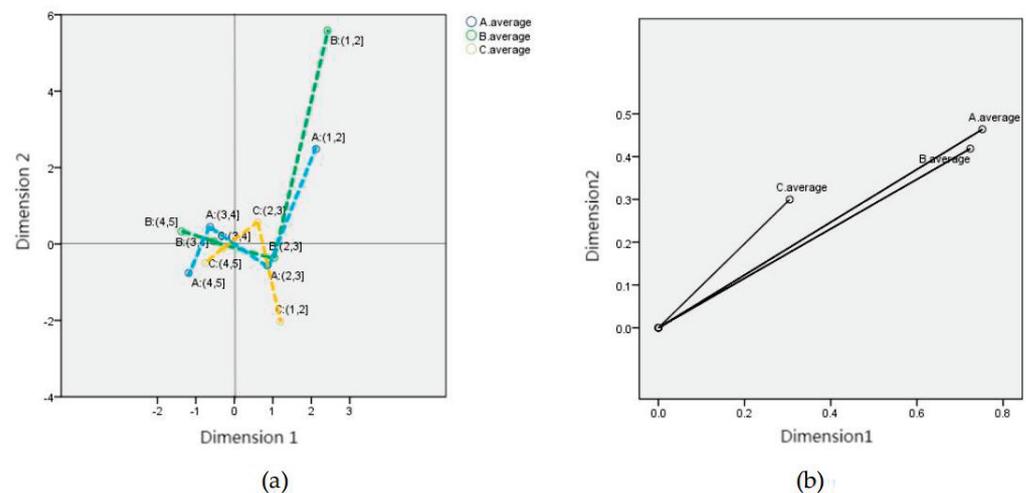


Figure 6. MCA results of the three variables: (a) Joint plot of the category points. (b) Discrimination measures of the three variables.

4.2. Statistical Analysis of Traffic Police Questionnaire Data

Table 8 shows the average and standard deviation of each item in the traffic police questionnaire. The results indicate that the average of most items is between 3.6 and 3.7, and the average of all items is 3.651. Among the traffic police questionnaire items, P7 has the lowest scores, which is like the public’s results. Thus, from the perspectives of the traffic police, FRT can not entirely improve the safety awareness of e-bikers. Nevertheless, overall, the traffic police have high support for the use of FRT.

Table 8. Average and standard deviation of each item in the traffic police questionnaire, $n = 90$.

Item	Item Average ¹	Standard Deviation
P1: I have a good understanding of applying FRT to manage the red-light running behavior of e-bikers.	3.656	0.791
P2: FRT can significantly reduce the red-light running behavior of e-bikers.	3.678	0.747
P3: FRT can reduce the management difficulty of traffic police.	3.656	0.733
P4: Need to take specific measures to punish the identified behavior.	3.700	0.781
P5: Require giving safety education to the identified violators.	3.600	0.712
P6: Most people will support and actively obey the application of FRT in traffic management.	3.689	0.709
P7: The application of FRT helps to strengthen personal traffic safety awareness.	3.589	0.759
P8: It is not a privacy violation to show red-light running behavior on the screen.	3.656	0.653
P9: Personal information identified by FRT will be strictly protected and will not be leaked.	3.633	0.767

¹ The average of all items is 3.651.

4.3. Comparative Analysis of Questionnaire Datasets

Extracting the same items in the two questionnaires, the Mann–Whitney U test was used to explore the attitude differences between the public and traffic police toward FRT. The test results in Table 9 illustrate the two groups differ in attitudes toward FRT ($U = 6958.500, p < 0.001$).

Table 9. Differences in attitudes toward FRT between the public and traffic police.

Variable	N	M	Mann-Whitney U	Wilcoxon W	Z	p
Public group	258	3.300	6958.500	40369.500	−5.691	<0.001
Traffic police group	90	3.657				

Table 10 shows the Mann–Whitney U test results of the same items by public and traffic police groups. There is a significant difference in the same items of the two groups, including B2/P7, B3/P4, B5/P6, C1/P8, and C3/P9, but there is no significant difference in B1/P2. The average of the same items also has a significant difference between the public and traffic police. Based on the results, it is concluded that there is a significant difference between public and traffic police attitudes toward FRT in general.

Table 10. Mann–Whitney U test results of the same items by public and traffic police groups.

Item	Public		Traffic Police		Mann–Whitney U	Wilcoxon W	Z	p
	M	S.D.	M	S.D.				
B1/P2	3.516	0.852	3.678	0.747	10592.000	44003.000	−1.329	0.184
B2/P7	3.012	0.766	3.589	0.763	7165.000	40576.000	−5.935	<0.001
B3/P4	3.287	0.848	3.700	0.785	8605.500	42016.500	−3.905	<0.001
B5/P6	3.264	0.855	3.689	0.713	8450.000	41861.000	−4.139	<0.001
C1/P8	3.376	0.947	3.656	0.656	10069.000	43480.000	−2.011	<0.05
C3/P9	3.349	0.901	3.633	0.771	9568.500	42979.500	−2.648	<0.05
Average	3.300	0.526	3.657	0.429	6958.500	40369.500	−5.691	<0.001

5. Discussion

5.1. Public Attitude toward FRT

In general, the public supports using FRT to manage the red-light running behavior of e-bikers. To understand which public characteristics are related to the attitudes toward FRT, we analyzed the correlation between the four individual characteristics and the three variables using the method of the Mann–Whitney U test, Kruskal–Wallis test, and MCA.

The results of the Kruskal–Wallis test and MCA indicate that members of the public with higher education levels are more resistant to the red-light running behavior of e-bikers ($\chi^2(3) = 114.730$, $p < 0.001$; Figure 4a). This finding is consistent with Wang et al. [53]. Under-educated e-bikers lack safety knowledge [53], and people with higher education backgrounds comprehend more traffic safety knowledge [39,45]. Members of the public with higher education levels are supportive towards the application effect of FRT ($\chi^2(3) = 103.534$, $p < 0.001$; Figure 4b), and they also show the trust of privacy protection ($\chi^2(3) = 90.292$, $p < 0.001$; Figure 4c). Because of more safety knowledge, people with higher education pay more attention to red-light running behavior and highly support FRT, perhaps due to their greater acceptance of new technologies. Moreover, their acceptance of FRT affects the trust of privacy protection.

Regarding whether or not holding a driving license affects the public's attitudes toward red-light running behavior and FRT ($U = 998.000$, $p < 0.001$; Figure 4a), people with driving licenses appeared to be more resistant to red-light running behavior. This is because e-bikers with driving licenses have lower perceived behavioral control and higher moral norm than those without driving licenses [23]. Moreover, e-bikers with driving licenses are also more supportive of the use of FRT ($U = 2865.5$, $p < 0.001$; Figure 4b). The strong correlation between the attitudes toward red-light running behavior and the application effect of FRT may indicate that people with driving licenses are more supportive of FRT.

5.2. Comparison of Public and Traffic Police Attitudes on FRT

The traffic police generally support the application of FRT (the average of all items is 3.651). Comparing the results of the same questions in public and the traffic police questionnaires shows that there are significant differences between the two groups in many items ($U = 6958.500$, $p < 0.001$), including “raise safety awareness, support for FRT applications, privacy issues of FRT, and information protection”. The support from the traffic police to FRT is significantly higher than that of the public.

For the traffic police, how to reduce the red-light running behavior of e-bikers has been a difficulty [54], and the appearance of FRT has solved the problem well [55]. Thus, reducing management difficulty may be the main reason why the traffic police support FRT. For example, Shenzhen started to use FRT in 2017, which had reduced the number of red light-running behavior at intersections from about 150 cases per hour to about 8 cases per hour within half a year [56]. Besides, FRT can realize real-time monitoring, which is difficult

for traffic police [57]. The application of FRT can protect traffic police from personal injury caused by violators [58].

5.3. Measures to Protect Public's Privacy

The public generally believes that FRT does not violate their privacy (the average score of variable C is 3.297), indicating that FRT is trustworthy for the public. However, the attitudes toward privacy violations differ in the education level of e-bikers ($\chi^2(3) = 90.292$, $p < 0.001$). Overall, highly-educated e-bikers have more confidence in privacy protection involved in FRT. Thus, under the circumstance that information can be completely protected, the public's concerns about privacy violation can be alleviated.

The privacy about personal data (e.g., facial images) consists of the right to control the access to and use of these data [59]. Regulation of the use of FRT is vital for privacy protection. In China, FRT used at signalized intersections ensures traffic safety and protects public interest. However, laws and regulations to standardize FRT use in China are still not complete. The official privacy-preserving policy could mitigate some of the privacy concerns which seem to be most troubling for the public, such as blurring people's faces, allowing officers to access only violation footage, and so on [40]. Besides, the public should be well informed about the facial recognition systems and should have consented to use these systems for the specific and justified purposes in question [59].

Updated technologies are conducive to privacy protection. For instance, FRT based on temporal features could preserve privacy [60]. A face recognition protocol, named PEEP is used to protect privacy by utilizing differential privacy [61]. The principal components of adversarial segmented image blocks can protect people's privacy and prevent the distinct face-related features of images from being easily extracted [62].

6. Conclusions

This research developed two questionnaires for the public and traffic police and analyzed their attitudes toward applying FRT and its effects and privacy issues. The results indicate that:

- (1) The public's attitudes toward FRT are related to two individual characteristics: Education level ($\chi^2(3) = 114.730$, $p < 0.001$; $\chi^2(3) = 103.534$, $p < 0.001$; $\chi^2(3) = 90.292$, $p < 0.001$) and driving license status ($U = 998.000$, $p < 0.001$; $U = 2865.5$, $p < 0.001$). The MCA results (Figure 4) show that a person with a higher education level or a driving license supports FRT.
- (2) There are significant differences between the public and traffic police in attitudes toward FRT ($U = 6958.500$, $p < 0.001$). Traffic police support FRT application more than the public, as the technology is conducive to reducing red-light running behavior of e-bikers and enforcement effort.
- (3) Based on data analysis, we make some suggestions about the application of FRT. Improving the education level and safety knowledge of the public helps to enhance their support to FRT under the circumstances that privacy is protected completely. There are also several suggestions about the use of FRT. For example, laws and regulations on applying FRT could protect the public from privacy invasion, updating the technology of FRT to protect information better, and that the public should consent before using these systems for the specific and justified purposes in question.
- (4) This research has some limitations. The questionnaire designed for the public is not comprehensive enough, and more detailed questions about privacy violations could be included in the future. In addition, this research is only based on the investigation of e-bikers in Fuzhou, China and e-bikers from other cities and other groups (e.g., pedestrians and bicyclists) may have different attitudes toward FRT.

This study investigates the application of FRT from the perspectives of the public and traffic police. We analyzed the rationality of the use of FRT in combination with the public's attitudes toward personal privacy invasion. Our research has a certain contribution to society and science. Based on the research results, we recommend that government

departments should carry out the following tasks for the public, including people with low education and without driving licenses: (1) Conduct safety education and training regularly, and (2) promote the fact that the final purpose of FRT application is to improve public security's awareness and publish information without violating privacy.

Author Contributions: Conceptualization, Y.Y. and J.L.; methodology, Y.Y. and J.L.; validation, Y.Y., J.L., and D.Y.; formal analysis, D.Y.; investigation, D.Y.; writing—original draft preparation, D.Y.; writing—review and editing, Y.Y., J.L., and D.Y.; visualization, J.L. and S.M.E.; supervision, J.L. and S.M.E.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: No applicable.

Informed Consent Statement: No applicable.

Data Availability Statement: The study did not report any data.

Acknowledgments: The authors are grateful to four anonymous reviewers for their thorough and most helpful comments. They also thank the Fujian Traffic Police Corps for funding this research and to Shihua Ni from the Fuzhou Traffic Police Detachment for his valuable support of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Weinert, J.; Ma, C.; Yang, X.; Cherry, C. Electric two-wheelers in China: Effect on travel behavior, mode shift, and user safety perceptions in a medium-sized city. *Transp. Res. Rec. J. Transp. Res. Board* **2007**, *43*, 62–68. [CrossRef]
- Tencent News. Available online: <https://new.qq.com/omn/20211203/20211203A02MNC00.html> (accessed on 3 December 2021).
- Lin, S.; He, M.; Tan, Y.; He, M. Comparison study on operating speeds of electric bicycles and bicycles: Experience from field investigation in Kunming, China. *Transp. Res. Rec. J. Transp. Res. Board* **2008**, *2048*, 52–59. [CrossRef]
- Ma, C.; Yang, D.; Zhou, J.; Feng, Z.; Yuan, Q. Risk riding behaviors of urban e-bikes: A literature review. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2308. [CrossRef] [PubMed]
- Wang, Z.; Neitzel, R.L.; Xue, X.; Zheng, W.; Jiang, G. Awareness, riding behaviors, and legislative attitudes toward electric bikes among two types of road users: An investigation in Tianjin, a municipality in China. *Traffic Inj. Prev.* **2019**, *20*, 72–78. [CrossRef]
- Bai, L.; Liu, P.; Chen, Y.; Zhang, X.; Wang, W. Comparative analysis of the effects of e-bikes and bicycles on safety of signalized intersections. *Transp. Res. Part D* **2013**, *20*, 48–54. [CrossRef]
- Schleinitz, K.; Petzoldt, T.; Kröling, S.; Gehlert, T.; Mach, S. (E-)Cyclists running the red light—The influence of bicycle type and infrastructure characteristics on red light violations. *Accid. Anal. Prev.* **2019**, *122*, 99–107. [CrossRef] [PubMed]
- Yang, X.; Abdel-Aty, M.; Huan, M.; Jia, B.; Peng, Y. The effects of traffic wardens on the red-light infringement behavior of vulnerable road users. *Transp. Res. Part F Psychol. Behav.* **2016**, *37*, 52–63. [CrossRef]
- Heilmann, S.; Petzoldt, T.; Gehlert, T.; Schleinitz, K. Traffic conflicts and their contextual factors when riding conventional vs. electric bicycles. *Transp. Res. Part F Traffic Psychol. Behav.* **2017**, *46*, 477–490. [CrossRef]
- Guo, Y.; Li, Z.; Wu, Y.; Xu, C. Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities. *Accid. Anal. Prev.* **2018**, *115*, 118–127. [CrossRef]
- Rodon, C.; Ragot-Court, I. Assessment of risky behaviours among E-bike users: A comparative study in Shanghai. *Transp. Res. Interdiscip. Perspect.* **2019**, *2*, 100042. [CrossRef]
- Huan, M.; Yang, X.; Jia, B. Crossing reliability of electric bike riders at urban intersections. *Math. Probl. Eng.* **2013**, *2013*, 831–842. [CrossRef]
- Haustein, S.; Møller, M. E-bike safety: Individual-level factors and incident characteristics. *J. Transp. Health* **2016**, *3*, 386–394. [CrossRef]
- Tang, T.; Wang, H.; Zhou, X.; Gong, H. Understanding electric bikers' red-light running behavior: Predictive utility of theory of planned behavior vs prototype willingness model. *J. Adv. Transp.* **2020**, *2020*, 1–13. [CrossRef]
- Tang, T.; Guo, Y.; Zhou, X.; Shi, Q.; Gong, H. Identifying psychological factors of e-bike riders' traffic rule violating intention and accident proneness in China. In Proceedings of the 20th COTA International Conference of Transportation Professionals, Xi'an, China, 14–16 August 2020.
- Lin, Y.; Wu, C. Traffic safety for electric bike riders in China—Attitudes, risk perception, and aberrant riding behaviors. *Transp. Res. Rec.* **2012**, *3*, 386–394. [CrossRef]
- Yang, X.; Huan, M.; Abdel-Aty, M.; Peng, Y.; Gao, Z. A hazard-based duration model for analyzing crossing behavior of cyclists and electric bike riders at signalized intersections. *Accid. Anal. Prev.* **2015**, *74*, 33–41. [CrossRef]

18. Du, W.; Yang, J.; Powis, B.; Zheng, X.; Ozanne-Smith, J.; Bilston, L.; He, J.; Ma, T.; Wang, X.; Wu, M. Epidemiological profile of hospitalised injuries among electric bicycle riders admitted to a rural hospital in Suzhou: A cross-sectional study. *Inj. Prev.* **2014**, *20*, 128–133. [[CrossRef](#)] [[PubMed](#)]
19. Zhang, Y.; Wu, C. The effects of sunshields on red-light running behavior of cyclists and electric bike riders. *Accid. Anal. Prev.* **2013**, *52*, 210–218. [[CrossRef](#)]
20. Guo, Y.; Liu, P.; Bai, L.; Xu, C.; Chen, J. Red light running behavior of electric bicycles at signalized intersections in China. *Transp. Res. Rec.* **2014**, *2468*, 28–37. [[CrossRef](#)]
21. Wu, C.; Yao, L.; Zhang, K. The red-light running behavior of electric bike riders and cyclists at urban intersections in China: An observational study. *Accid. Anal. Prev.* **2012**, *49*, 186–192. [[CrossRef](#)] [[PubMed](#)]
22. Gao, X.; Zhao, J.; Gao, H. Red-light running behavior of delivery-service e-cyclists based on survival analysis. *Traffic Inj. Prev.* **2020**, *21*, 558–562. [[CrossRef](#)]
23. Yang, H.; Liu, X.; Su, F.; Cherry, C.; Liu, Y.; Li, Y. Predicting e-bike users' intention to run the red light: An application and extension of the theory of planned behavior. *Transp. Res. Part F Psychol. Behav.* **2018**, *58*, 282–291. [[CrossRef](#)]
24. Gitelman, V.; Korchatov, A.; Elias, W. Speeds of young e-cyclists on urban streets and related risk factors: An observational study in Israel. *Safety* **2020**, *6*, 29. [[CrossRef](#)]
25. Siman-Tov, M.; Radomislensky, I.; Peleg, K.; Bahouth, H.; Becker, A.; Jeroukhimov, I.; Karawani, I.; Kessel, B.; Klein, Y.; Lin, G.; et al. A look at electric bike casualties: Do they differ from the mechanical bicycle? *J. Transp. Health* **2018**, *11*, 176–182. [[CrossRef](#)]
26. Ma, C.; Zhou, J.; Yang, D.; Fan, Y. Research on the relationship between the individual characteristics of electric bike riders and illegal speeding behavior: A questionnaire-based study. *Sustainability* **2020**, *12*, 799. [[CrossRef](#)]
27. Wang, Z.; Huang, S.; Wang, J.; Sulaj, D.; Hao, W.; Kuang, A. Risk factors affecting crash injury severity for different groups of e-bike riders: A classification tree-based logistic regression model. *J. Saf. Res.* **2021**, *76*, 176–183. [[CrossRef](#)]
28. Tang, T.; Guo, Y.; Zhou, X.; Labi, S.; Zhu, S. Understanding electric bike riders' intention to violate traffic rules and accident proneness in China. *Travel Behav. Soc.* **2021**, *23*, 25–38. [[CrossRef](#)]
29. Shaaban, K.; Pande, A. Evaluation of red-light camera enforcement using traffic violations. *J. Traffic Transp. Eng.* **2018**, *5*, 66–72. [[CrossRef](#)]
30. Brasil, R.H.; Machado, A. Automatic detection of red light running using vehicular cameras. *IEEE Lat. Am. Trans.* **2017**, *15*, 81–86. [[CrossRef](#)]
31. Zhang, M.; Wang, C.; Ji, Y. A Recognition method of pedestrians' running in the red light based on image. *J. Softw. Eng. Appl.* **2014**, *7*, 452–460. [[CrossRef](#)]
32. Dow, C.; Ngo, H.; Lee, L.; Lai, P.; Wang, K.; Bui, V. A crosswalk pedestrian recognition system by using deep learning and zebra-crossing recognition techniques. *Softw. Pract. Exp.* **2020**, *50*, 630–644. [[CrossRef](#)]
33. Semenova, E.I.; Kyarimova, S.D.; Kukartsev, V.V.; Leonteva, A.A.; Ogol, A.R.; Bondarev, A.S. Moving object recognition system on a pedestrian crossing. *J. Phys. Conf. Ser.* **2021**, *1728*, 012004. [[CrossRef](#)]
34. Bai, Y. Research on design and implementation of pedestrian red light automatic early warning system based on face recognition. *J. Comput. Methods Sci. Eng.* **2019**, *19*, 235–240. [[CrossRef](#)]
35. Wei, Y.; Wan, X.; Xu, H.; Shen, B. Design and implementation of pedestrians' red-light running evidence-collection system based on face tracking and recognition. *Mod. Electron. Tech.* **2018**, *41*, 36–39.
36. Zheng, J.; Peng, J. A novel pedestrian detection algorithm based on data fusion of face images. *Int. J. Distrib. Sens. Netw.* **2019**, *15*. [[CrossRef](#)]
37. Wang, Z.; Sun, X.; Zhang, X.; Han, T.; Gao, F. Algorithm improvement of pedestrians' red-light running snapshot system based on image recognition. *Lect. Notes Electr. Eng.* **2020**, *571*, 1718–1726. [[CrossRef](#)]
38. Cheng, Y.; Meng, H.; Lei, Y.; Tan, X. Research on privacy protection technology in face identity authentication system based on edge computing. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Industrial Design, Guangzhou, China, 28–30 May 2021.
39. Eriksson, L.; Bjørnskau, T. Acceptability of traffic safety measures with personal privacy implications. *Transp. Res. Part F Psychol. Behav.* **2012**, *15*, 333–347. [[CrossRef](#)]
40. Rosenfeld, A. Are drivers ready for traffic enforcement drones? *Accid. Anal. Prev.* **2019**, *122*, 199–206. [[CrossRef](#)] [[PubMed](#)]
41. McNally, B.; Bradley, G.L. Predicting young, novice drivers' intentions to install in-vehicle data recorders. *Transp. Res. Part F Psychol. Behav.* **2018**, *59*, 401–417. [[CrossRef](#)]
42. Lai, X.; Rau, P.P. Has facial recognition technology been misused? A user perception model of facial recognition scenarios. *Comput. Hum. Behav.* **2021**, *124*, 106894. [[CrossRef](#)]
43. Akhtar, Z.; Rattani, A. A face in any form: New challenges and opportunities for face recognition technology. *Computer* **2017**, *50*, 80–90. [[CrossRef](#)]
44. Liu, Y.; Yan, W.; Hu, B. Resistance to facial recognition payment in China: The influence of privacy-related factors. *Telecommun. Policy* **2021**, *45*, 102155. [[CrossRef](#)]
45. Broadbent, E.; Stafford, R.; MacDonald, B. Acceptance of healthcare robots for the older population: Review and future directions. *Int. J. Soc. Robot.* **2009**, *1*, 319–330. [[CrossRef](#)]
46. Mark, A.; Neil, S. Facial recognition technology in schools: Critical questions and concerns. *Learn. Media Technol.* **2020**, *45*, 115–128. [[CrossRef](#)]

47. Phillip, M.K.; Christine, L. Gardiner. Public support for body-worn cameras: The need for inclusion of more comprehensive measures of public concerns. *Crim. Justice Stud.* **2021**, *34*, 289–305. [CrossRef]
48. Ropiequet, J.L.; Doss, A.F.; Pennacchio, V.G. An introduction to the Illinois Biometric Information Privacy Act. *Bank. Financ. Serv. Policy Rep.* **2019**, *38*, 8–15.
49. Denise, A.; Konstantin, S.; Elizabeth, L. The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: A comparative analysis of US, EU, and UK regulatory frameworks. *AI Ethics* **2021**, 1–11. [CrossRef]
50. Zhang, J.; Zhao, Z. Legal Regulation on the application of face recognition technology. *J. Southwest Pet. Univ. Soc. Sci. Ed.* **2021**, *23*, 100–105.
51. Reason, J.; Manstead, A.; Stradling, S.; Baxter, J.; Campbell, K. Errors and violations on the roads: A real distinction? *Ergonomics* **1990**, *33*, 1315–1332. [CrossRef]
52. Nunnally, J.C.; Bernstein, I.H. Psychometric theory. *Psychom. Theory* **1994**, *3*, 248–292.
53. Wang, C.; Xu, C.; Xia, J.; Qian, Z. The effects of safety knowledge and psychological factors on self-reported risky driving behaviors including group violations for e-bike riders in China. *Transp. Res. Part F Psychol. Behav.* **2018**, *56*, 344–353. [CrossRef]
54. Xu, C.; Yu, X. Modeling and analysis of crash severity for electric bicycle. *Smart Innov. Syst. Technol.* **2019**, *98*, 367–375. [CrossRef]
55. Wei, J.; Li, L. Research on application of face recognition technology in road traffic management. *China Public Secur.* **2018**, 84–87.
56. China Daily. Available online: <https://baijiahao.baidu.com/s?id=1597330054767076132&wfr=spider&for=pc> (accessed on 10 April 2018).
57. Zheng, H.; Yuan, S.; Tian, S.; Yin, Y.; Lv, X. Research on face recognition in restricting pedestrian traffic regulations. *Inf. Rec. Mater.* **2018**, *19*, 95–96.
58. Tian, D. How to deal with the traffic accidents happened or arrested by the electron police system. Master Thesis, Guizhou University, Guiyang, China, 2008.
59. Marcus, S. Seumas Miller. The ethical application of biometric facial recognition technology. *AI Soc.* **2021**, 1–9. [CrossRef]
60. Leong, S.; Phan, R.; Baskaran, V.M.; Ooi, C. Privacy-preserving facial recognition based on temporal features. *Appl. Soft Comput.* **2020**, *96*, 106662. [CrossRef]
61. Chamikara, M.A.P.; Bertok, P.; Khalil, I.; Liu, D.; Camtepe, S. Privacy preserving face recognition utilizing differential privacy. *Comput. Secur.* **2020**, *97*, 101951. [CrossRef]
62. Yang, J.; Liu, J.; Wu, J. Facial image privacy protection based on principal components of adversarial segmented image blocks. *IEEE Access* **2020**, *8*, 103385–103394. [CrossRef]

Article

Identifying the Importance of Criteria for Passenger Choice of Sustainable Travel by Train Using ARTIW and IHAMCI Methods

Lijana Maskeliūnaitė * and Henrikas Sivilevičius

Department of Mobile Machinery and Railway Transport, Faculty of Transport Engineering, Vilnius Gediminas Technical University, LT-10105 Vilnius, Lithuania; henrikas.sivilevicius@vilniustech.lt

* Correspondence: lijana.maskeliunaite@vilniustech.lt

Abstract: Nowadays, travelers can use different modes of transport, and they usually choose the most suitable and reliable mode available. The choice of one mode of transport as an alternative to another is subjective. It is usually built upon passenger attitude toward the advantages and disadvantages of using a particular mode. This article proposes analytical methods for and research results on passenger choices for sustainable train journeys as an alternative to traveling by bus. The rank averages of all criteria and their normalized subjective weights were calculated with reference to new linear (ARTIW-L) and nonlinear (ARTIW-N) methods of average rank transformation into weight. A correlation between sub-criteria rank averages and normalized weights is presented, based on the minimum number of passengers required to be interviewed to provide reliable results. The average ranks assigned by passengers to the evaluation sub-criteria and their global weights were used for determining and describing the most and least important key criteria by applying the inverse hierarchy for assessment of main criteria importance (IHAMCI) method. The analysis shows that the most important key criterion belonged to the sub-criteria characterizing economy, while the less important key criteria included ride comfort. The least important key criteria described safety and environmental protection, whose normalized subjective overall weights were the lowest. Rail transport authorities and companies involved in transporting passengers can make this mode of transport more attractive to people by giving priority to improving the services they provide to passengers.

Citation: Maskeliūnaitė, L.; Sivilevičius, H. Identifying the Importance of Criteria for Passenger Choice of Sustainable Travel by Train Using ARTIW and IHAMCI Methods. *Appl. Sci.* **2021**, *11*, 11503. <https://doi.org/10.3390/app112311503>

Academic Editor:
Anselme Muzirafuti

Received: 27 October 2021
Accepted: 2 December 2021
Published: 4 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: railway transport; passengers; sustainable travel; ARTIW method; IHAMCI method; MCDM

1. Introduction

The sustainable development of the European Union (EU) is strongly influenced by transport, which is polluting from an environmental point of view. Transport infrastructure accessibility criteria and accessibility distance have a positive effect on sustainable development [1]. In order to mitigate transport sector greenhouse gas emissions, it is necessary to assess the efficiency of transport policy [2]. One of the most effective ways of improving the sustainability of the transport sector is the choice to use a less-polluting mode of transport. Passenger satisfaction is an important factor in choosing a mode of transport to travel in municipalities and especially in big cities [3].

Mobility is crucial for the development of a country's internal market and for maintaining the desired life quality of citizens, as it is important for people to exercise their freedom to travel. This means more frequent travel by bus, rail, and air [4]. Viable options can only be available through better integration of modal networks, which means that airports, railway stations, and metro and bus stations should be increasingly interconnected and transformed into multimodal passenger transport platforms [4].

There will be more possibilities for passengers to choose a particular mode of transport when various systems of transportation become more closely integrated. Experience has

shown that passengers usually choose the mode of transport that best suits their personal needs, habits and understanding of the expected quality of traveling. The mode of transport that is the most acceptable to a passenger can be determined by developing a system of criteria and calculating their weights using multi-criteria decision-making (MCDM) methods [5–9]. Passenger loyalty development increases the profits of airline, rail, and road transport companies. This is undoubtedly important for the success of the enterprise [10].

Traveling allows us to learn more about the world. However, it brings about some inconveniences and discomfort. People who travel by bus for long distances spend much time on a bus and become tired.

For many people, it causes stress, unless they are inveterate travelers. Therefore, when planning a trip, people should learn more about particular transport modes, trip duration, possibilities for rest, and the cost of the trip. Various transport modes (by bus, train or airplane) have advantages and disadvantages [11]. It is important to inform urban public transport passengers on the estimated time of arrival, as that influences the outlook for traveling by bus [12]. Train traffic has increased over the last decade and is likely to continue to do so, as more passengers and freight are transported by train rather than by car. This will reduce energy consumption and pollutant emissions [13].

Travel by train is still popular worldwide. The railways, rolling stock, and services provided to passengers as well as safety are being constantly improved, while the environmental pollution is being reduced. This helps rail transport to compete with other modes of transport. Matuška [14] examined the accessibility of rail transport by providing ways to assess the accessibility of railway infrastructure and trains. He applied a two-step model to assess the availability of departure halls. Train services are barrier-free for passengers, especially those traveling long distances, but they are still less accessible to disabled passengers in the suburbs and outer regions.

2. Literature Review

A U.S. interregional travel study focused on regional long-distance (100 to 500 miles) passenger transportation. Consideration was put on travel by car, plane, motor bus, and train. Attention was paid to high-speed and conventional passenger trains [15]. In the freight and passenger transport sector in the Slovak Republic, competition in the rail transport market differs. The total number of passenger-kilometers has increased by 12 percent due to an increase in free tickets for students and retired people [16].

Studies aimed at evaluating technical parameters of roads and rolling stocks and improving their interaction and safety of travel as well as risk factors and accidents have been performed. Rail transport must ensure a high level of reliability and safety of travel. Since the wheel is one of the main subsystems of the railway vehicle, it can make a significant contribution to the reliability and safety of the train. One of the main measures to meet the requirements is to implement proper maintenance procedures. The quality of railway tracks has an impact on train safety and passenger comfort. In practice, the quality of railway tracks is measured by a track recording vehicle, which measures seven key geometrical parameters of tracks. Traditionally, track gauge, vertical and lateral alignments, and cross-level (angular variation in the track section, i.e., cant or superelevation) are measured [17]. Xin et al. [18] presented a model for predicting railway track damage. Railway violations have been proved to be the most important determinants of train safety and passenger travel comfort [19]. Unauthorized changes in railway track geometry can have a negative impact on train traffic safety [20].

Compared to other modes of transport, rail travel is safe. Afazov et al. [21] provided a more detailed understanding of modern modeling techniques that can be used in the design of railway vehicles. Lin et al. [22] presented a probabilistic risk assessment methodology for analyzing adjacent-track accident risk. Research in [23] was conducted to establish the safety of glazing systems for passenger railway equipment. Kovandová and Válka [24] investigated traffic safety as a major social problem related to accidents on railways and at track crossings.

Some studies have been performed regarding the possibilities of increasing the power and frequency of rail vehicles. Xu et al. [25] presented strategies to increase train frequency and rail capacity that would be helpful to metro dispatchers.

The model presented by Sun and Schönefeld [26] made it possible to identify gaps in the capacity of the train network and to assess the impact of schedule adjustments on passenger route choice. Xiang and Zhu [27] proposed multifunctional optimization to improve the economic performance of heavy rail.

Increasing the volume of passenger transportation by rail and its effectiveness in terms of expenses is a priority task. Determining the market shares and new offers for passenger service requires the study of the specific regional features of demand and the relevant state of transport services provided by different types of passenger transportation companies. Makarova and Muktepavel [28] presented a system of calculation and analytical indicators for analyzing regional passenger traffic. It allows specialists in this area to investigate passenger flow tendencies, determining the demand and cost of transport services as well as their dispersion across the region and to perform a comparative analysis of internal passenger traffic and the total amount of passenger traffic in the network. In practice this analytical information can be used to determine the optimal passenger train length, to assess the profitability of introducing local trains into operation, to define the optimal number of stops, and to calculate the amount of passenger cars needed to satisfy the demand. The route in the model that was offered by Tang et al. [29] is divided into sections that can be independently updated, and the target function is expressed in terms of minimizing driving time. This model can help to quickly and efficiently develop a strategic plan to reduce running time in passenger rail corridors.

Liao and Liu [30] used microscopic simulation models to investigate passenger behavior in the non-payment area.

Allen and Levinson [31] studied passenger train schedules and their average speed on North American railways in the period between 1965 and 2015. These train traffic parameters were used because their values were easy to obtain.

Passenger transportation systems are being upgraded and expanded around the world. Experimental studies have been carried out to improve the quality and efficiency of high-speed train services. Lee et al. [32] investigated the aerodynamic properties of a high-speed train pantograph and made suggestions for their improvement. Ou et al. [33] investigated the reasons for the development of a comprehensive railway system in China and its impact on the development of intercity passenger railways. Teixeira and Prodan [34] reviewed railway taxation systems and their development in 2007–2012. They assessed the importance of taxation for the single European railway market.

Oh et al. [35] conducted an analysis of covariance and analysis of regression and identified the effect of wagon door width on passenger boarding time on Korean city railways. Holloway et al. [36] presented the results obtained from an experiment in determining the time required for passengers to board or deboard a train. They found out that steps had little or no effect on the time to board a train for younger luggage-carrying people, while senior passengers, on the contrary, needed more time.

Multi-criteria decision making (MCDM) methods are used to solve problems related to the use of different modes of transport [37–45]. Chen et al. [46] investigated the process of rail passenger transfer at large terminals by comparing different alternatives. Stoilova [47] presented a combination of multi-criteria models for rating railway passenger transport development.

The MCDM methods used for modeling and evaluating the quality of passenger transportation on an international route allow researchers to identify the opinions of passengers, staff and the administration of the train about the weights (significances) of various criteria describing this complicated process [48,49]. Improving various aspects of this process can help rail transport to compete with other modes of transport more effectively. Based on an investigation of existing market research practices, three main approaches were identified for a comparative analysis of the influence of different parameters of transport

services on passenger satisfaction in order to define priority directions for implementation of administrative decisions concerning service quality. They included the method of obtaining priorities from passengers, as well as calculations based on mechanisms of correlation and regression analysis using the method of smallest squares and calculations based on the application of various nonparametric methods of statistics. Methodical and practical approaches based on modeling and intended for identifying promising areas for improving the quality of public services have also been presented [50]. The comparison and visualization of the results of assessing the impact of different transport service parameters on the overall quality of service by the methods of ordinal logistic regression were also presented. Customer perceptions of the quality of service provided by the operator and the level satisfaction are key parameters to monitor performance. Kesten and Ögüt [51] provided a practical way to monitor the functioning of the public transport system as a result of passenger evaluation. The passenger-oriented efficiency index was developed, and it employed 22 indicators and 6 different tools. Time, cost, ease of transfer, security and quality of service were assessed.

The aim of this study is to provide a set of criteria and show the advantages of rail transport compared to road transport (buses). By using MCDM methods, we determined the mean ranks, global, and overall weights of these criteria, employed a reverse hierarchy model and correlation of values. Finally, we calculated indicators showing the consistency of passenger views.

3. The Methods of the Average Rank Transformation into Weight (ARTIW-L and ARTIW-N)

The weights of the evaluation criteria (sub-criteria or key criteria) largely determine the evaluation result. In practice, the subjective weights assigned by experts or respondents to the considered criteria are commonly used. These weights present the judgments of highly qualified experts with long-term practical experience and theoretical knowledge in the considered field [52,53]. Passengers themselves make decisions about the mode of transport they choose for travel and, therefore, are experts themselves. However, because of their low competence, they should be referred to as respondents answering the survey questions rather than experts.

Most of the widely known and used methods for evaluating the weights of multiple criteria (factors) are based on experts' judgments. These methods embrace a thorough problem analysis by experts, the organization of this process as well as quantitative evaluation of decisions, and the arrangement of the obtained results. Therefore, the problem of practical determination of the accurate weights of the considered criteria arises. The subjective weights of the evaluation criteria can also be found from the ranks assigned to these criteria by experts. The estimates (judgments) of various experts differ considerably, often being inconsistent, which implies that the obtained weights (significances) of the criteria as well as their order of preference may be different.

The result of the experts' evaluation largely depends on their qualifications and experience in assessing the objectives, as well as their responsibility for providing the appropriate estimates of criterion significance and readiness to take part in the experimental study. The judgments of specialists and respondents about the relative significance of the criteria and their arrangement by order of priority (preference) often differ; therefore, the ranks and weights expressed in terms of the average values of the experts' estimates can be used in multi-criteria evaluation only if the consistency of the estimates has been proved. The consistency of the estimates given by a group of experts in terms of ranks is based on the idea of compactness.

In the case of expert evaluation, the average estimate obtained from a group of experts (respondents) is the problem solution (a result of decision-making) only when the judgments of all the experts are consistent. If a decision should be made based on the average estimate of the experts or respondents, the level of consistency of the experts' estimates is described by the concordance coefficient W . To determine the concordance coefficient W , the ranks of the evaluation criteria assigned by the experts or respondents

are required. If their estimates are given in other units (for example, in points), they should be ranked.

The consistency of the weights of the criteria describing an object and the estimates provided by experts are usually determined by using the analytic hierarchy process (AHP) approach [54–56]. The consistency of group evaluation results is determined by using the method of rank correlation [57,58].

The AHP approach is rather complicated [59,60] because not all of the experts can properly fill in the questionnaire (i.e., a pairwise comparison matrix), which would allow them to calculate the weights of the criteria and the consistency ratio (C.R.). The AHP method also allows for determining the consistency ratio of each expert’s estimates, which should be smaller than $C.R. \leq 0.1$. Moreover, the AHP approach is used for calculating each criterion’s eigenvector, i.e., the procedure of normalizing the geometric mean of the rows. The maximal eigenvalue (λ_{max}), the consistency index (C.I.) and the consistency ratio (C.R.) should also be calculated [41,53,55,61,62].

Experts usually assign the ranks R_{ij} to the criteria by arranging them according to their significance and giving them the appropriate numbers. This method of determining the criteria weights is logical; however, its accuracy is low. Therefore, it can be used only at the initial stage of analysis. Using more accurate and complicated methods still requires preliminary ranking of the criteria.

The weights ω_j of the ranked indicators (criteria) can be determined by applying different methods (algorithms) that do not have theoretical advantages over one another. However, the general principle of all algorithms is the same: the most important criterion is assigned the highest weight. The values of the weights ω_j must correspond to the criteria ranks (lower rank–higher weight). The sum of the weights ω_j of all the criteria describing the research object must be equal to one, i.e., weights must be normalized.

It is convenient to transform the ranks assigned to the criteria by a group of experts into weights by using the new ARTIW-L and ARTIW-N methods, whose sequence of operations and calculations is given below (see Figure 1). The average rank \bar{R}_j , representing the ranks assigned by all i -th experts ($i = 1, 2, \dots, m$) is calculated for each j -th criterion ($j = 1, 2, \dots, n$) by the formula:

$$\bar{R}_j = \frac{\sum_{i=1}^m R_{ij}}{m}. \tag{1}$$

The more important the criterion, the smaller its average rank \bar{R}_j . In practice, it is more convenient to use the estimates of the criteria significance, whose numerical values show higher importance. For this purpose, the normalized weights of criteria j , expressing relative importance, are used.

Significances (weights) of the evaluation criteria of an object can be determined in the process of their normalization (setting their sum equal to one) by transforming the average ranks into weights (the ARTIW method). This method was first proposed in 2011 [63]; however, at that time it was not called ARTIW. A relative weight ω_j of the criterion is calculated as follows:

$$\omega_j = \frac{(n + 1) - \bar{R}_j}{\sum_{j=1}^n \bar{R}_j}, \tag{2}$$

where n is the number of criteria describing the quality of the considered object, \bar{R}_j is the average j -th criterion rank calculated by Equation (1).

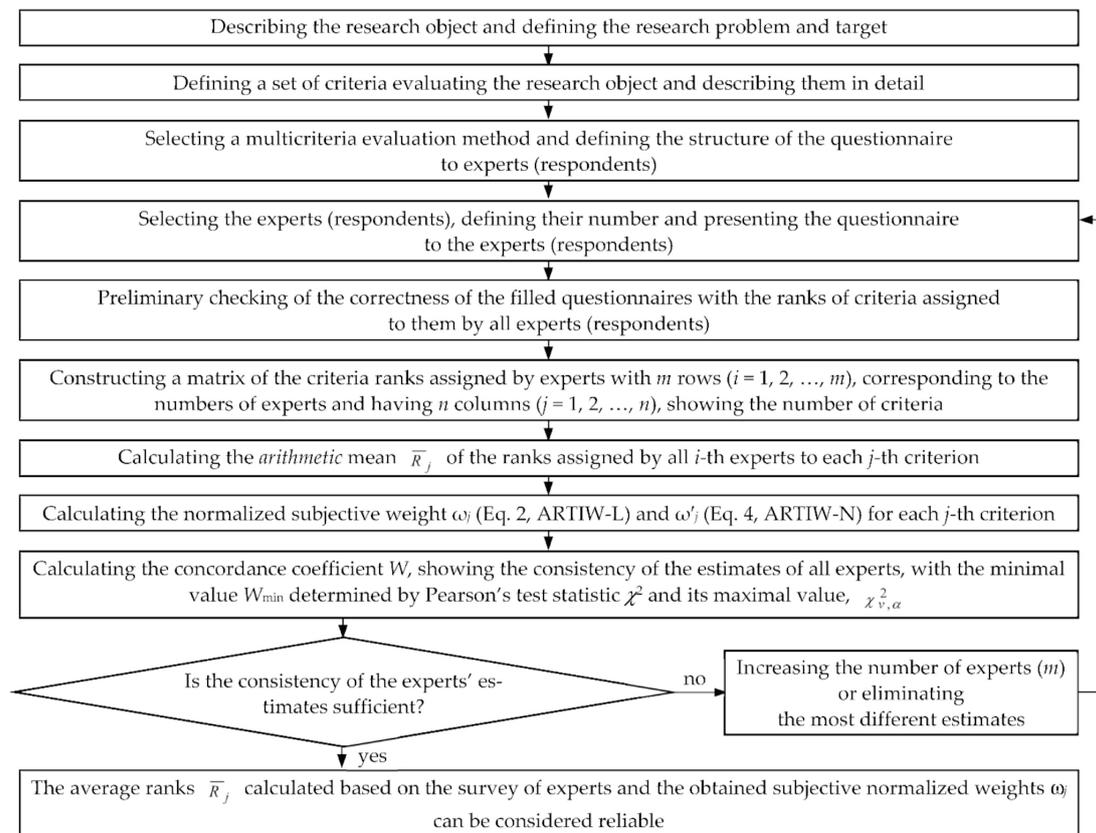


Figure 1. The algorithm of ARTIW-L and ARTIW-N methods of the average rank transformation into weight used for calculating normalized subjective weights of the criteria describing the research object.

The normalized weights ω_j of the j -th criteria calculated according to formula (2) have a linear inverse correlation (functional) relationship with the rank averages (average ranks) \bar{R}_j of these criteria calculated according to Formula (1). Therefore, this method is called the average rank transformation into weight-linear (ARTIW-L).

The normalized weights ω'_j of the criteria can be calculated by using another method of transforming rank averages into weights. The criterion weights calculated according to Formulas (3) and (4) are related to criteria rank averages \bar{R}_j by a non-linear inverse correlation (functional) dependence. Therefore, this method is called the average rank transformation into weight-non-linear (ARTIW-N).

Using the ARTIW-N method, the ratio of the min \bar{R}_j of the most important criterion (with the lowest average ranks \bar{R}_j) to the average of the ranks \bar{R}_j of all other j -th criteria is initially calculated:

$$u_j = \frac{\min_j \bar{R}_j}{\bar{R}_j}, \tag{3}$$

After normalizing the values u_j for each criterion, their subjective significances ω'_j are calculated:

$$\omega'_j = \frac{u_j}{\sum_{j=1}^n u_j}. \tag{4}$$

Neither of these two methods (ARTIW-L and ARTIW-N) can be considered more accurate than the other, and neither of them can be looked at as the reference method.

The average of the weights $\bar{\omega}_j$ calculated for each criterion by these two methods can be considered as the result of the task calculation:

$$\bar{\omega}_j = \frac{\omega_j + \omega'_j}{2}. \tag{5}$$

The consistency of expert group estimates is determined by the concordance coefficient W .

The concordance coefficient W in the absence of tied ranks is expressed in terms of the relationship between the obtained sum S and the respective largest sum S_{\max} [58]:

$$W = \frac{12S}{m^2n(n^2 - 1)} = \frac{12S}{m^2(n^3 - n)}. \tag{6}$$

When the estimates provided by the experts or respondents are in agreement, the Kendall coefficient of concordance, W , is about one. When the estimates differ considerably, the value of W is close to zero.

The deviations of the ranks R_{ij} of each criterion from the sum of squares of the average rank can be calculated as follows:

$$S = \sum_{j=1}^n \left[\sum_{i=1}^m R_{ij} - \frac{1}{2}m(n+1) \right]^2, \tag{7}$$

where n is the number of criteria ($j = 1, 2, \dots, n$), m is the number of experts (respondents) ($i = 1, 2, \dots, m$).

The random value for S was calculated by Equation (7), adding the squared values of all the criteria given in parentheses.

The concordance coefficient may be used in practice when its ultimate value, showing when expert estimates can be considered consistent, has been found. Kendall [57] has shown that if the number of criteria is $n > 7$, the significance of the concordance coefficient W can be determined using Pearson's chi-squared test statistic χ^2 . The random value

$$\chi^2 = Wm(n - 1) = \frac{12S}{mn(n + 1)}, \tag{8}$$

is distributed according to χ^2 distribution, with the degree of freedom $\nu = n - 1$.

When the number of the compared criteria n ranges from 3 to 7, the distribution χ^2 cannot be used in all cases because sometimes the critical value of $\chi^2_{\nu, \alpha}$ may be larger than the calculated value (even though the consistency of the estimates is still sufficiently high). In this case, the probability tables of the concordance coefficient or the tables of critical values S (with $3 \leq n \leq 7$) can be used [64].

The smallest value of the concordance coefficient W_{\min} allowing the authors to consider that the estimates of m experts of the quality of the research object based on n criteria, with the assigned (required) significance level α and degree of freedom $\nu = n - 1$, are consistent, can be calculated as follows [63]:

$$W_{\min} = \frac{\chi^2_{\nu, \alpha}}{m(n - 1)}, \tag{9}$$

where $\chi^2_{\nu, \alpha}$ is the critical Pearson's statistic found in the table [65], assuming the degree of freedom $\nu = n - 1$ and the significance level α .

The quality of the research object is evaluated by the additive mathematical model used for calculating its comprehensive quality index, which allows for describing the quality of the object by a single number. It also allows for comparing it with the quality of other similar objects, and the coefficients $\bar{\omega}_j$ of the normalized criteria weights (rather than the average criteria ranks \bar{R}_j , which cannot show how one criterion is more important than another) are used.

The weights of the criteria describing the research object (the selection of rail transport rather than road transport by passengers) can be calculated by using a very popular but complicated approach referred to as the analytic hierarchy process (AHP) offered by T. L. Saaty [54–56,66]. Passengers are not highly qualified experts and, therefore, can hardly fill in a pairwise comparison matrix properly, particularly if the number of the criteria compared is large. This number may be more than nine (e.g., fifteen) criteria. In the study [67], passengers completed pairwise comparison matrices with 32 criteria, 22 of which were rejected because their C.R. (consistency ratio) was greater than 0.1. Only 10 matrices were applicable for the study on the quality of passenger transport by train. Therefore, it is not rational to apply the AHP method in passenger interviews. Not every passenger can complete the pairwise comparison matrix properly. The AHP method can only be applied to interview highly qualified experts.

The objective weights of the criteria and sub-criteria can be calculated by using the entropy method [61,68] as well as the new IDOCRIW method [52], which combines (integrates) the entropy and the criterion impact loss (CILOS) methods.

4. The Structure of the Hierarchy Model, the Questionnaire, and the Respondents

The famous American writer Mark Twain wrote that “Travel is fatal to prejudice, bigotry, and narrow-mindedness”. On the other hand, people become tired when traveling and, therefore, the choice of an appropriate mode of transport is very important. Now, there is a wide choice of modes of travel, which include pedestrian traveling, cycling and traveling by automobile, by bus as well as by rail, air or water transport. A passenger decides which mode of transport is most safe and comfortable for travel. The criteria determining the choice of a particular mode of transport can be identified when a set of the evaluation criteria is defined and a certain number of passengers are surveyed. The passengers, who chose a particular mode of transport (e.g., rail transport) as an alternative to another mode of transport, assign the ranks to the considered criteria. All the criteria describing a particular mode of transport have some advantages over the criteria describing another means of transport.

The significance of hierarchically unstructured criteria or sub-criteria is identified using a two-level model (Figure 2a). In a three-level hierarchy model, which is used in multiple criteria decision-making, the goal of the study is given first, then the criteria are presented, and, finally, sub-criteria are provided [69–75] (see Figure 2b). In this work, the inverse (not classical) hierarchy model (see Figure 2c) was used for determining the ranks of the criteria and their weights. Level 1 of the model presents the goal, Level 2 the factors and sub-criteria, and Level 3 provides a group of factors and criteria. First, the average ranks and global weights of particular sub-criteria were calculated without their division into groups (Figure 2a). Then, they were grouped into three groups, and the reduced weights of the criteria groups were calculated, considering the fact that each group had a different number of criteria (Figure 2c).

The study was based on a survey of passengers traveling from Vilnius (Lithuania) to Moscow (Russia) and back to Vilnius. There is a regular rail and road service between the capital of Lithuania, one of the Baltic states (and a member-state of the EU) and the capital of Russia (Moscow). Therefore, passengers can choose between the two modes of transport in covering a distance of 944 km between these cities.

A set of criteria (sub-criteria) was defined to determine their influence on passengers' choice to travel by train rather than by bus. For this purpose, passengers had to rank the considered criteria according to their importance for their choice of this mode of transport. The following sub-criteria were included in the questionnaire presented to the passengers (Figure 2a):

- A. A trip by train is safer than a trip by bus (according to the statistical data).
- B. Trains take preference over buses at road crossings, which helps them to cut the time of stops.
- C. Trains pollute the environment less heavily.

- D. A trip by train is cheaper than a trip by bus.
- E. Passengers can lie down on berths, rest and change clothes in the compartment.
- F. The compartment of a passenger car is a closed space (ensuring fewer disturbances from other passengers),
- G. There is a possibility for passengers to use WC on the train and smoke on the platform of a passenger car at any time.
- H. Traveling by train does not depend on weather conditions.
- I. There is less noticeable rocking and vibration than in a passenger car.
- J. A dining car is available for passengers.
- K. There is freedom of movement on the train.
- L. Trains have fewer stops than buses and are rarely overdue.
- M. Trains are traveling day and night.
- N. There is a possibility for passengers to order food or newspapers and magazines to the compartment.
- O. There is a simpler border control for passengers on the train (they do not need to leave the cars and their luggage is with them).

A questionnaire for ranking the sub-criteria by using the method of correlation was prepared by the authors. It was also translated into Russian language. An anonymous survey was carried out, with 52 questionnaires presented to passengers on the Vilnius–Moscow–Vilnius train. About 48% of the trip, which lasts for 14 h and 05 min (944 km), took place during the night. Respondent characteristics are presented in Table 1.

The same passengers completed questionnaires and assessed the sub-criteria that determine the choice to travel by train as an alternative to aircraft. The results of this research were published in the article [11].

Table 1. Details of 52 respondents who gave the judgement on ranks.

Items	Descriptions
Gender	28 male respondents, 24 female respondents
Age	Average age is 46 years
Citizenship	Lithuanian (24), Ukrainian (1), Russian (26), passenger with dual Russian–Lithuanian citizenship took part (1)
Education	Most respondents (46) had higher education; 6 respondents had secondary education
Aim (goal) of trip	Business (21), tourism (15), visiting relatives and friends (10), medical treatment (5), research (1)

The number of respondents (52) was three times that of the criteria (sub-criteria) (15). Therefore, it was sufficient because $m \geq n$. A description of 15 sub-criteria was presented in the questionnaire, and the respondents assigned different ranks to them (all the ranks had different assignable values).

When applying expert research methods to assess the significance of criteria, there is a problem of determining the required (necessary) minimum number of experts. In practice, the mathematically unsound provision (principle) that the number of experts must be equal to or greater than the number of criteria is often observed. There is another common position that is often applied in practice, which maintains that the amount of data required for studies $m \geq 30$ is also not substantiated, because in some cases the number m is sufficient (if the group range is small), while in other cases it is too small. The credibility of expert group assessments depends on the level of knowledge of individual experts and their number. Having assumed that the experts are accurate assessors, it can be stated that as their number increases, the reliability of the expertise of the whole group of experts

(average of the opinion estimate) also increases. The minimum number of experts to be interviewed m_{\min} can be calculated according to the sample size formula [69]:

$$m_{\min} = \frac{t^2 \sigma_{R_j}^2}{\Delta_j^2}, \tag{10}$$

where t is the value of t (Students) distribution, which depends on the probability taken to assess the importance of the criterion in deciding to go by train as an alternative to the bus. When the probability $P = 95\%$ (significance level $\alpha = 0.05$ for one-sided test), $t = 1.96$; σ_{R_j} —standard deviation of the ranks R_{ij} of the evaluated j -th criterion; Δ_j is the absolute error of passengers (respondents) rank values j -th criterion, indicating the accuracy of the survey results.

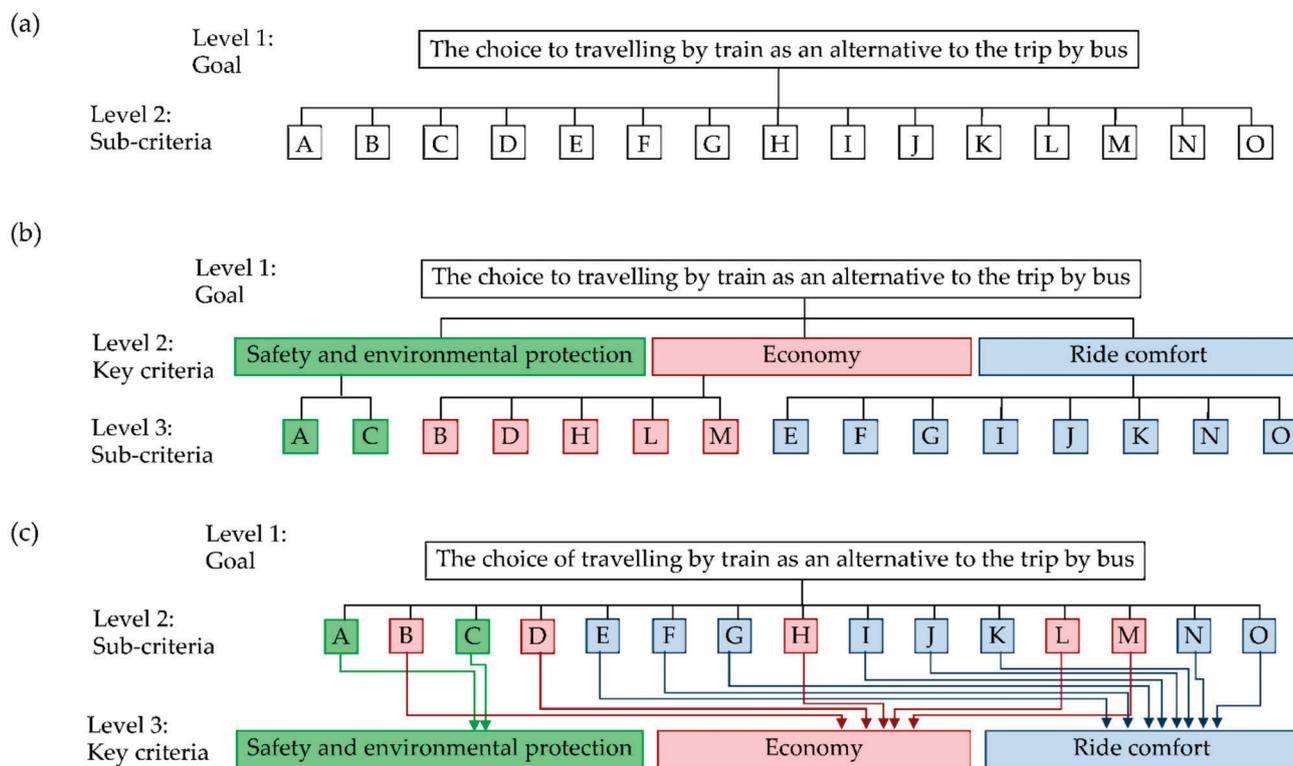


Figure 2. Calculating the weights of criteria determining the choice of passengers to travel by train as an alternative mode of transport to travel by bus using: (a)—non-hierarchical model; (b)—a classical (direct) hierarchy model; (c)—an inverse hierarchy model.

The absolute error of the survey shows how much the average of the ranks \bar{R}_j calculated for the j -th criterion of m surveyed passengers may differ from the average of the population set \bar{R}_{jp} that would be determined by surveying all passengers. Due to the limited sample size m , \bar{R}_j always differs from \bar{R}_{jp} by no more than plus or minus Δ_j . This difference is greater the smaller the m and the larger σ_{R_j} .

By interviewing m passengers and calculating the standard deviation σ_{R_j} of the ranks of j -th criterion with the 95% probability recommended in practice, the absolute error Δ_j of the j -th criterion value can be determined from formula (10) and compared with the permissible value (if any).

5. Calculating the Average Rank, the Consistency of Expert Estimates and the Criteria Weights

All 15 sub-criteria presented in the questionnaire, which determined the choice of passengers to travel by train rather than by bus, were divided into three groups and named key criteria (Figure 2b,c). The group of safety and environmental protection included two sub-criteria (A and C), the economy group embraced five sub-criteria (B, D, H, L, M) and the ride comfort group consisted of eight sub-criteria (E, F, G, I, J, K, N, O).

The ranks of the significance of sub-criteria, determining the choice by passengers of a trip by train rather than traveling by bus, were used for calculating the average values of the ranks \bar{R}_j , the concordance coefficient W , Pearson's chi-square statistic χ^2 and sub-criteria weights ω_j and ω'_j (Table 2). The following table also presents the mean values $\bar{\omega}_j$ of the sub-criteria weights ω_j and ω'_j and the standard deviations of the ranks R_{ij} .

The total of sub-criteria ranks was $\sum_{j=1}^{15} R_j = 6240$, while the sum of average ranks \bar{R}_j of all j sub-criteria was $\sum_{j=1}^{15} \bar{R}_j = 120.0$. The average value $\bar{R} = 416$ of sub-criteria ranks was calculated by Equation (6) or $\bar{R} = 6240/15 = 416$. The sum of squared deviations $S = 147,172$ (Equation (7)). The concordance coefficient $W = 0.194$, showing the consistency of the estimates of respondents (52 passengers), was calculated by Equation (6).

Based on the data from the passengers' survey and using Equation (11), $\chi^2 = 141.5$ was obtained. The critical value $\chi^2_{\alpha, \nu}$ taken from the table of chi-squared distribution with $\nu = 15 - 1 = 14$ degrees of freedom and the significance level $\alpha = 0.01$ was equal to 29.1413. The empirical value $\chi^2 = 141.5$ was 4.8 times the critical value $\chi^2_{\nu, \alpha} = 29.1$, which allowed the researchers to assume that the respondents' estimates were consistent.

The smallest value of the concordance coefficient W_{\min} , with the significance level $\alpha = 0.01$ and the degree of freedom $\nu = n - 1 = 15 - 1 = 14$, allowing the authors to assume that the respondents' estimates were consistent, was calculated by Equation (9). The smallest value of the concordance coefficient $W_{\min} = 0.0400$ corresponded to only about one-fifth of the calculated concordance coefficient $W = 0.194$.

The estimates of 52 passengers that took part in the survey were in agreement (or consistent) because the calculated concordance coefficient was equal to 0.194, while the value of Pearson's chi-squared statistic, equal to 141.5, was considerably larger than the critical value of 29.14, corresponding to degrees of freedom of 14 and a significance level of 0.01. The smallest concordance coefficient still allowing the estimates of all respondents to be considered consistent was equal to 0.0400, which was equivalent to only one-fifth of 0.194. It was hardly possible to expect very high consistency of the respondents' estimates because of their highly different experiences, wishes, habits and means.

A bar diagram of the calculated average ranks \bar{R}_j of the 15 sub-criteria determining the passengers' choice of traveling by train rather than by bus is given in Figure 3.

By applying the new ARTIW-L and ARTIW-N methods, the passengers' reasons for selecting a trip by train rather than a trip by bus, which were described by criteria (sub-criteria) and their weights ω_j , ω'_j , and $\bar{\omega}_j$, were determined. The calculation data for sub-criteria beginning from the most important (E) to the least important (D) ones are shown in Figure 4.

The calculated average ranks \bar{R}_j (Figure 3) and global weights ω_j and ω'_j (Figure 4) of sub-criteria, determining the choice by the respondents to travel by train rather than by bus, show that sub-criteria E, M and H were much more important than sub-criteria N, C and D. This implies that their priority order should be as follows:

E ∨ M ∨ H ∨ A ∨ L ∨ G ∨ F ∨ O ∨ B ∨ K ∨ J ∨ I ∨ N ∨ C ∨ D.

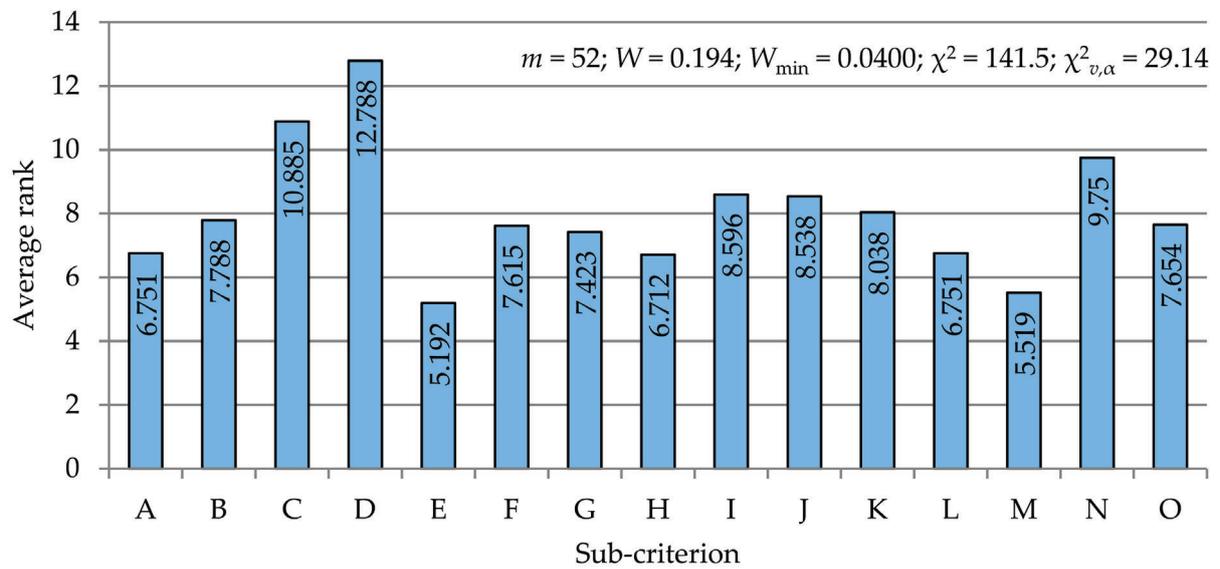


Figure 3. The average ranks of sub-criteria \bar{R}_j .

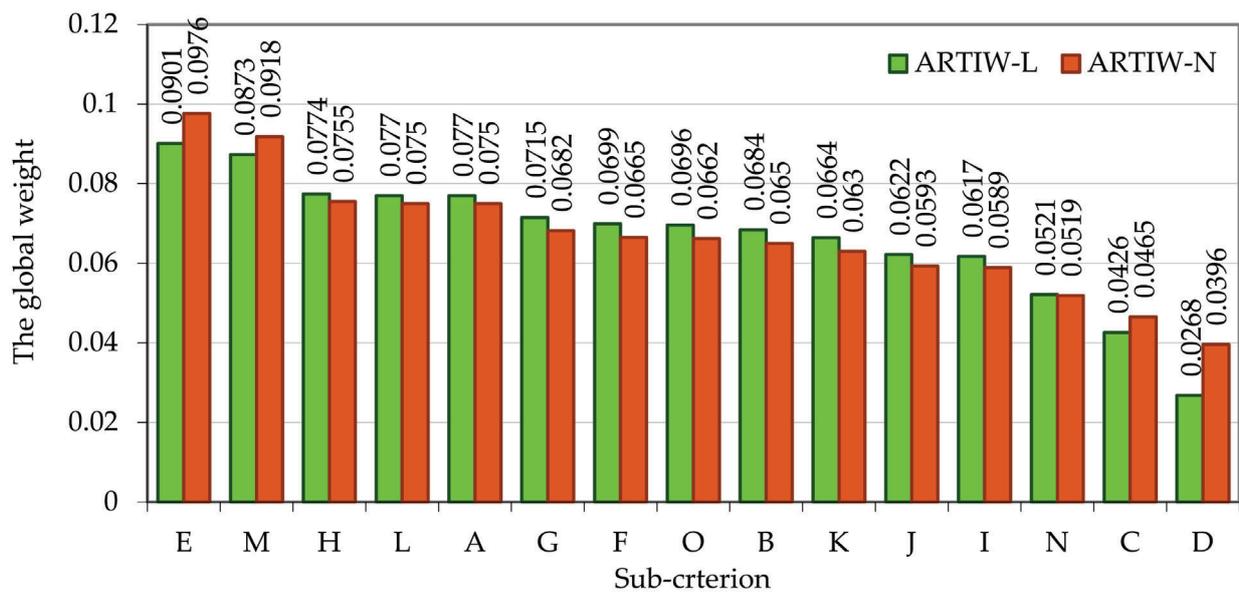


Figure 4. Relative weights of sub-criteria in descending order.

Moreover, there should be an inverse straight-line relationship between the average ranks \bar{R}_j and the global weights ω_j calculated by the ARTIW-L method. The determination coefficient of the regression equation of 15 sub-criteria, $R^2 = 1$, and coefficient of correlation $r = -1$, show that this is a functional linear relationship $\omega_j = -0.0083\bar{R}_j + 0.1333$. The correlation between \bar{R}_j and the weights ω'_j calculated using the ARTIW-N method is non-linear (Figure 5a). Ranks and weights are related by the quadratic regression equation $\omega'_j = 0.0009 \bar{R}_j^2 - 0.0234 \bar{R}_j + 0.1932$, coefficient of determination, which is $R^2 = 0.9955$.

Table 2. Averages of sub-criteria ranks \bar{R}_j of passenger choice to travel by train as an alternative to travel by bus and calculated weights ω_j and ω'_j with reference to linear ARTIW-L and non-linear ARTIW-N methods (model Figure 2a).

Equation	Sub-Criterion $j = 1, 2, \dots, n$														Sum	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N		O
$\sum_{j=1}^m R_{ij}$	351	405	566	665	270	396	386	349	447	444	418	351	287	507	398	6240
$\bar{R}_j = \frac{\sum_{i=1}^m R_{ij}}{m}$ (1)	6.751	7.788	10.885	12.788	5.192	7.615	7.423	6.712	8.596	8.538	8.038	6.751	5.519	9.750	7.654	120.000
$\sum_{j=1}^m \left[\sum_{i=1}^m R_{ij} - \frac{m(i+1)}{2} \right]^2$ (7)	4225	121	22,500	62,001	21,316	400	900	4489	961	784	4	4225	16,641	8281	324	147,172
ARTIW-L $\omega_j = \frac{\sum_{i=1}^m \bar{R}_j}{\sum_{j=1}^m \bar{R}_j}$ (2)	0.077	0.0684	0.0426	0.0268	0.0901	0.0699	0.0715	0.0774	0.0617	0.0622	0.0664	0.077	0.0873	0.0521	0.0696	1.0000
ARTIW-N $u_j = \frac{\sum_{i=1}^m R_{ij}}{\bar{R}_j}$ (3)	0.7691	0.6667	0.4770	0.4060	1	0.6818	0.6994	0.7735	0.6040	0.6081	0.6459	0.7691	0.9408	0.5325	0.6783	10.2522
$\omega'_j = \frac{u_j}{\sum_{j=1}^m u_j}$ (4)	0.0750	0.0650	0.0465	0.0396	0.0976	0.0665	0.0682	0.0755	0.0589	0.0593	0.0630	0.0750	0.0918	0.0519	0.0662	1.0000
$\bar{\omega}_j = \frac{\omega_j + \omega'_j}{2}$ (5)	0.076	0.0667	0.0445	0.0332	0.0939	0.0682	0.0699	0.0764	0.0603	0.0607	0.0647	0.076	0.0896	0.052	0.0679	1.0000
Priority	4-5	9	14	15	1	7	6	3	12	11	10	4-5	2	13	8	120
Standard deviation σ_{R_j}	4.414	4.267	4.081	3.392	4.039	4.303	3.957	3.972	3.604	4.377	3.242	3.725	3.654	3.602	3.875	-

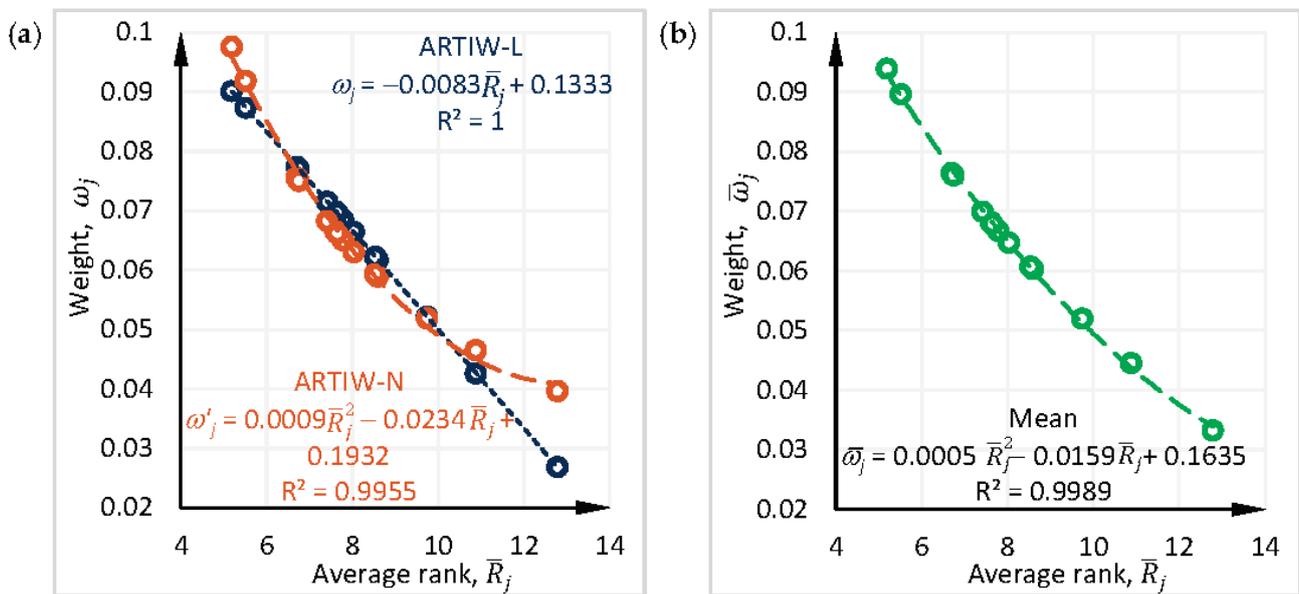


Figure 5. Correlation between the means of the sub-criteria ranks and the normalized weights of these sub-criteria calculated using: (a)—ARTIW-L and ARTIW-N methods; (b)—average values of two methods.

The data obtained in the performed study show that the estimates of the significance (importance) of sub-criteria determining the passenger choice of traveling by train rather than by bus were consistent (in agreement) and reflect their general opinion, shown by the averages $\bar{\omega}_j$ of criteria weights ω_j and ω'_j calculated according to Equation (5) and their correlation with the rank averages \bar{R}_j (Figure 5b).

With reference to the principle of determining the sample size, the absolute error Δ_j for determining the average rank \bar{R}_j of each j -th sub-criterion was calculated from formula (10) (Table 3). For the calculation of Δ_j , the values of the standard deviation σ_{R_j} of each sub-criterion were taken from Table 2 when 52 passengers were interviewed and $\alpha = 0.05$ significance level.

Table 3. The absolute error Δ_j in determining the averages of the 15 sub-criteria ranks \bar{R}_j of the group of 52 surveyed passengers.

Sub-Criterion $j = 1, 2, \dots, n$															
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1.20	1.16	1.11	0.92	1.10	1.17	1.08	1.08	0.98	1.19	0.88	1.01	0.99	0.98	1.05	

The results (Table 3) show that the mean ranks \bar{R}_j of the 15 sub-criteria were identified with an absolute error Δ_j , the sample range of which was $\Delta_{j\max} - \Delta_{j\min} = \Delta_{jA} - \Delta_{jK} = 1.20 - 0.88 = 0.32$, and the mean value $\bar{\Delta}_j = 1.06$. By taking Δ_j for each sub-criterion or the mean value $\bar{\Delta}_j$ of 15 sub-criteria, it was possible to calculate the confidence interval $\bar{R}_j \pm \Delta_j$ of ranks average \bar{R}_j , with the population mean rank \bar{R}_{jp} of 95% confidence. For example, the population mean rank \bar{R}_{jA} of the sub-criterion A with the highest rank variation was in the range 6.751 ± 1.20 , i.e., between 5.551 and 7.951. The range would decrease if more than 52 passengers were interviewed. We believe that $\bar{\Delta}_j = 1.06$ is close to one, so the number of passengers $m = 52$ who completed the survey was sufficient and allowed us to reliably assess the factors determining the choice of passengers to travel by train as an alternative to bus.

6. Calculating the Overall Weights of Key Criteria

It was rather difficult to determine the weights ω_j of 15 sub-criteria by using the AHP approach because the optimal number of criteria for this method was seven plus or minus two [55,66].

When the global weights $\bar{\omega}_j$ of all 15 sub-criteria were determined and sub-criteria were divided into three groups (key criteria) as shown in Figure 2c, the overall weights of the key criteria $\tilde{\omega}_g$ (see Table 4) were calculated as follows using the inverse hierarchy for assessment main criteria importance (IHAMCI) method:

$$\tilde{\omega}_g = \frac{\sum_{j=1}^k \omega_j/k}{\sum_{b=1}^g \sum_{j=1}^k \omega_j/k}, \tag{11}$$

where ω_j is the global weight of j -th sub-criterion, k is the number of sub-criteria in the group ($j = 1, 2, \dots, k$), g is the number of groups of criteria describing the research object ($b = 1, 2, \dots, g$).

The overall weight $\tilde{\omega}_{Sa}$ of two sub-criteria, A + C, included in the travel safety and environmental protection group (key criterion), which was calculated by Equation (11), was the smallest:

$$\tilde{\omega}_{Sa} = \frac{0.1205 : 2}{0.0602 + 0.0684 + 0.0672} = 0.3077.$$

The overall weight $\tilde{\omega}_{Ec}$ of five sub-criteria, B + D + H + L + M, included in the key criterion describing economy, was calculated in the same way, as follows:

$$\tilde{\omega}_{Ec} = \frac{0.3419 : 5}{0.0602 + 0.0684 + 0.0672} = 0.3492.$$

The overall weight $\tilde{\omega}_{Co}$ of the key criterion of eight sub-criteria, E + F + G + I + J + K + N + O, describing ride comfort, was the largest:

$$\tilde{\omega}_{Co} = \frac{0.5376 : 8}{0.0602 + 0.0684 + 0.0672} = 0.3431.$$

The results of calculation show that the choice of passengers to travel by rail transport rather than by road transport (a bus) was determined by the criteria describing ride comfort and economy (about 35%) as well as safety and environmental protection (only about 31%).

The obtained global weights $\bar{\omega}_j$ of sub-criteria and the overall weights $\tilde{\omega}_g$ of criteria divided into three groups, as shown in Figure 6, allowed the authors to identify the criteria determining the choice of passengers to travel by train (as an alternative to travel by bus). The obtained data can be used by companies engaged in passenger transportation by rail to enhance the quality of services provided by this more environmentally friendly mode of land transport so that it will have a competitive edge over rival modes of transportation.

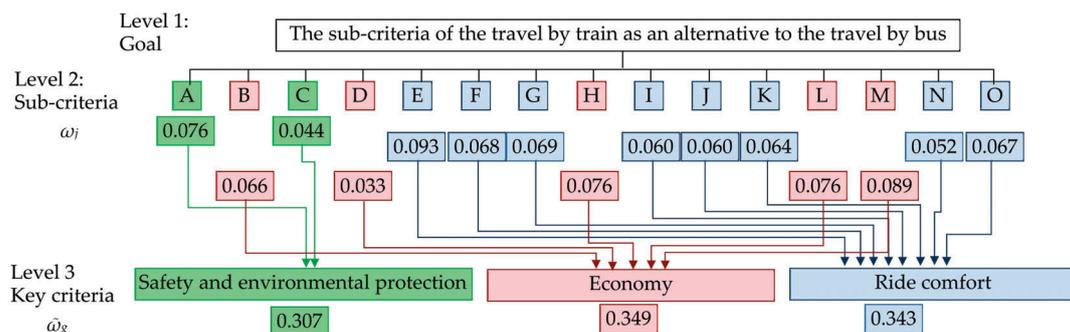


Figure 6. The calculated subjective weights of sub-criteria and the key criteria.

Table 4. The calculated global, local and overall weights of the criteria determining the choice of passengers to travel by train.

Factors (Sub-Criteria)		Factors (Sub-Criteria)				Factors (Sub-Criteria)			Factors (Sub-Criteria)					
Global Weights of Sub-Criterion ω_j	Priority	Size of Key Criterion k	Global Weights of Sub-Criterion ω_j	The Sum of Global Weights $\omega_g = \sum_{j=1}^k \omega_j$	Normalized Local Weights $\frac{\omega_j}{\omega_g}$	The Sum of Local Weights $\sum_{j=1}^k \omega_j$	Average Weight of Sub-Criteria $\frac{\omega_g}{k}$	The Overall Weight of Key Criterion $\tilde{\omega}_g$						
A	4-5	Safety and environmental protection $k_{Sa} = 2$	0.076	0.1205	0.6307	1.0000	0.1205:2 = 0.0602	0.3077						
B	9		0.0667	0.0445	0.3693									
C	14	Economy $k_{Ec} = 5$	0.0445	0.3419	0.1951	1.0000	0.3419:5 = 0.0684	0.3492						
D	15		0.0332		0.0971									
E	1		0.0939		0.2235									
F	7		0.0699		0.2223									
G	6		0.0764		0.2620									
H	3	Ride comfort $k_{Co} = 8$	0.0603	0.5376	0.1747	1.0000	0.5376:8 = 0.0672	0.3431						
I	12		0.0607		0.1269									
J	11		0.0647		0.1300									
K	10		0.0647		0.1122									
L	4-5		0.076		0.1129									
M	2		0.0896		0.1203									
N	13		0.052		0.0967									
O	8		0.0679		0.1263									
Total	-		-		1.0000				1.0000	3.0000	3.0000	0.1958	1.0000	

The sub-criteria and key criteria weights calculated in this study, which show why passengers choose train travel as an alternative to the bus, are not absolutely accurate and constant. When interviewing citizens of other countries traveling on international trains, the significance of the sub-criteria and key criteria may differ. Their values can be influenced by the economic development of the country, passenger habits, the reliability factors relating to different modes of transport, and risks.

The most important part of the study consisted of the original sub-criteria system and their weighting methodology, applying the new ARTIW-L and ARTIW-N methods, as well as the method of inverse hierarchy for assessment main criteria importance (IHAMCI). These methods can be used by other researchers to calculate normalized subjective weights when ranking the results of an expert or respondent survey.

7. Discussion and Conclusions

Passengers usually make a decision to travel by a particular mode of transport by evaluating the criteria describing it, whose weights reflecting their significance seem to be different to them. The selection of a particular (alternative) mode of transport is based on the significances (subjective weights) of the considered criteria, which can be determined by using expert evaluation methods. The average value of the estimates given by a considerable number of passengers (respondents) in ranking the criteria can be used as a result, presenting public opinion about a particular transport mode chosen for a particular route if their opinions (judgments) are considered.

In the present work, the reasons behind the passengers' choice to travel by train rather than by bus were identified by considering fifteen sub-criteria. The significances of these sub-criteria for choosing travel by train were evaluated by 52 respondents (passengers on the Vilnius–Moscow–Vilnius train) against a 15 rank scale. The subjective total normalized weights of sub-criteria based on the new ARTIW-L and ARTIW-N methods allowed the authors to rank them by order of priority (preference). A functional or close and strong correlation between the means of the sub-criteria ranks and the normalized weights of the sub-criteria calculated from them indicated that the ARTIW-L and ARTIW-N methods were satisfactory to assess the significance of the sub-criteria. The mean of the sub-criteria weights calculated by these two methods was taken as the final significance of the sub-criteria. The sub-criteria determining the passengers' choice of traveling by train rather than by bus (as an alternative mode of transport) included ride comfort (the availability of berths in passenger cars for sleeping and relaxation) (0.0939), the selected time of travel (0.0896) and a negligible effect of weather conditions on it (0.0764). The sub-criteria describing rail transport as safer than road transport, as well as such advantages as a smaller number of stops and delays on the way (0.0760), the availability of WCs and places for smoking (0.0699), a lower probability that passengers would disturb each other (0.0682), and simpler border control (0.0679) were less important for passengers. However, even less important for them were sub-criteria describing the priority given to rail vehicles when crossing motor roads (0.0667) and freedom of movement (0.0647), the availability of a dining car (0.0607) and unpleasant slight rocking and vibration (0.0603). The least important sub-criteria for passengers included the possibility to order food or newspapers and magazines to the compartment (0.0520), lower environmental pollution by rail transport (0.0445), and sometimes cheaper railway tickets (0.0332). The ratio of the largest total weight of sub-criteria (0.0939) to their smallest total weight (0.0332), which was equal to 3.36, showed that the significance of particular sub-criteria was different for passengers choosing a particular mode of transport.

The fifteen considered sub-criteria were divided into three groups using the inverse hierarchy for assessment main criteria importance (IHAMCI) method suggested by the second author. The normalized overall weights for these groups (criteria) were calculated. The overall weight of two sub-criteria describing safety and environmental protection was equal to 0.3077, while the overall weight of five sub-criteria forming the 'economy' group

was 0.3492, and the overall weight of eight sub-criteria referring to ride comfort was the largest, at 0.3431. The overall weights of the key criteria of any sub-criteria group were calculated using a new IHAMCI method, which allowed assessment of different numbers of sub-criteria in a key criterion.

The minimum number of experts or respondents to be consulted in order to obtain reliable results was calculated using the sample size principle. After interviewing 52 passengers, the values of the standard deviations of the sub-criteria ranks were identified, and then used to determine the absolute error of the mean of the ranks of each sub-criterion. The results showed that with 95% probability, the sub-criteria sample rank averages differed from the population averages by no more than 1.20–0.88 rank (on average 1.06 rank). This difference was close to unity and indicated that the significance of the sub-criteria was determined with sufficient accuracy.

The decision-makers in the countries engaged in passenger transportation by the considered international train should primarily improve the services described by the criteria that most strongly influence decisions by passengers to choose a trip by train rather than by bus. A company providing passenger transportation by any particular mode of transport can win the competition in this area only if its provided services are of the highest quality and satisfy the ever-growing demands of passengers.

Author Contributions: Conceptualization, H.S.; methodology, H.S.; investigation, H.S. and L.M.; writing—original draft preparation, H.S.; writing—review and editing, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Reviewer Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Analyzed datasets are available from the corresponding author.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. Sakib, N.; Appiotti, F.; Magni, F.; Maragno, D.; Innocenti, A.; Gissi, E.; Musco, F. Addressing the passenger transport and accessibility enablers for sustainable development. *Sustainability* **2018**, *10*, 903. [[CrossRef](#)]
2. Pijoan, A.; Kamara-Esteban, O.; Alonso-Vicario, A.; Borges, C.E. Transport choice modeling for the evaluation of new transport policies. *Sustainability* **2018**, *10*, 1230. [[CrossRef](#)]
3. Li, X.-H.; Huang, L.; Li, Q.; Liu, H.-C. Passenger satisfaction evaluation of public transportation using Pythagorean fuzzy MULTIMOORA method under-large group environment. *Sustainability* **2020**, *12*, 4996. [[CrossRef](#)]
4. European Commission. *White Paper on Transport. Roadmap to a Single European Transport Area-Towards a Competitive and Resource-Efficient Transport System*; Directorate-General for Mobility and Transport: Brussels, Belgium, 2011.
5. Al-Atawi, A.; Kumar, R.; Saleh, W. Transportation sustainability index for Tabuk city in Saudi Arabia: An Analytic Hierarchy Process. *Transport* **2016**, *31*, 47–55. [[CrossRef](#)]
6. Paha, J.; Rompf, D.; Warnecke, C. Customer choice patterns in passenger rail competition. *Transp. Part A Pol. Pract. Res.* **2013**, *50*, 209–227. [[CrossRef](#)]
7. Bauk, S.; Šekularac-Ivošević, S.; Jolič, N. Seaport positioning supported by the combination of some quantitative and qualitative approaches. *Transport* **2015**, *30*, 385–396. [[CrossRef](#)]
8. Földes, D.; Csiszár, C. Route plan evaluation method for personalised passenger information service. *Transport* **2015**, *30*, 273–285. [[CrossRef](#)]
9. Buliček, J.; Drdla, P. Quality of urban public transport websites according to extent of provided information. *Transport* **2015**, *30*, 202–216. [[CrossRef](#)]
10. Skačkauskienė, I.; Vilkaitė-Vaitonė, N.; Raudeliūnienė, J.; Davidavičienė, V. A model for measuring passenger loyalty. *Transport* **2016**, *31*, 100–107. [[CrossRef](#)]
11. Sivilevičius, H.; Maskeliūnaitė, L. Multiple criteria evaluation and the inverse hierarchy model for justifying the choice of rail transport mode. *PROMET-Traffic Transp.* **2018**, *30*, 57–69. [[CrossRef](#)]
12. Kumar, S.; Dogiparthi, K.C.; Vanajakshi, L.; Subramanian, S.C. Integration of exponential smoothing with state space formulation for bus travel time and arrivaltime prediction. *Transport* **2017**, *32*, 358–367. [[CrossRef](#)]

13. Stenström, C.; Parida, A.; Kumar, U. Measuring and monitoring operational availability of rail infrastructure. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 1457–1468. [[CrossRef](#)]
14. Matuška, J. Railway system accessibility evaluation for wheelchair users: Case study in the Czech Republic. *Transport* **2017**, *32*, 32–43. [[CrossRef](#)]
15. *TRB Special Report 320: Interregional Travel: A New Perspective for Policy Making*; Transport Research Board: Washington, DC, USA, 2016; p. 187.
16. Dolinayova, A.; Camaj, J.; Kanis, J. Charging railway infrastructure models and their impact to competitiveness of railway transport. *Transp. Probl.* **2017**, *12*, 139–150. [[CrossRef](#)]
17. Barbosa, R.S. Evaluation of railway track safety with a new method for track quality identification. *J. Transp. Eng.* **2016**, *142*, 04016053. [[CrossRef](#)]
18. Xin, T.; Famurewa, S.M.; Gao, L.; Kumar, U.; Zhang, Q. Grey-system-theory-based model for the prediction of track geometry quality. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 1735–1744. [[CrossRef](#)]
19. Bai, L.; Liu, R.; Sun, Q.; Wang, F.; Wang, F. Classification learning-based framework for predicting railway track irregularities. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 598–610. [[CrossRef](#)]
20. Lazarević, L.; Vučković, D.; Popović, Z. Assessment of sleeper support conditions using micro-tremor analysis. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 1828–1841. [[CrossRef](#)]
21. Afazov, S.; Denmark, W.; Yaghi, A. Modelling aspects of the design of railway vehicle structures and their crashworthiness. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 1575–1589. [[CrossRef](#)]
22. Lin, C.-Y.; Saat, M.R.; Barkan, C.P.L. Fault tree analysis of adjacent track accidents on shared-use rail corridors. *Transp. Res. Rec.* **2016**, *2546*, 129–136. [[CrossRef](#)]
23. Gordon, J.; Llana, P.; Severson, K.; Tyrell, D. Research into integrity of glazing for passenger rail equipment. *Transp. Res. Rec.* **2016**, *2546*, 88–93. [[CrossRef](#)]
24. Kovandová, H.; Válka, R. Experimental study non-compatible collision of rail and road vehicle. *PROMET–Traffic Transp.* **2014**, *26*, 459–466. [[CrossRef](#)]
25. Xu, X.; Li, K.; Li, X. Research on passenger flow and energy consumption in a subway system with fuzzy passenger arrival rates. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2015**, *229*, 863–874. [[CrossRef](#)]
26. Sun, Y.; Schonfeld, P.M. Schedule-based rail transit patch-choice estimation using automatic fare collection data. *J. Transp. Eng.* **2016**, *142*, 04015037. [[CrossRef](#)]
27. Xiang, X.; Zhu, X. An optimization strategy for improving the economic performance of heavy-haul railway networks. *J. Transp. Eng.* **2016**, *142*, 04016003. [[CrossRef](#)]
28. Makarova, E.A.; Muktepavel, S.V. Razrabotka sistemy raschetno-analiticheskikh pokazateley, kharakterizujushchikh regional'nye passazhirskie perevozki. *Vest VNIIZhT* **2016**, *3*, 169–173. [[CrossRef](#)]
29. Tang, H.; Dick, C.T.; Caughron, B.M.; Feng, X.; Wang, Q.; Barkan, C.P.L. Model for optimal selection of projects to improve running time and operating cost efficiency on passenger rail corridors. *Transp. Res. Rec.* **2016**, *2546*, 33–42. [[CrossRef](#)]
30. Liao, M.; Liu, G. Modeling passenger behaviour in nonpayment areas at rail transit stations. *Transp. Res. Rec.* **2015**, *2534*, 101–108. [[CrossRef](#)]
31. Allen, J.G.; Levinson, H.S. What happened to speed. *Transp. Res. Rec.* **2016**, *2546*, 60–68. [[CrossRef](#)]
32. Lee, Y.; Rho, J.; Kim, K.H.; Lee, D.H.; Kwon, H.B. Experimental studies on the aerodynamic characteristics of a pantograph suitable for a high-speed train. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2015**, 136–149. [[CrossRef](#)]
33. Ou, X.; Zhang, G.; Chang, X.Y. Renewal and development of intercity passenger rail system. *Transp. Res. Rec.* **2016**, *2546*, 53–59. [[CrossRef](#)]
34. Teixeira, P.; Prodan, A. Railway infrastructure pricing in Europe for high-speed and intercity services. *Transp. Res. Rec.* **2014**, *2448*, 1–10. [[CrossRef](#)]
35. Oh, S.; Kim, S.; Hong, J. Comprehensive analysis of the influence of door width on the passenger flow time on Korean urban railways. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 1531–1539. [[CrossRef](#)]
36. Holloway, C.; Thoreau, R.; Roan, T.-R.; Boampong, D.; Clarke, T.; Watts, D.; Tyler, N. Effect of vertical step height on boarding and alighting time of train passengers. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2016**, *230*, 1234–1241. [[CrossRef](#)]
37. Barić, D.; Pilko, H.; Strujić, J. An Analytic Hierarchy Process model to evaluate road section design. *Transport* **2016**, *31*, 312–321. [[CrossRef](#)]
38. Mardani, A.; Zavadskas, E.K.; Khalifah, Z.; Jusoh, A.; Nor, K.M. Multiple Criteria Decision-Making techniques in transportation systems: A systematic review of the state of the art literature. *Transport* **2016**, *31*, 359–385. [[CrossRef](#)]
39. Sayers, T.M.; Jessop, A.T.; Hills, P.J. Multi-criteria evaluation of transport options-flexible, transparent and user-friendly. *Transp. Pol.* **2003**, *10*, 95–105. [[CrossRef](#)]
40. Pandian, P.; Sundaram, V.D.; Sivaprakasam, R. Development of fuzzy based intelligent decision model to optimize the blind spots in heavy transport vehicles. *PROMET–Traffic Transp.* **2016**, *28*, 1–10. [[CrossRef](#)]
41. Podvezko, V.; Sivilevičius, H. The use of AHP and rank correlation methods for determining the significance of the interaction between the elements of a transport system having a strong influence on traffic safety. *Transport* **2013**, *28*, 389–403. [[CrossRef](#)]
42. Lin, C.-T.; Hsu, S.-C. Evaluating travel risk from the freelance perspective. *J. Test. Eval.* **2016**, *44*, 2033–2044. [[CrossRef](#)]

43. Liao, M.-S.; Ding, J.-F.; Liang, G.-S.; Lee, K.-L. Key criteria for evaluating the green performance of ports. *J. Test. Eval.* **2016**, *44*, 1791–1801. [[CrossRef](#)]
44. Camargo-Pérez, J.; Carrillo, M.H.; Montoya-Torres, J.R. Multi-criteria approaches for urban passenger transport systems: A literature review. *Ann. Oper. Res.* **2015**, *226*, 69–87. [[CrossRef](#)]
45. Di Gravio, G.; Patriarca, R.; Costantino, F. Safety assessment for an ATM system change: A methodology for the ANSPs. *PROMET–Traffic Trans.* **2017**, *29*, 99–107. [[CrossRef](#)]
46. Chen, S.; Leng, Y.; Mao, B.; Liu, S. Integrated weight-based multi-criteria evaluation on transfer in large transport terminals: A case study of the Beijing South railway station. *Transp. Res. Part A Pol. Ptact.* **2014**, *66*, 13–26. [[CrossRef](#)]
47. Stoilova, S. Study of railway passenger transport in the European Union. *Tehnički Vjesnik* **2018**, *25*, 587–595. [[CrossRef](#)]
48. Maskeliūnaitė, L.; Sivilevičius, H. The model for evaluating the criteria describing the quality of the trip by international train. *Technol. Econ. Develop. Econ.* **2014**, *20*, 484–506. [[CrossRef](#)]
49. Sivilevičius, H.; Maskeliūnaitė, L. The numerical example for evaluating the criteria describing the quality of the trip by international train. *E+M Ekon. Manag.* **2014**, *17*, 73–86. [[CrossRef](#)]
50. Pastukhov, S.S. Izuchenie vliyaniya parametrov transportnogo servisa na uroven' udovletvorennosti passazhirov na osnove modelirovaniya metodami poryadkovoy logisticheskoy peggessii. *Vest VNIIZhT* **2016**, *2*, 108–115.
51. Kesten, A.S.; Ögüt, K.S. A new passenger-oriented performance measurement framework for public rail transportation systems. *PROMET–Traffic Transp.* **2014**, *26*, 299–311. [[CrossRef](#)]
52. Zavadskas, E.K.; Podvezko, V. Integrated determination of objective criteria weights in MCDM. *Int. J. Inf. Technol. Decis. Mak.* **2016**, *15*, 267–283. [[CrossRef](#)]
53. Stanujkic, D. An extension of the ratio system approach of MOORA method for group decision-making based on interval-valued triangular fuzzy numbers. *Technol. Econ. Develop. Econ.* **2016**, *22*, 122–141. [[CrossRef](#)]
54. Saaty, T.L. *The Analytic Hierarchy Process*; McGraw-Hill: New York, NY, USA, 1980; p. 287.
55. Saaty, T.L. How to make a decision: The analytic hierarchy process. *Europ. J. Operat. Res.* **1990**, *48*, 9–26. [[CrossRef](#)]
56. Saaty, L.; Vargas, L.G. *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*, 2nd ed.; Springer: New York, NY, USA, 2012; p. 343. ISBN 978-1-4614-3596-9.
57. Kendall, M.E. *Rank Correlation Methods*, 4th ed.; Griffin and Co: London, UK, 1970; p. 365.
58. Kendall, M.; Gibbons, J.D. *Rank Correlation Methods*; Edward Arnold: London, UK, 1990; p. 260.
59. Li, F.; Phoon, K.K.; Du, X.; Zhang, M. Improved AHP method and its application in risk identification. *J. Const. Eng. Manag.* **2013**, *139*, 312–320. [[CrossRef](#)]
60. Orugbo, E.E.; Alkali, B.M.; DeSilva, A.; Harrison, D.K. RCM and AHP hybrid model for road network maintenance prioritization. *Balt. J. Road Bridge Eng.* **2015**, *10*, 182–190. [[CrossRef](#)]
61. Chen, S.; Hung, H.-F. Formative measurement test of supplier evaluation indicators and weight models. *J. Test. Eval.* **2016**, *44*, 1350–1363. [[CrossRef](#)]
62. Chang, K.-H.; Chain, K.; Wen, T.-C.; Yang, G.K. A novel general approach for solving a supplier selection problem. *J. Test. Eval.* **2016**, *44*, 1911–1924. [[CrossRef](#)]
63. Sivilevičius, H. Application of expert evaluation method to determine the importance of operating asphalt neixing plant quality criteria and rank correlation. *Balt. J. Road Bridge Eng.* **2011**, *6*, 48–58. [[CrossRef](#)]
64. Podvezko, V. Determining the level of agreement of expert estimates. *Technol. Econ. Develop. Econ.* **2005**, *11*, 101–107. [[CrossRef](#)]
65. Montgomery, D.C. *Statistical Quality Control: A Modern Introduction*, 7th ed.; John Wiley & Sons Inc.: Arizona, AZ, USA, 2013; p. 752.
66. Saaty, T.L.; Ozdemir, M.S. Why the magic number seven plus or minus two. *Mathemat. Comput. Model.* **2003**, *38*, 233–244. [[CrossRef](#)]
67. Maskeliūnaitė, L.; Sivilevičius, H.; Podvezko, V. Research on the quality of passenger transportation by railway. *Transport* **2009**, *24*, 100–112. [[CrossRef](#)]
68. Sun, C.-C. A novel approach for evaluating TFT-LCD manufacturer operational performance. *J. Test. Eval.* **2016**, *44*, 1421–1429. [[CrossRef](#)]
69. Navikas, D.; Sivilevičius, H.; Bulevičius, M. Investigation and evaluation of railway ballast properties variation during technological processes. *Constr. Build. Mater.* **2018**, *185*, 325–337. [[CrossRef](#)]
70. Madanu, S.K.; Mattingly, S.P.; Kam, K.A.; Ardekani, S.A. Integrated fuzzy technique for order preference by similarity to ideal solution framework for evaluating high-speed passenger rail corridor alternatives. *Transp. Res. Rec.* **2015**, *2499*, 1–19. [[CrossRef](#)]
71. Saaty, T.L.; Shang, J.S. An innovative orders of-magnitude approach to AHP-based multi-criteria decision making: Prioritizing divergent intangible humane acts. *Eur. J. Operat. Res.* **2011**, *214*, 703–715. [[CrossRef](#)]
72. Gudienė, N.; Banaitis, A.; Podvezko, V.; Banaitienė, N. Identification and evaluation of the critical success factors for construction projects in Lithuania: AHP approach. *J. Civ. Eng. Manag.* **2014**, *20*, 350–359. [[CrossRef](#)]
73. Chang, S.-C.; Tsai, P.-H. A hybrid financial performance evaluation model for wealth management banks following the global financial crisis. *Technol. Econ. Develop. Econ.* **2016**, *22*, 21–46. [[CrossRef](#)]

74. Beiler, M.O.; Wakszynski, E. Measuring the sustainability of shared-use paths: Development of the Green Paths rating system. *J. Transp. Eng.* **2015**, *141*, 04015026. [[CrossRef](#)]
75. Rabbani, A.; Zamani, M.; Yardani-Chamzini, A.; Zavadskas, E.K. Proposing a new integrated model based on balanced scorecard (SBSC) and MCDM approaches by using linguistic variables for the performance evaluation of oil producing companies. *Expert Syst. Appl.* **2014**, *41*, 7316–7327. [[CrossRef](#)]

Article

Semi-Markov Model of the System of Repairs and Preventive Replacements by Age of City Buses

Klaudiusz Migawa, Sylwester Borowski *, Andrzej Neubauer and Agnieszka Sołtysiak

Faculty of Mechanical Engineering, UTP University of Science and Technology, Al. Prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland; klaudiusz.migawa@pbs.edu.pl (K.M.); aneub@umk.pl (A.N.); agnieszka.soltysiak@pbs.edu.pl (A.S.)

* Correspondence: sylwester.borowski@pbs.edu.pl

Abstract: The paper presents a mathematical model of the system of repairs and preventive replacements by age of city buses. The mathematical model was developed using the theory of semi-Markov processes. In the model developed, four types of city bus renewal processes are considered and three types of corrective repairs and preventive replacement. Corrective repairs are considered in two types: minimal repairs (repairs carried out by the Technical Service units) and perfect repairs (repairs carried out at the stations of the Service Station). The models of restoration systems that use semi-Markov processes in which minimal repairs, perfect repairs, and preventive replacements by age, have been examined in the literature to a limited extent. The system under consideration is analysed from the point of view of two criteria: profit per time unit and availability of city buses to carry out the assigned transport tasks. Conditions of criterion functions' extremum (maximum) existence were formulated for the adopted assumptions. The considerations presented in the paper are illustrated by exemplary results of calculations.

Keywords: city buses; semi-Markov processes; preventive maintenance; corrective maintenance; age-replacement; minimal repair; perfect repair; profit per time unit; availability

Citation: Migawa, K.; Borowski, S.; Neubauer, A.; Sołtysiak, A. Semi-Markov Model of the System of Repairs and Preventive Replacements by Age of City Buses. *Appl. Sci.* **2021**, *11*, 10411. <https://doi.org/10.3390/app112110411>

Academic Editors:

Giovanni Randazzo,
Anselme Muzirafuti and Dimitrios
S. Paraforos

Received: 5 October 2021

Accepted: 1 November 2021

Published: 5 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The basic task of transport systems is to transport people, animals and goods. The transport tasks, due to their particular specification, are carried out by different types of transport means. A very important branch of the transport system is the road passenger transport, which can generally be divided into international transport, interurban transport and urban transport. Urban transport systems usually operate in medium and large cities, in suburban areas and in industrialised areas. One of the important types of urban transport is the urban bus transport. The task of this type of transport system is to reliably and punctually carry out transport tasks along defined routes in accordance with an accepted timetable of courses [1]. The basic characteristics for evaluating the functioning of this type of transport system are economic efficiency characteristics (e.g., profit per unit time) and operational-technical efficiency characteristics (e.g., readiness of the city buses to carry out the assigned transport tasks) [2,3]. Any kind of disruption in the implementation of the assigned transport tasks, including downtime caused by damages to the means of transport (city buses) causes a decrease in the reliability and readiness of technical facilities and generates additional costs (losses). These losses arise as a result of corrective maintenance (CM) conducted after damage, losses caused by fines for non-performance of transport services, and costs related to the maintenance of reserve buses whose task is to replace damaged buses. One of the ways of ensuring the correct and efficient fulfilment of transport tasks in urban bus operation systems is the implementation of preventive maintenance (PM). The implementation of these activities consists of planning the timing and scope of preventive maintenance in such a way as to keep their costs lower than the costs of repairs after the damage. For this reason, the determination of optimal times for

preventive maintenance of technical objects is an important problem of planning strategies in the systems of exploitation of the means of transport [4].

During the operation of technical objects, their elements are subject to wear and tear processes and the impact of external factors, which causes damage to these objects. The resulting damages are the cause of lowering the effectiveness of functioning of the analysed systems. In order to ensure an appropriate level of reliability of technical facilities, different types of strategies are applied in the subsystems for ensuring serviceability. These activities are divided into two types: preventive maintenance and corrective maintenance. In practice, corrective maintenance is carried out in two variants. Firstly, the perfect repair (PR), which makes the system "As-Good-As-New" (AGAN), and secondly the minimal repair (MR), which makes the system "As-Bad-As-Old" (ABAO) are conducted. Normally, CM corrective repair costs and times are higher than PM preventive repair costs. This is due to the fact that, in general, CM activities require prior identification of the damage and high skills of personnel (diagnosticians and mechanics). In addition, there are the costs associated with the unplanned downtime of technical facilities caused by the damage to them. From this, it follows that it is possible to plan preventive actions (the scope and frequency of preventive repairs and replacements) in such a way as to ensure the required level of readiness of technical facilities and to reduce system maintenance costs. This requires the development and application of rational preventive repair and replacement strategies. For these reasons, the development of various preventive action strategies with the application of optimal decision models to reduce the system maintenance costs and the risk of adverse events is an important research topic in reliability engineering.

For the first time, the concept of minimal repair can be found in Morse's study [5]. In this study, a repair model is considered in which the criterion function is the monthly revenue generated for the technical facility under consideration. However, this model was developed on the basis of the queueing theory and not the reliability theory. However, the concept of minimum repairs in relation to the reliability theory was introduced by Barlow and Hunter in their paper [6]. In this paper, the model of periodic replacements and minimum repairs is considered, in which it is assumed that after each minimum repair the damaged technical system is restored only to the same failure condition as before the damage. In a formal way, the concept of minimum repairs was defined by Nakagawa and Kowada in their paper [7]. In the paper [8] Brown and Proschan also consider the issue of minimal repair. In this paper, the authors assume that when a technical object is damaged, a perfect repair is performed with probability p , while a minimal repair is performed with probability $q = 1 - p$. A modified version of such a model was proposed by Fontenot and Proschan [9]. In the model developed, the object is replaced with a new one after time T and either a perfect repair or a minimal repair is performed with probabilities p and q , respectively, for intermittent failures.

In the literature one can find descriptions of models of systems with minimum repairs, which have been developed with the use of various methods and mathematical tools. An overview of the used modelling methods and the construction of criterion functions in models of minimum repairs with preventive maintenance can be found, for example, in the papers [10,11]. The papers classify and discuss models of maintenance strategies for technical objects developed for both finite and infinite time horizons, in which the criterion functions are total costs, unit costs, reliability and readiness. Most of the models presented in the literature have been developed on the basis of renewal theory, while less frequently with the use of stochastic processes, including Markov and semi-Markov process models. For example, in the paper [12] the criterion functions cost per unit time and system availability were determined on the basis of a semi-Markov model in an infinite time horizon, and in the paper [13] the model of the imperfect maintenance system was developed using the theory of Markov processes, and the readiness function is a criterion for optimisation.

In practice, the effectiveness of the realised repair is between AGAN and ABAO repair and it concerns the so-called imperfect maintenance/repairs. The methods concerning

preventive repairs and replacements using the repair mechanism with an imperfect maintenance model with the (p, q) rule are extensively discussed in the paper [10]. The paper [14] presents the problem of imperfect repair with periodic preventive replacement. Models of preventive replacements by age are presented in papers [15,16]. In this type of model, it was assumed that the probabilities p and q depend on the age of the technical object at the time of failure, and that a thorough repair restores the technical object to the reliability state as for a new object, while a minimal repair restores the technical object to the reliability state just before failure. In the paper [17] it was shown that the PM policy limiting the possibility of failure can be more cost-effective than the PM policy implemented according to age, while the authors of the papers [18,19] analysed the imperfect repair system model with a delayed time concept.

Models of imperfect maintenance systems, using different age replacement policies that take into account different types of repairs after failure and their cost structures, have been presented in a number of papers. In [20], the authors consider replacement policies depending on the age of the system and the minimisation of repair costs. The authors of the papers [21,22] consider the age replacement policy of system subject to shocks in their models. Other papers consider policies that assume randomness of model parameters, e.g., papers [23,24] assume random repair costs. On the other hand, the paper [25] describes an age replacement policy with Bayesian imperfect repair model, in which the probability of an exact repair is a random variable with a specified distribution. A similar approach is adopted in the paper [26], where the optimal age replacement policy is determined in the case of minimising the cost per unit time. The results were obtained both for an infinite time horizon and for a single replacement cycle. The sequential imperfect preventive maintenance model for city buses is presented in the paper [27]. In this model, the optimal decision-making concerning the efficiency of maintenance of city buses is realized on the basis of the evaluation of the difference between the actual and expected increments of the intensity of damage.

The results presented in this paper are a continuation of the considerations presented in papers [4,28,29]. Similarly, to this paper, the results presented in these papers were obtained based on the study of the semi-Markov models. In the paper [4] a 4-state model of replacements according to the age of technical objects with a guarantee was analysed, in which the criterion function is the cost of preventive replacement determined per unit of time, while in the paper [28] a multi-state model of exploitation decisions was developed, in which corrective repairs (after a damage) and preventive replacements are carried out, and the profit per unit of time was used as the criterion function. A direct continuation of the conducted research are the results presented in the paper [29], in which a 4-state model of a service system with minimum repair was considered. The model was developed with the application of the theory of semi-Markov processes, and the theoretical considerations were illustrated with numerical examples on the basis of the assumed sample data. In this paper, on the other hand, the 5-state semi-Markov model of the system of preventive repairs and replacements according to the age of city buses is considered. The aim of the paper is to develop and study a semi-Markov model of preventive repairs and replacements according to the age of city buses using two criterion functions: profit per unit time and the coefficient of readiness of city buses to carry out the assigned transport tasks, and also to formulate sufficient conditions for the existence of the maximum of these functions. The results of studies of the model, developed on the basis of real data, can be the basis for decision-making in the analysed system of exploitation of urban transport means. In the developed mathematical model, the basis for the construction of the criterion function is the limit theorem for semi-Markov processes [30,31]. In the developed model, four types of implemented urban bus renewal processes are considered. The conditions for the existence of the extremum (maximum) of the criterion functions were formulated for the assumed assumptions. The theoretical considerations presented in the paper are illustrated by the results of calculations. The calculation examples have been developed on the basis of operational data obtained from a real urban bus operation system. In the first example, for

the estimated input data, the profit per unit time and the value of the readiness factor are maximised. In the second example, the analysed criterion functions are investigated, in case when the number (frequency) of preventive replacements will be higher than in the first example. In both the calculation examples, it is assumed that the time to failure of the technical object (city bus) has a Weibull distribution.

2. Description of the States of the Model of the Repair and Preventive Replacement System

In the study the object of research is the bus exploitation system of public transport, in which technical objects (city buses) can stay in one of the five states of the considered model of the renewal system (repairs and preventive replacements):

State 1—the state of operational availability of the technical object—it is the state, when the technical object (a city bus) is fully fit and supplied and can perform the assigned transport task in accordance with the accepted assumptions, i.e., in accordance with the accepted plan and schedule of realization of the courses (in the paper this state is considered as the state of failure-free work of the technical object);

State 2—the state of repair by the Technical Rescue Service without the loss of the course—this is the state, when a damaged technical object (a city bus) is repaired during the realization of the transport task (on the route), which is carried out by the Technical Rescue Service; this type of repair is carried out in a “short” time interval or in the gaps between the successive realizations of the transport task (between the next courses), i.e., it does not cause loss of the course in accordance with the accepted plan and schedule of transport realization—it is assumed that this condition does not cause any disturbance (break) in realization of assigned transport tasks (in the paper this condition is considered as a condition of minimal repair of a technical object);

State 3—the state of the repair by the Technical Rescue Unit with the loss of the course—it is the state, when a damaged technical object (a city bus) is repaired during the realization of the transport task (on the route), which is realized by the Technical Rescue Unit; this type of repair is realized in a “longer” time interval than the repair realized in the state 2 of the model, thus the repair causes the loss of the course in accordance with the plan and schedule of transportation realization—it is assumed that this state causes a disruption (break) in the realization of the assigned transportation tasks (in the paper this state is considered as the state of the minimal repair of the technical object);

State 4—state of repair at the Service Station—it is the state, when a damaged technical object (a city bus) is subject to repair at the specialized service and repair stands of the Service Station assigned for this purpose—it is assumed that this state causes a disturbance (break) in the realization of the assigned transport tasks (in the paper, this state is considered as the state of perfect repair of a technical object);

State 5—the state of preventive replacement—it is the state, when the technical object (a city bus) is subject to preventive maintenance after a specific hourly mileage and in accordance with the adopted operational strategy (according to the resour) in the technical object the elements and sub-assemblies are replaced—it is assumed, that this state does not cause disruption (break) in the realization of the assigned transport tasks (in the work this state is considered as the state of preventive age-replacement.).

Figure 1 shows a directed graph of the mapping of the state changes of the renewal system model (preventive repairs and replacements) of the considered technical objects (city buses).

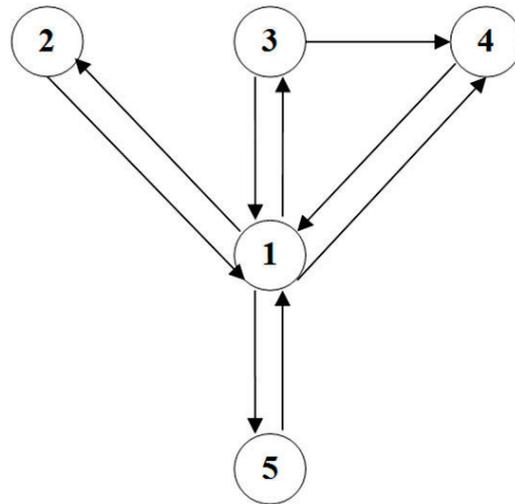


Figure 1. Directed graph representation of the state changes of a city bus renewal system model with state spaces $S = \{1, 2, 3, 4, 5\}$.

3. Determination of the Criterion Function

For the directed graph shown in Figure 1, a mathematical model was built assuming that it is a stochastic process $X(t)$. The mathematical model was developed using the theory of semi-Markov processes [30,31]. The paper considers a 5-state semi-Markov model of renewals (preventive repairs and replacements) with a state space $S = \{1, 2, 3, 4, 5\}$. If $X(t) = i$, then the technical object under consideration at time t is in state i .

In the case when the transition probabilities between the states of the modelled process are known, it is possible to determine the Markov chain inserted in the semi-Markov process. The transition matrix of the Markov chain for the model under consideration has the form

$$P = \begin{bmatrix} 0 & P_{12} & P_{13} & P_{14} & P_{15} \\ p_{21} & 0 & 0 & 0 & 0 \\ p_{31} & 0 & 0 & p_{34} & 0 \\ p_{41} & 0 & 0 & 0 & 0 \\ p_{51} & 0 & 0 & 0 & 0 \end{bmatrix} \quad (1)$$

where

$p_{ij}, i, j = 1, 2, 3, 4, 5$ —probability of transition from state i to state j .

To determine the limiting probabilities for a Markov chain, the following matrix system must be solved:

$$P^T \cdot \Pi = \Pi \begin{bmatrix} 0 & p_{21} & p_{31} & p_{41} & p_{51} \\ p_{12} & 0 & 0 & 0 & 0 \\ p_{13} & 0 & 0 & 0 & 0 \\ p_{14} & 0 & p_{34} & 0 & 0 \\ p_{15} & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix} \quad (2)$$

where

$\pi_i, i = 1, 2, 3, 4, 5$ —the limiting probability of a Markov chain inserted in a semi-markov process.

The Matrix System (2) can be replaced by a system of linear Equation (4) in which, in order to obtain an unambiguous solution, a normalization condition is introduced (3)

$$\sum_i \pi_i = 1 \quad (3)$$

then the system of linear Equation (4) takes the form

$$\begin{cases} \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1 \\ P_{12} \cdot \pi_1 = \pi_2 \\ P_{13} \cdot \pi_1 = \pi_3 \\ P_{14} \cdot \pi_1 + P_{34} \cdot \pi_3 = \pi_4 \\ P_{15} \cdot \pi_1 = \pi_5 \end{cases} \quad (4)$$

As a result of solving the system of linear Equation (4), formulas representing the limiting probabilities for the analysed Markov chain were obtained:

$$\pi_1 = \frac{1}{m} \pi_2 = \frac{P_{12}}{m} \pi_3 = \frac{P_{13}}{m} \pi_4 = \frac{P_{13} \cdot P_{34} + P_{14}}{m} \pi_5 = \frac{P_{15}}{m} \quad (5)$$

where

$$m = 1 + P_{12} \cdot P_{13} \cdot (1 + P_{34}) + P_{14} + P_{15}$$

In this paper, a semi-Markov model of renewals (preventive corrective maintenance and age-replacement) is analysed. A 5-state semi-Markov process $X(t)$ with state space $S = \{1, 2, 3, 4, 5\}$ is considered. By $z_i, i = 1, 2, 3, 4, 5$ means the profit (cost) per unit (per time unit) for the state i . It is assumed in the paper, that $z_1 > 0, z_i < 0$ for $i = 2, 3, 4, 5$. This means that if the technical object is in state 1, then a profit is generated, while if the technical object is in state $i = 2, 3, 4, 5$, then a cost (loss) is generated. In the paper [28] it was proved, that the summary profit (loss) per unit of time generated in the system is expressed by the formula

$$Z = \frac{\sum_i \pi_i \cdot ET_i \cdot z_i}{\sum_i \pi_i \cdot ET_i} \quad (6)$$

where

$ET_i, i = 1, 2, 3, 4, 5$ —average time spent in state i .

A technical object is subject to renewal at age T or when it is damaged, whatever comes first. By $T_1(x)$ we define the time to replace or damage (repair) a technical object. The variable $T_1(x)$ is defined as follows

$$T_1(x) = \begin{cases} T_1, & \text{gdy } T_1 < x \\ x, & \text{gdy } T_1 \geq x \end{cases} \quad (7)$$

It is assumed that after time x , if the technical object has not failed, it transitions to the preventive replacement state. The process of changing states $i = 1, 2, 3, 4, 5$, given preventive replacement after time x is a new semi-Markov process with a matrix $P(x)$ of transition probabilities of the Markov chain inserted into the semi-Markov process. With respect to the matrix $P(1)$ shown above, only the first row of the matrix P changes, then the matrix $P(x)$ takes the form

$$P(x) = \begin{bmatrix} 0 & P_{12}(x) & P_{13}(x) & P_{14}(x) & P_{15}(x) \\ P_{21} & 0 & 0 & 0 & 0 \\ P_{31} & 0 & 0 & P_{34} & 0 \\ P_{41} & 0 & 0 & 0 & 0 \\ P_{51} & 0 & 0 & 0 & 0 \end{bmatrix} \quad (8)$$

while the limiting probabilities determined for a Markov chain (determined analogously to Formulae (5) can be presented as:

$$\begin{aligned} \pi_1(x) &= \frac{1}{n} \\ \pi_2(x) &= \frac{p_{12}(x)}{n} \\ \pi_3(x) &= \frac{p_{13}(x)}{n} \\ \pi_4(x) &= \frac{p_{13}(x) \cdot p_{34} + p_{14}(x)}{n} \\ \pi_5(x) &= \frac{p_{15}(x)}{n} \end{aligned} \tag{9}$$

where

$$n = 1 + p_{12}(x) \cdot p_{13}(x) \cdot (1 + p_{34}) + p_{14}(x) + p_{15}(x)$$

Based on the paper [28], the Criterion Function (6) is of the form:

$$Z = g(x) = \frac{\pi_1(x) \cdot ET_1(x) \cdot z_1 + \pi_2(x) \cdot ET_2 \cdot z_2 + \pi_3(x) \cdot ET_3 \cdot z_3 + \pi_4(x) \cdot ET_4 \cdot z_4 + \pi_5(x) \cdot ET_5 \cdot z_5}{\pi_1(x) \cdot ET_1(x) + \pi_2(x) \cdot ET_2 + \pi_3(x) \cdot ET_3 + \pi_4(x) \cdot ET_4 + \pi_5(x) \cdot ET_5} \tag{10}$$

where

$ET_1(x)$ —average value of time spent in state 1, calculated from the formula [30,31]

$$\begin{aligned} ET_1(x) &= \int_0^x dF_1(T) + xP\{T_1 \geq x\} \\ ET_1(x) &= \int_0^x R_1(T)dT \end{aligned} \tag{11}$$

ET_2, ET_3, ET_4 and ET_5 ,—average values of times spent in states 2, 3, 4 and 5, respectively. In particular, based on the paper [28], it can be written:

$$\begin{aligned} p_{12}(x) &= p_{12} \cdot F_{12}(x) \\ p_{13}(x) &= p_{13} \cdot F_{13}(x) \\ p_{14}(x) &= p_{14} \cdot F_{14}(x) \\ p_{15}(x) &= p_{15} \cdot F_{12}(x) + R_1(x) \end{aligned} \tag{12}$$

where:

$F_{1j}(x), j = 2, 3, 4, 5$ —conditional distributions of the time spent in state 1, provided that the next state is state j , defined as follows [30,31]

$$F_{ij}(t) = P\{\tau_{k+1} - \tau_k < t | X(\tau_{k+1}) = j, X(\tau_k) = i\}, \text{ dla } i, j = 1, 2, 3, 4, 5 \tag{13}$$

$R_1(x) = 1 - F_1(x)$ —random variable reliability function T_1 .

Additionally, in order to simplify further considerations, it has been assumed that the following equations are true

$$F_{12}(x) = F_{13}(x) = F_{14}(x) = F_{15}(x) = F_1(x) \tag{14}$$

Considering the above, the Criterion Function (10) is expressed by the formula

$$\begin{aligned} g(x) &= \frac{ET_1(x) \cdot z_1 + p_{12} \cdot F_1(x) \cdot ET_2 \cdot z_2 + p_{13} \cdot F_1(x) \cdot ET_3 \cdot z_3 + [(p_{13} \cdot p_{34} + p_{14}) \cdot F_1(x)] \cdot ET_4 \cdot z_4 +}{ET_1(x) + p_{12} \cdot F_1(x) \cdot ET_2 + p_{13} \cdot F_1(x) \cdot ET_3 + [(p_{13} \cdot p_{34} + p_{14}) \cdot F_1(x)] \cdot ET_4 +} \\ &\quad + \frac{[1 - (p_{12} + p_{13} + p_{14}) \cdot F_1(x)] \cdot ET_5 \cdot z_5}{+ [1 - (p_{12} + p_{13} + p_{14}) \cdot F_1(x)] \cdot ET_5} \end{aligned}$$

or after regrouping, it can be represented as

$$g(x) = \frac{z_1 \cdot ET_1(x) + [p_{12} \cdot ET_2 \cdot z_2 + p_{13} \cdot ET_3 \cdot z_3 + (p_{13} \cdot p_{34} + p_{14}) \cdot ET_4 \cdot z_4 + ET_1(x) + [p_{12} \cdot ET_2 + p_{13} \cdot ET_3 + (p_{13} \cdot p_{34} + p_{14}) \cdot ET_4 + \frac{-(p_{12} + p_{13} + p_{14}) \cdot ET_5 \cdot z_5 \cdot F_1(x) + ET_5 \cdot z_5}{-(p_{12} + p_{13} + p_{14}) \cdot ET_5 \cdot F_1(x) + ET_5}]}{ET_1(x) + [p_{12} \cdot ET_2 + p_{13} \cdot ET_3 + (p_{13} \cdot p_{34} + p_{14}) \cdot ET_4 + \frac{-(p_{12} + p_{13} + p_{14}) \cdot ET_5 \cdot z_5 \cdot F_1(x) + ET_5 \cdot z_5}{-(p_{12} + p_{13} + p_{14}) \cdot ET_5 \cdot F_1(x) + ET_5}}$$

Representing the numerator and the denominator of the criterion function as follows:

$$L(x) = A_1 \cdot ET_1(x) + B_1 \cdot F_1(x) + C_1$$

$$M(x) = A \cdot ET_1(x) + B \cdot F_1(x) + C$$

the Criterion Function (10) can be represented analogously as

$$g(x) = \frac{A_1 \cdot ET_1(x) + B_1 \cdot F_1(x) + C_1}{A \cdot ET_1(x) + B \cdot F_1(x) + C}$$

where:

$$A_1 = z_1$$

$$B_1 = p_{12} \cdot ET_2 \cdot z_2 + p_{13} \cdot ET_3 \cdot z_3 + (p_{13} \cdot p_{34} + p_{14}) \cdot ET_4 \cdot z_4 - (p_{12} + p_{13} + p_{14}) \cdot ET_5 \cdot z_5$$

$$C_1 = ET_5 \cdot z_5$$

$$A = 1$$

$$B = p_{12} \cdot ET_2 + p_{13} \cdot ET_3 + (p_{13} \cdot p_{34} + p_{14}) \cdot ET_4 - (p_{12} + p_{13} + p_{14}) \cdot ET_5$$

$$C = ET_5$$

4. Conditions for the Existence of a Maximum of the Criterion Function

4.1. Maximum of the Criterion Function—General Analysis

The conditions for the existence of an extremum (maximum) of the Criterion Function (10) will be formulated depending on the parameters of the developed semi-Markov model of the renewal system (repairs and preventive replacements), i.e., the elements of the matrix of probabilities of changes in the states of the model $P = [p_{ij}]$, $i, j = 1, 2, 3, 4, 5$, the average staying times in the states of the model ET_i , $i = 1, 2, 3, 4, 5$, and the unit profits (costs) generated in the states of the model z_i , $i = 1, 2, 3, 4, 5$. The considered parameters are the input data of the model, and their values depend on the category and type of the analysed technical objects, the adopted operation strategy, and specific operating conditions in which the repair and preventive replacement processes are carried out.

The assumptions regarding the values of the parameters of the examined system are defined below. The adopted assumptions must take into account the real relations occurring between the parameters characterizing the implemented processes of repair of the damaged technical objects and the preventive replacement:

- Z1: $z_1 > 0, z_2 < 0, z_3 < 0, z_4 < 0, z_5 < 0$; means that a technical object staying in state 1 generates profit (+), whereas staying in states 2, 3, 4 and 5 generates costs (-);
- Z2: $ET_2 < ET_3$; means, that the value of the average time of repair carried out by the Technical Service units without the loss of the course is lower than the value of the average time of repair carried out by the Technical Service units with the loss of the course;
- Z3: $ET_2 < ET_4$; means that the value of the average repair time performed by the Technical Service units without the loss of the course (minimal repair) is lower than the value of the average repair time performed at the Service Stations (perfect repair));
- Z4: $ET_3 < ET_4$; means that the value of the average repair time realized by the Technical Service units with the loss of the course (minimal repair) is lower than the value of the average repair time realized at the Service Stations (perfect repair);

- Z5: $ET_4 > ET_5$; means that the value of the average time of repair performed at Service Stations (precise repair) is higher than the value of the average time of preventive replacement;
- Z6: $ET_2 < ET_5$; means, that the value of the average time of repair carried out by the Technical Service units without the loss of the course (minimal repair) is lower than the value of the average time of preventive replacement;
- Z7: $z_2 < z_3$; means, that the unit cost generated in the state 2 (the state of repair realized by the Technical Service units without the loss of the course) is lower than the unit cost generated in the state 3 (the state of repair realized by the Technical Service unit with the loss of the course);
- Z8: $z_2 < z_4$; means, that the unit cost generated in the state 2 (the state of the repair carried out by the Technical Service units without the loss of the course) is lower than the unit cost generated in the state 4 (the state of the repair carried out at the Service Stations);
- Z9: $z_3 < z_4$; means, that the unit cost generated in the state 3 (the state of the repair carried out by the Technical Service units with the loss of the course) is lower than the unit cost generated in the state 4 (the state of the repair carried out at the Service Stations);
- Z10: $z_4 > z_5$; means, that the unit cost generated in the state 4 (the state of repair performed at the Service Stations) is higher than the unit cost generated in the state 5 (the state of preventive replacement);
- Z11. $z_2 < z_5$; means, that the unit cost generated in the state 2 (the state of repair carried out by the Technical Service units without the loss of the course) is lower than the unit cost generated in the state 5 (the state of preventive replacement).

The above assumptions do not define the relationship between the state of repair performed by the Technical Service unit with loss of the course (state 3) and the state of preventive replacement (state 5). In the analysed system it is very difficult to unambiguously define the relation between the average values of ET_3 and ET_5 times and unit costs z_3 and z_5 . However, based on the results of research on other systems of exploitation of this class of technical objects (means of transport), an additional assumption can be made regarding the unit costs generated in states 3 and 5:

- Z12: $z_3 < z_5$; means that the unit cost generated in the state 3 (the state of repair performed by the Technical Service unit with the loss of the course) is lower than the unit cost generated in the state 5 (the state of preventive replacement).

In the subsequent part of the paper, the following coefficients have been introduced to formulate the conditions for the existence of the extremum (maximum) of the criterion function (10):

$$\alpha = AB_1 - A_1B = B_1 - z_1B$$

$$\beta = A_1C - AC_1 = z_1C - C_1$$

$$\gamma = B_1C - BC_1$$

where:

$$A_1 = z_1$$

$$B_1 = p_{12}ET_2z_2 + p_{13}ET_3z_3 + (p_{13}p_{34} + p_{14})ET_4z_4 - (p_{12} + p_{13} + p_{14})ET_5z_5$$

$$C_1 = ET_5z_5$$

$$A = 1$$

$$B = p_{12}ET_2 + p_{13}ET_3 + (p_{13}p_{34} + p_{14})ET_4 - (p_{12} + p_{13} + p_{14})ET_5$$

$$B = p_{12}ET_2 + p_{12}p_{23}ET_3 - p_{12}ET_4$$

$$C = ET_5$$

The coefficients α , β , and γ play an important role in formulating sufficient conditions for the existence of the extremes of a criterion function. For this purpose, the following formulates the sufficient conditions for the inequalities to be true $\alpha < 0$, $\beta > 0$, $\gamma < 0$.

In regard to the above:

- the coefficient α is determined by the formula

$$\alpha = p_{12}ET_2(z_2 - z_1) + p_{13}ET_3(z_3 - z_1) + (p_{13}p_{34} + p_{14})ET_4(z_4 - z_1) + (p_{12} + p_{13} + p_{14})ET_5(z_1 - z_5) \quad (15)$$

The inequality $\alpha < 0$ is equivalent to the inequality

$$(p_{13} + p_{34} + p_{14}) \frac{ET_4}{z_1 - z_4} > p_{12}ET_2(z_2 - z_1)/(z_1 - z_4) + p_{13}ET_3(z_3 - z_1)/(z_1 - z_4) + (p_{12} + p_{13} + p_{14})ET_5(z_1 - z_5)/(z_1 - z_4) \quad (16)$$

- the coefficient β is determined by the formula

$$\beta = ET_5(z_1 - z_5) \quad (17)$$

Based on the assumption Z1 made, it follows that $\beta > 0$.

- the coefficient γ is determined by the formula

$$\gamma = [p_{12}ET_2(z_2 - z_5) + p_{13}ET_3(z_3 - z_5) + (p_{13}p_{34} + p_{14})ET_4(z_4 - z_5)]ET_5 \quad (18)$$

The inequality $\gamma < 0$ is equivalent to the inequality

$$(p_{13} + p_{34} + p_{14})ET_4 > p_{12}ET_2(z_2 - z_5)/(z_5 - z_4) + p_{13}ET_3(z_3 - z_5)/(z_5 - z_4) \quad (19)$$

In practice, it is difficult to unambiguously determine what is the relation between the state of repair by the Technical Service unit with course loss (state 3) and the state of preventive replacement (state 5), i.e., the relation between the average values of the times ET_3 and ET_5 and between the unit costs z_3 and z_5 is unknown. With respect to the coefficient γ , inequality (19) must be considered similarly to inequality (16) regarding the coefficient α . In this case, the right-hand sides of inequalities (16) and (19) are denoted by δ_1 and δ_2 , respectively. Let $\delta = \max\{\delta_1, \delta_2\}$, then the condition $(p_{13} p_{34} + p_{14}) ET_4 > \delta$ and formulas (15), (16), (18) and (19) imply the inequalities $\alpha < 0$, $\gamma < 0$. From this, the following conclusion can be made:

Conclusion 1. *If $p_{34} > [\delta/(ET_4 - p_{14})/p_{13}]$, then the inequalities $\alpha < 0$, $\gamma < 0$ are true.*

4.2. The Maximum of the Criterion Function—The Distributions of the Random Variable of the IFR Classes and MTFR

In this subsection of the paper, sufficient conditions for the existence of the maximum of the Criterion Function (10) will be formulated in two cases. In the first case, the considerations apply to a class of random variable distributions for which the time to failure of a technical object T_1 is assumed to be a random variable with increasing damage intensity function $\lambda_1(t)$, i.e., $T_1 \in$ IFR (Increasing Failure Rate). In the second case, a class of random variable distributions with a unimodal failure intensity function, the $T_1 \in$ MTFR (Mean Time to Failure or Repair), is considered. The results of testing the properties of the random variable distributions of the MTFR class are presented in detail in the papers [32–34].

Conclusion 2. *If $T_1 \in$ IFR, $\lambda_1(t)$ is differentiable, $\alpha < 0$, $\beta > 0$, $\gamma < 0$, $\beta + \gamma f_1(0+) > 0$, $\lambda_1(\infty) \alpha ET_1 + \beta - \alpha < 0$, then the criterion function $g(x)$ reaches its maximum value.*

Proof of Conclusion 2. The derivative of the criterion function $g(x)$ has the form

$$g'(x) = \{\alpha[f_1(x)ET_1(x) - R_1(x)F_1(x)] + \beta R_1(x) + \gamma f_1(x)\}/M^2(x)$$

where $M(x)$ is the denominator of the criterion function $g(x)$.

It is known, that if the time to failure T_1 belongs to the class of distributions of the random variable MTFR, then the equality $H(x) = \lambda_1(x) ET_1(x) - F_1(x) \geq 0$ for $x \geq 0$ is true. The class of distributions of the random variable MTFR has been studied in the papers [33,34]. Some lifetime distributions with unimodal damage intensity function belong to the class of MTFR [33,34]. From the fact, that the derivative $H'(x) = \lambda_1'(x) ET_1(x)$, it follows that if the damage intensity function $\lambda_1(t)$ increases, the function $H(x)$ also increases. The class of distributions of a random variable with a non-decreasing damage intensity function (IFR) is contained in the MTFR class. The sign of the derivative is the same as the sign of the function

$$h(x) = \alpha[\lambda_1(x)ET_1(x) - F_1(x)] + \beta + \gamma\lambda_1(x)$$

It is known, that $H(0+) = 0$, hence $h(0+) = \beta + \gamma f_1(0+) > 0$. From the fact that $\alpha < 0$, $\beta > 0$, $\gamma < 0$ and the function $H(x)$ increases, it follows that the function $h(x)$ decreases from the value $h(0+) = \beta + \gamma f_1(0+) > 0$ to the value $h(\infty) = \lambda_1(\infty) \alpha ET_1 + \beta - \alpha < 0$. It follows from this that the derivative of $g'(x)$ changes sign exactly once from “+” to “-”. Hence, it is concluded that the criterion function $g(x)$ reaches exactly one maximum.

If $\lambda_1(\infty) = \infty$, then the following conditions suffice for the existence of the maximum of the criterion function $g(x)$: $T_1 \in \text{IFR}$, differentiability of $\lambda_1(t)$, $\alpha < 0$, $\beta > 0$, $\gamma < 0$, $\beta + \gamma f_1(0+) > 0$. An example of such a distribution of a random variable is a Weibull distribution with an increasing damage intensity function. □

From the conclusions 1 and 2, the following sufficient condition for the existence of a maximum of the criterion function follows:

Conclusion 3. *If $T_1 \in \text{IFR}$, $\lambda_1(t)$ is differentiable, $\beta + \gamma \lambda_1(0+) > 0$, $p_{34} > [\delta/(ET_4 - p_{14})/p_{13}]$, $\lambda_1(\infty) \alpha ET_1 + \beta - \alpha < 0$, then the criterion function $g(x)$ reaches the maximum value.*

A sufficient condition for the existence of the asymptotic maximum of the availability factor is formulated below. To obtain the availability factor from the criterion function $g(x)$, it suffices to assume the following conditions: $z_1 = 1$, $z_2 = z_3 = z_4 = z_5 = 0$. After considering these conditions in formula (10), $B_1 = 0$, $C_1 = 0$ are obtained. Hence, based on (6), (7) and (9) for α , β , γ , one can calculate:

$$\begin{aligned} \alpha &= -B = -p_{12}ET_2 - p_{13}ET_3 - (p_{13} + p_{34} + p_{14})ET_4 + (p_{12} + p_{13} + p_{14})ET_5 \\ \beta &= C = ET_5; \beta > 0 \\ \gamma &= 0 \end{aligned}$$

The inequality $\alpha < 0$ is equivalent to the inequality:

$$p_{34} > [p_{12}(ET_5 - ET_2) + p_{13}(ET_5 - ET_3) + p_{14}(ET_5 - ET_4)] / p_{13}ET_4$$

Given that $\beta > 0$ and $\gamma = 0$, one can now formulate a sufficient condition for the existence of a maximum of the availability factor.

Conclusion 4. *If $T_1 \in \text{IFR}$, $\lambda_1(t)$ is differentiable, $\lambda_1(\infty) \alpha ET_1 + \beta - \alpha < 0$, $p_{34} > [p_{12} (ET_5 - ET_2) + p_{13} (ET_5 - ET_3) + p_{14} (ET_5 - ET_4)]/p_{13} ET_4$, then the availability factor reaches exactly one maximum value.*

Proof of Conclusion 4. For the availability factor, the derivative of the criterion function has the form:

$$g'(x) = \{\alpha[f_1(x)ET_1(x) - R_1(x)F_1(x)] + \beta R_1(x)\} / M^2(x)$$

where $M(x)$ is the denominator of the criterion function $g(x)$.

If the damage intensity function $\lambda_1(t)$ is increasing, then the function $H(x)$ is increasing. The sign of the derivative is the same as the sign of the function:

$$h(x) = \alpha[\lambda_1(x)ET_1(x) - F_1(x)] + \beta$$

It is known that $H(0+) = 0$, hence $h(0+) = \beta > 0$.

From the fact that $p_{34} > [p_{12}(ET_5 - ET_2) + p_{13}(ET_5 - ET_3) + p_{14}(ET_5 - ET_4)]/p_{13}ET_4$, it follows that $\alpha < 0$ and the function $h(x)$ decreases from $h(0+) = \beta > 0$ to $h(\infty)$. If $h(\infty) = \lambda_1(\infty)\alpha ET_1 + \beta - \alpha < 0$, it means that the derivative $g'(x)$ changes sign exactly once from “+” to “-”. Hence, it is concluded, that the availability factor $g(x)$ reaches exactly one maximum.

If $\lambda_1(\infty) = \infty$, then the following conditions are sufficient for the existence of a maximum of the availability factor:

$$T_1 \in iRF, p_{34} > [p_{12}(ET_5 - ET_2) + p_{13}(ET_5 - ET_3) + p_{14}(ET_5 - ET_4)]/p_{13}ET_4$$

□

5. Exemplary Calculation Results

Example 1. In Figure 2, the plots of the criterion function $g(x)$ are shown when $g(x)$ represents profit per unit time, and in Figure 3, when $g(x)$ represents availability for transportation tasks.

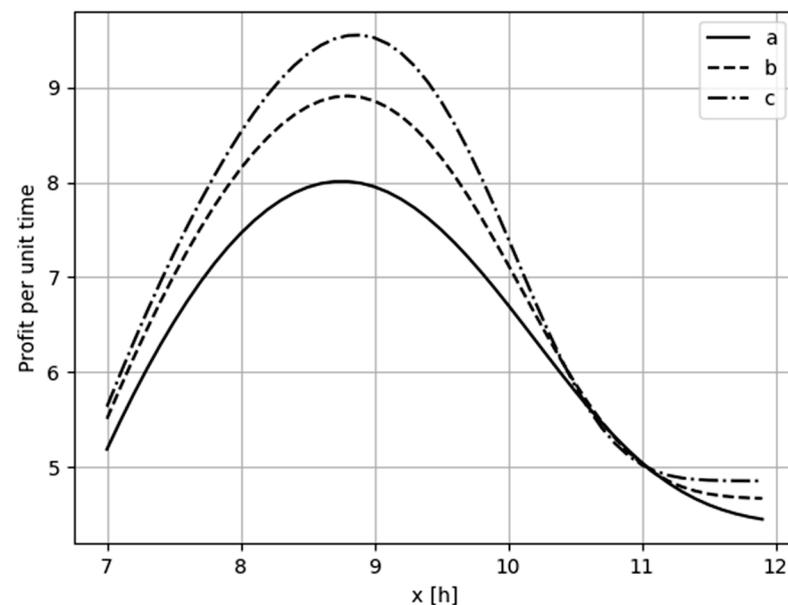


Figure 2. The graphs of the function $g(x)$ —profit per unit time as a function of time to preventive maintenance x [h], determined for the Weibull distribution with the following parameter values: scale = 10 and shape = 9 (curve a), shape = 11.5 (curve b), shape = 14 (curve c), the values of the parameters of the distribution of the random variable ET_1 were determined for the considered values of the service life of the tested city buses, respectively, 9.25, 9.5, 9.75 [h].

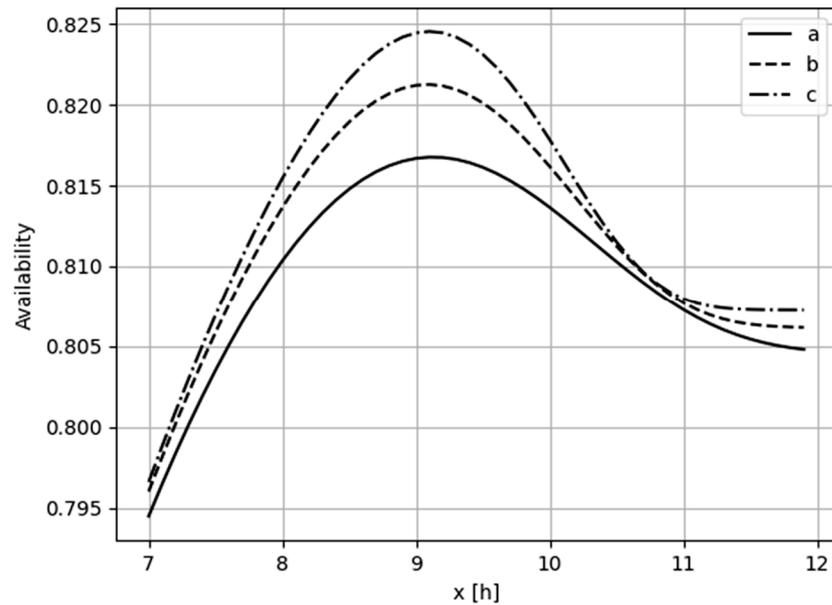


Figure 3. The graphs of the function $g(x)$ —availability to perform transport tasks as a function of time to preventive maintenance x [h], determined for the Weibull distribution with the following parameter values: scale = 10 and shape = 9 (curve a), shape = 11.5 (curve b), shape = 14 (curve c), the values of the parameters of the distribution of the random variable ET_1 were determined for the considered values of the service life of the tested city buses, respectively, 9.25, 9.5, 9.75 [h].

The calculations were conducted for the following data:

- (1) values of the matrix of probabilities of changes of states of the model P :

$$P = \begin{bmatrix} 0 & 0.2033 & 0.0811 & 0.4124 & 0.3032 \\ 1 & 0 & 0 & 0 & 0 \\ 0.7858 & 0 & 0 & 0.2142 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- (2) the average values of the technical object residence times in the model states in [h]: $ET_2 = 0.389$, $ET_3 = 1.538$, $ET_4 = 3.621$, $ET_5 = 1.783$; for the airworthiness time (time to failure) ET_1 , a Weibull distribution was assumed for which the value of the scale parameter = 10; three cases were analysed when the value of the shape parameter of the Weibull distribution is, respectively $\text{shape} \in \{9, 11.5, 14\}$;
- (3) the average values of profits (costs) per unit time in each state of the model in [PLN/h]: $z_1 = 38$, $z_2 = -71$, $z_3 = -117$, $z_4 = -143$, $z_5 = -121$. Both when the criterion function $g(x)$ denotes the profit per unit time (Figure 2) and when $g(x)$ denotes the availability to carry out transport tasks (Figure 3) the criterion function reaches its maximum value. In each of the three cases analysed, for particular values of the shape parameter of the Weibull distribution, there is an optimal value of time to preventive replacement x [h]. Based on the analysis of the value of x_{\max} , for which the criterion function $g(x)$ reaches its maximum value, it can be concluded that as the value of the shape parameter increases, the value of x_{\max} and the maximum value of the criterion function $g(x)$.

Example 2. Figures 4 and 5 show, respectively, the graphs of the criterion function $g(x)$ in the case where $g(x)$ represents profit per time unit and where $g(x)$ represents availability to complete transportation tasks. The calculations were performed for the data of Example 1, assuming that the uptime (time to failure) ET_1 has a Weibull distribution, for which the value of the scale

parameter = 10 and the shape parameter = 11.5. The graphs show four cases: case *d*—when the number of preventive replacements is the same as in Example 1, and cases *a*, *b*, and *c*, when the number of preventive replacements is increased by 10%, 20%, and 30%, respectively, with respect to case *d*.

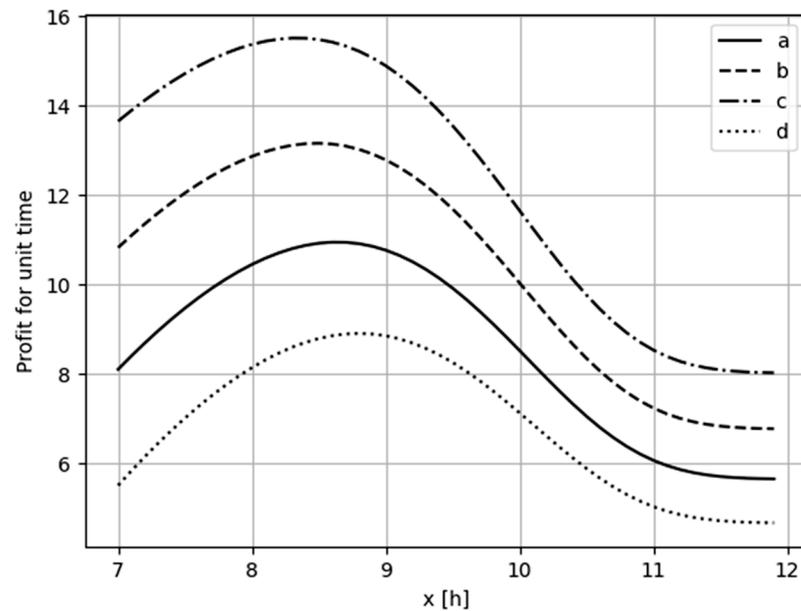


Figure 4. The graphs of the function $g(x)$ —profit per time unit as a function of time to preventive replacement x [h], determined when the number of preventive replacements is as in Example 1 (curve *d*) and when the number of preventive replacements is increased by 10%, 20%, and 30%, respectively (curves *a*, *b*, and *c*).

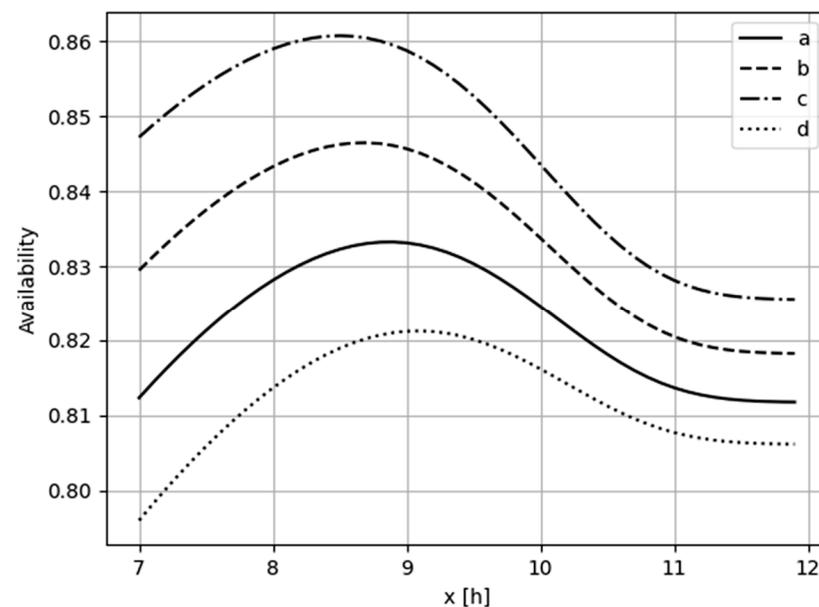


Figure 5. Graphs of the function $g(x)$ —availability to perform transportation tasks as a function of time to preventive replacement x [h], determined when the number of preventive replacements is as in example 1 (curve *d*) and when the number of preventive replacements is increased by 10%, 20%, and 30%, respectively (curves *a*, *b*, and *c*).

Based on the analysis of the graphs presented in Figures 4 and 5, it can be seen that as the number of performed preventive exchanges increases (their frequency in-

creases), the value of the criterion function $g(x)$ increases, both in the case of profit per time unit and availability to perform transport tasks, and the maximum value of the function $g(x)$ is reached for the increasingly smaller values of x_{\max} (the optimum time value for preventive replacement).

6. Conclusions

The mathematical model presented in the article makes it possible to determine the optimum values of the preventive replacement time in such a way that the criterion functions (profit per unit time and readiness to carry out transport tasks) reached the maximum values. On the basis of the analysis of the results obtained, it can be noted that for the considered input data of the model, the increase in the value of the serviceability time of the examined city buses (increase in the value of the shape parameter of the Weibull distribution) causes an increase in the value of the considered criterion functions (both profit per time unit and readiness), while increasing the optimum time to preventive replacement (Figures 2 and 3, respectively). Decreasing the time to preventive replacement (increasing the frequency of preventive replacements) by 10, 20 and 30%, causes a significant increase in the values of the criterion functions: profit per unit time (from 8.9 to 15.5 [PLN/h]) and readiness to carry out transport tasks from 0.821 to above 0.86 (Figures 4 and 5, respectively). In the paper, the criterion functions are considered over an infinite time horizon. The formulation of stronger conditions requires the establishment of relations between the average stay times of a technical object and the unit profits (costs) in the states of repair by the Technical Emergency Service with the loss of course (state 3) and the preventive replacement (state 5). It has been proved that under general assumptions the criterion functions considered in the paper have exactly one extremum (maximum). On the basis of the conducted analysis, sufficient conditions were formulated for the existence of the maximum of these functions when the time to failure of a technical object is a random variable with an increasing damage intensity function. The assumptions adopted in the model and the formulated conditions define the relations between the input parameters of the developed model and verify the possibility of applying a specific set of input data for determining the optimum preventive replacement times (determining the maximum of the criterion functions). The presented research results constitute the next stage of works on modelling the exploitation systems of technical objects, in which preventive replacements by age are carried out. In the next stages, the models of preventive replacements will be developed for technical objects of other classes than the means of transport, e.g., for power equipment. These models will use both the criterion functions describing economic efficiency (e.g., profit per unit time), operational and technical efficiency (e.g., readiness) and safety (e.g., risk of loss). On this basis, it is planned to develop a comprehensive control method for the subsystem for ensuring the serviceability of technical objects using decision-making semi-Markov processes and non-deterministic methods for determining optimal (sub-optimal) solutions).

Author Contributions: Conceptualization, K.M. and S.B.; methodology, K.M. and A.S.; software, A.N.; validation, K.M., investigation, S.B. and A.S.; data curation, S.B. and A.S.; writing—original draft preparation, K.M., writing—review and editing, S.B.; visualization, A.S. and A.N.; supervision, K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Migawa, K. Control of Availability in the System of Operation of Means of Transport. Ph.D. Thesis, University of Science and Technology, Bydgoszcz, Poland, 2013.
2. Knopik, L.; Migawa, K.; Wdzięczny, A. Profit optimalization in operation systems. *Pol. Marit. Res.* **2016**, *23*, 93–98.
3. Migawa, K.; Knopik, L.; Wawrzyniak, S. Application of genetic algorithm to control the availability of technical systems. In Proceedings of the 22nd International Conference Engineering Mechanics, Svratka, Czech Republic, 9–12 May 2016; pp. 386–389.
4. Knopik, L.; Migawa, K. Optimal age-replacement policy for non-repairable technical objects with warranty. *Maint. Reliab.* **2017**, *19*, 172–178.
5. Morse, P.C. *Queues, Inventories, and Maintenance*; Wiley: Hoboken, NJ, USA, 1958.
6. Barlow, R.; Hunter, L. Optimum preventive maintenance policies. *Oper. Res.* **1960**, *8*, 90–110. [[CrossRef](#)]
7. Nakagawa, T.; Kowada, M. Analysis of a system with minimal repair and its application to replacement policy. *Eur. J. Oper. Res.* **1983**, *12*, 176–182.
8. Brown, M.; Proschan, F. Imperfect repair. *J. Appl. Probab.* **1983**, *20*, 851–859. [[CrossRef](#)]
9. Fontenot, R.A.; Proschan, F. Some imperfect maintenance models. In *Reliability Theory and Models*; Abdel-Hameed, M., Ed.; Academic Press: Orlando, FL, USA, 1984.
10. Pham, H.; Wang, H. Imperfect maintenance. *Eur. J. Oper. Res.* **1996**, *94*, 425–438. [[CrossRef](#)]
11. Tadj, L.; Ouali, M.S.; Yacount, S.; Ait-Kadi, D. A Survey of Replacement Models with Minimal Repair. In *Replacement Models with Minimal Repair*; Springer: London, UK, 2011.
12. Chan, P.K.W.; Downs, T. Two criteria for preventive maintenance. *IEEE Trans. Reliab.* **1978**, *27*, 272–273. [[CrossRef](#)]
13. Goel, L.R.; Agnihotri, R.K.; Gupta, R. Two-unit redundant system with inspection and adjustable rates. *Microelectron. Reliab.* **1991**, *31*, 11–14. [[CrossRef](#)]
14. Zequeira, R.; Brenguer, C. Periodic imperfect preventive maintenance with two categories of competing failure modes. *Reliab. Eng. Syst.* **2006**, *91*, 460–468.
15. Huynh, K.; Castro, I.; Barros, A.; Brenguer, C. Modeling age-based maintenance strategies with minimal repairs for systems subject to competing failure modes due to degradation and shocks. *Eur. J. Oper. Res.* **2012**, *218*, 140–151.
16. El-Ferik, S. Economic production lot-sizing for an unreliable machine under imperfect age-based maintenance policy. *Eur. J. Oper. Res.* **2008**, *186*, 150–163.
17. Gilardoni, G.L.; de Toledo, M.L.G.; Freitas, M.A.; Colosimo, E.A. Dynamics of an optimal maintenance policy for imperfect repair models. *Eur. J. Oper. Res.* **2016**, *248*, 1104–1112. [[CrossRef](#)]
18. Yang, L.; Ye, Z.S.; Lee, C.G.; Yang, S.F.; Peng, R. A two-phase preventive maintenance policy considering imperfect repair and postponed replacement. *Eur. J. Oper. Res.* **2019**, *274*, 966–977.
19. Wu, T.; Ma, X.; Yang, L.; Zhao, Y. Proactive maintenance scheduling in consideration of imperfect repairs and production wait time. *J. Manuf. Syst.* **2019**, *53*, 183–194. [[CrossRef](#)]
20. Cleroux, R.; Dubuc, S.; Tilquin, C. The age replacement problem with minimal repair and random repair costs. *Oper. Res.* **1979**, *27*, 1158–1167. [[CrossRef](#)]
21. Sheu, S.H.; Chien, Y.H. Optimal age-replacement policy of a system subject to hocks with random lead-time. *Eur. J. Oper. Res.* **2004**, *159*, 132–144. [[CrossRef](#)]
22. Chien, Y.H.; Sheu, S.H. Extended optimal age-replacement policy with minimal repair of a system subject to shocks. *Eur. J. Oper. Res.* **2006**, *174*, 169–181. [[CrossRef](#)]
23. Jhang, J.P.; Sheu, S.H. Opportunity-base dagere placement policy with minimal repair. *Reliab. Eng. Syst. Saf.* **1999**, *64*, 339–344. [[CrossRef](#)]
24. Sheu, S.H.; Yeh, R.H.; Lin, Y.B.; Juang, M.G. A Bayesian perspective on age replacement with minimal repair. *Reliab. Eng. Syst. Saf.* **1999**, *65*, 55–64. [[CrossRef](#)]
25. Lim, J.H.; Lu, K.L.; Park, D.H. Bayesian imperfect repair model. *Commun. Stat. Theory Methods* **1998**, *27*, 965–984.
26. Lim, J.H.; Qu, J.; Zuo, M.J. Age replacement policy based on imperfect with random probability. *Reliab. Eng. Syst.* **2016**, *149*, 24–33.
27. Zhou, Y.; Kou, G.; Xiao, H.; Peng, Y.; Alsaadi, F.E. Sequential imperfect preventive maintenance model with failure intensity reduction with an application to urban buses. *Reliab. Eng. Syst.* **2020**, *198*, 965–984. [[CrossRef](#)]
28. Knopik, L.; Migawa, K. Multi-state model of maintenance policy. *Maint. Reliab.* **2018**, *20*, 125–130. [[CrossRef](#)]
29. Knopik, L.; Migawa, K. Semi-Markov system model for minimal repair maintenance. *Maint. Reliab.* **2019**, *21*, 256–260. [[CrossRef](#)]
30. Grabski, F. *Semi-Markowskie Modele Niezawodności i Eksploatacji [Semi-Markov Models of Reliability and Maintenance]*; IBS PAN: Warszawa, Poland, 2002.
31. Grabski, F. *Semi-Markov Processes: Applications in System Reliability and Maintenance*; Elsevier: Amsterdam, The Netherlands, 2014.
32. Asha, G.; Unnikrishanan, N.N. Reliability properties of means time to failure in age replacement models. *Int. J. Reliab. Qual. Saf. Eng.* **2010**, *17*, 15–26. [[CrossRef](#)]
33. Knopik, L. Some results on ageing class. *Control. Cybern.* **2005**, *34*, 1175–1180.
34. Knopik, L. Characterization of a class of lifetime distributions. *Control. Cybern.* **2006**, *35*, 407–414.

Article

Long-Range Dependence and Multifractality of Ship Flow Sequences in Container Ports: A Comparison of Shanghai, Singapore, and Rotterdam

Chan-Juan Liu ^{1,*}, Jinran Wu ^{2,*}, Harshanie Lakshika Jayetileke ³ and Zhi-Hua Hu ⁴¹ School of Business Administration and Customs Affairs, Shanghai Customs College, Shanghai 201204, China² School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD 4001, Australia³ Department of Mathematics, Faculty of Science, University of Ruhuna, Matara 81000, Sri Lanka; harshaniejayetileke@gmail.com⁴ Logistics Research Center, Shanghai Maritime University, Shanghai 201306, China; zhhu@shmtu.edu.cn

* Correspondence: liuchanjuan@shcc.edu.cn (C.-J.L.); j73.wu@qut.edu.au (J.W.)

Abstract: The prediction of ship traffic flow is an important fundamental preparation for layout and design of ports as well as management of ship navigation. However, until now, the temporal characteristics and accurate prediction of ship flow sequence in port are rarely studied. Therefore, in this study, we investigated the presence of long-range dependence in container ship flow sequences using the Multifractal Detrended Fluctuation Analysis (MF-DFA). We considered three representative container ports in the world—including Shanghai, Singapore, and Rotterdam container ports—as the study sample, from 1 January 2013 to 31 December 2017. Empirical results suggested that the ship flow sequences are deviated from normal distribution, and the sequences with different time scales exhibited varying degrees of long-range dependence. Furthermore, the ship flow sequences possessed a multifractal nature, where the larger the time scale of ship flow time series, the stronger the multifractal characteristics are. The weekly ship flow sequence in the port of Singapore owned the highest degree of multifractality. Furthermore, the multifractality presented in the ship flow sequences of container ports are due to the correlation properties as well as the probability density function of the ship flow sequences. The study outlines the importance of adopting these features for an accurate modeling and prediction for maritime ship flow series.

Keywords: container ship traffic flow; volatility; generalized Hurst exponents; long-range dependence; multifractality

Citation: Liu, C.-J.; Wu, J.; Jayetileke, H.L.; Hu, Z.-H. Long-Range Dependence and Multifractality of Ship Flow Sequences in Container Ports: A Comparison of Shanghai, Singapore, and Rotterdam. *Appl. Sci.* **2021**, *11*, 10378. <https://doi.org/10.3390/app112110378>

Academic Editor: Giovanni Randazzo

Received: 28 September 2021

Accepted: 29 October 2021

Published: 5 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The analysis of the time series characteristics of port ship flow sequences and the accurate prediction of port ship flow can provide references for the port layout design and the management of ship navigation. Port congestion has been recognized as a serious problem in all large ports in the world, which has a significant effect on the shipping date, the transportation cost, economic loss of the owner of the goods, and even the development of ports [1,2]. Nevertheless, understanding the arrival laws and accurate prediction of the port ship traffic flow are two keys to solve this problem. Therefore, this paper aims to study the arrival laws of ship traffic flow in container ports based on the long-range correlation and multifractality, and then, to provide a reference and theoretical basis for effective modeling and prediction of port ship flows.

There is an abundance of literature on long-range dependence for time series data, such as biomedical data [3–5], stock returns [6–8], hydrology [9–11], and climatology [12–14]. However, only a few studies focus on traffic flow sequences [15,16] and the research on maritime traffic flow sequences are very limited. This is mainly due to the difficulty in obtaining data on port ship traffic flow in the maritime sector. However, in recent years,

there are more and more relevant studies on the application of ship AIS data [17,18]. This makes it possible to study the time series of maritime ship flow.

Studies on maritime traffic flow, especially the ship flow sequences in the ports, are very important as they can provide basic decisions for the allocation of port operation infrastructures, rational port planning, and port investment. Furthermore, if the ship flow sequences in ports are long-range dependent and multifractal, then the traditional ships flow prediction models should revalue to incorporate this type of volatility. Unfortunately, there is no universally accepted theory to define the volatility of traffic flow sequence. In order to understand the irregular patterns of ship flow time series, especially for prediction, we need to know whether the maritime traffic flow system follows chaotic, random, or deterministic structural patterns. The complex pattern is the motivation behind the study of maritime ship flow series through the Multifractal Detrended Fluctuation Analysis (MF-DFA).

The contributions of this study are threefold: First, we present a descriptive statistics of the ship traffic flow time series. Second, we analyze the long-range dependence correlation characteristics of the ship traffic flow series using the Hurst exponent. Third, we determine the degree of multifractality of the ship traffic flow of the different ports through the Multifractal Detrended Fluctuation Analysis, and compare them based on the MF-DFA results. Our results suggested that the ship flow sequences at different time scales in the ports of Shanghai, Singapore, and Rotterdam showed different degrees of long-range dependency. Therefore, the ship flow prediction models should incorporate the long-range dependency in forecasting. In addition, the results indicated that the ship flow sequences in container ports are multifractal, where the degree of multifractality is much higher for the port of Singapore compared with Shanghai and Rotterdam.

The rest of the paper is organized as follows: Section 2 presents a brief review on the literature. Section 3 introduces the methodology used in this paper. Section 4 describes the sample data. Section 5 presents empirical results. Finally, Section 6 provides concluding observations based on the findings of the study.

2. Literature Review

In the last few decades, researchers discovered more characteristics of volatility in terms of long-range dependence and multifractality of the data in numerous fields including DNA sequences [3,4,19], climatology and hydrological time series [11,20,21], and stocks and other financial market data [6,7,22,23]. In most of these studies, the complex long-range correlations and multifractality behaviors of the time series were measured by the so called Hurst exponent, which was originally developed in hydrology for the practical matter of determining optimum dam sizing for the Nile river's volatile rain and drought conditions [22,24].

The long-range correlations can be captured using several methods, including Rescaled Range Analysis (R/S) [22], Detrended Fluctuation Analysis (DFA) [3], Wavelet analysis [25], Multifractal Detrended Fluctuation Analysis [25], and so on. Coronado et al. [26] compared various methods on Hurst exponent and pointed that DFA is superior to other methods since it is less influenced by the time series finite size than others. As a generalization of DFA, MF-DFA is a popular method for the nonstationary time series, which has been applied with great success in several areas of research. Several literatures have also demonstrated the possibility of detecting the multifractal properties in time series through MF-DFA method [9,27].

MF-DFA is a good method to study the characteristics of time series of stock market and complex traffic flow. For example, Mensi et al. [28] and Ali et al. [29] stated that MF-DFA is an acceptable choice to study comparative efficiency and the multifractality of stock markets. They found that the Islamic stock markets' adjustment to speculative activity is, in fact, higher than their conventional counterparts, and all stock market returns exhibited multifractal features. Besides, some scholars used MF-DFA to examine the highway traffic flow time series in Beijing, Shanghai and other places and discovered that the long-range

dependence behavior is ubiquitous in time series of road traffic flows. Moreover, the length of the time scale was significantly impacted on the multifractal characteristics of traffic flow sequences [30–32].

At present, a broad consensus has emerged that long-range dependence and multifractality are somewhat realistic phenomena in traffic flow series [15,33,34]. However, there is no research on the long-range correlations of the time series of maritime traffic flow. Therefore, motivated by the importance of temporal structure and long-range correlations for modeling and prediction of maritime ship flow series, we investigated the complex temporal structure and long-range correlation behaviors of the ship flow sequences of container ports from the multifractal perspective using the MF-DFA method. Evaluation of such results for container ship flow in different time scales and different container ports will facilitate the production of more insights on evolution dynamics of these ports and global trade.

3. Methodology

We analyzed the ship flow sequences of three representative container ports by the MF-DFA method, which is a generalization of DFA method [35,36]. To obtain the generalized Hurst exponent, we followed the five-step procedure introduced by [9,36] to measure the multifractality and nonstationary behavior of Brazilian rivers. Rego et al. [9] pointed out that the periodic components in the sequence should be removed at the first stage before beginning with the general procedure of Kantelhardt et al. [36].

For a record $x(i), i = 1, 2, \dots, N$, where N denotes the length of the record, the MF-DFA consists of the following steps [35,36]:

Step 1: We first integrate the series and obtain the profile $y(j)$,

$$y(j) = \sum_{i=1}^j [x_i - \mu], \tag{1}$$

where μ is the mean value of the entire series.

Step 2: The integrated series $y(j)$ is divided into boxes of equal length s .

Step 3: In each box of length s , we calculate a polynomial fitting of $y(j)$, which represents the trend in that box. The shape of the polynomial trend is defined by the order m . A higher order m yield a more complex shape of the trend, but might lead to overfitting for a time series within small segment sizes. Therefore, in this study, we choose m equal to 2 as suggested by Ihlen (2012) [37]. The y coordinate of the fit line in each box is denoted by $y_s(j)$.

Step 4: The integrated series $y(j)$ is detrended by subtracting the local trend $y_s(j)$ in each box of length s .

Step 5: For a given box size,

$$F_q(s) = \left[\frac{1}{N} \sum_{j=1}^N |y(j) - y_s(j)|^q \right]^{\frac{1}{q}}. \tag{2}$$

If the series is long-range and the power-law is correlated for large values of the time scales, the fluctuation functions $F_q(s)$ can be written as Equation (3):

$$F_q(s) \sim s^{H(q)}, \tag{3}$$

where $H(q)$ is the generalized Hurst exponent. The generalized Hurst exponent $H(q)$ can be obtained by observing the slope of the log–log plots of $F_q(s)$ and scale s through the method of ordinary least squares (OLS). If $H(q)$ is independent of q , the time series is monofractal, otherwise, it is multifractal. Table 1 shows the relationship between the long-range dependence of time series and the generalized Hurst exponent.

Table 1. The relationship between the long-range dependence of time series and the generalized Hurst exponent.

Hurst Exponent	Long-Range Dependence Characteristics of Time Series
$H(q) = 0.5$	There is no long-range dependence in time series and the time series is a completely independent process.
$0.5 < H(q)$	The time series has positive long-range dependence and the overall trend in the future will be consistent with past trend.
$H(q) < 0.5$	The time series has negative long-range dependence and the overall trend in the future will be contrary to the past trend.

The value of the generalized Hurst exponent equal to 0.5 indicates an uncorrelated time series. A generalized Hurst exponent value larger than 0.5 indicates a positive long-range dependence and persistence of the series. In other words, the larger the H value is, the stronger the persistence is. A generalized Hurst exponent value smaller than 0.5 indicates a negative long-range dependence and antipersistence of the series. This means the closer the H value is to 0, the stronger the antipersistence is.

When q is equal to two, $H(2)$ is identical to the well-known Hurst exponent. Generally, the Hurst exponent is between 0 and 1. However, it is worth noting that the generalized Hurst index obtained by applying the MF-DFA method in this study may be greater than 1 [36,38].

The singularity spectrum $f(\alpha)$ is introduced to measure the degree of multifractality of the series and can be obtained through Legendre transform:

$$\alpha = H(q) + qH'(q), \tag{4}$$

$$f(\alpha) = q[\alpha - H(q)] + 1. \tag{5}$$

Here, α is the singularity strength and used to characterize the singularities of the time series. $f(\alpha)$ indicates the dimension of the subset of sequences that is characterized by α . The strength of multifractality can be estimated by the spans of singularity given by

$$\Delta\alpha = \alpha_{\max} - \alpha_{\min}. \tag{6}$$

4. Data

As this research investigates the long-range dependence and multifractality for ship flow sequences in container ports, we extracted required container ship flow data from the Automatic Identification System (AIS) database. The data span is from 1 January 2013 to 31 December 2017 for three representative container ports in the world—that is, Shanghai, Singapore, and Rotterdam. Figure 1 represents the original data of ship flow at different time scales for these three container ports.

The port of Shanghai is not only the largest container port in China, but also the world’s largest container port. Its shipping routes reach the world’s 12 largest shipping areas, and it has established business contacts with more than 500 ports in nearly 200 countries and regions. As the world’s second largest container port, Singapore is also the largest transit port in the Asia-Pacific region. The Port of Rotterdam is the largest port in Europe, as well as the European Gateway. Therefore, in this study, the ports of Shanghai, Singapore, and Rotterdam were chosen to represent all the ports around the globe.

We analyzed and compared the long-range correlation and multifractality characteristics of the ship flow sequences of both Asian and European ports and the gateway ports and transit ports. In general, the ship flow sequences for the three ports depicted different characteristics at different time scales. Among them, the ship flow series of Shanghai and Rotterdam with daily and weekly time scales showed a more significant upward trend than the ship flow series of Singapore port. Ship flow sequences of the three ports with monthly scale fluctuated up and down along the time line, indicating a nonlinear

and nonstationary feature. Generally, these fluctuations are not random, but relate to the seasonal and monthly cycles.

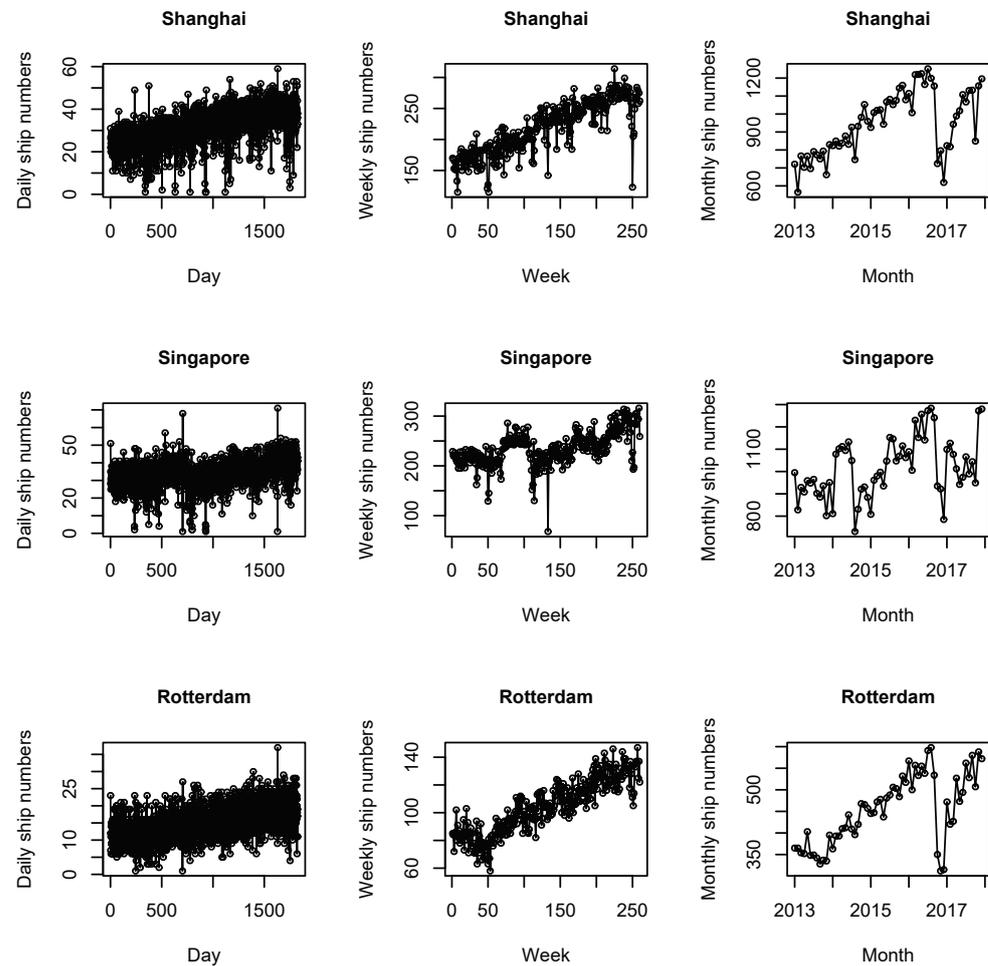


Figure 1. Original data of ship flow of different time scales in three ports.

5. Empirical Results

5.1. Descriptive Statistics

It is a well-known fact that the statistical properties of time series vary with time and depend on time windows. Table 2 presents the descriptive statistics for the original ship flow series at different time scales.

Table 2. Descriptive statistics of the original ship flow series.

Time Scale	Port	Average	Standard Deviation	Skewness	Kurtosis	ADF(<i>p</i>)	KPSS(<i>p</i>)
Daily	Shanghai	31	8.437	−0.420	0.464	−8.152(0.105)	14.418(0.01)
	Singapore	34	7.567	−0.393	2.156	−7.639(0.108)	0.598(0.01)
	Rotterdam	15	4.557	0.263	0.161	−8.989(0.290)	15.130(0.01)
Weekly	Shanghai	218	42.618	−0.177	−0.926	−4.962(0.301)	5.777(0.01)
	Singapore	235	34.244	−0.367	2.004	−3.675(0.226)	0.307(0.01)
	Rotterdam	104	19.488	−0.008	−0.820	−4.085(0.201)	6.009(0.01)
Monthly	Shanghai	950	176.878	−0.110	−1.053	−2.385(0.419)	1.443(0.01)
	Singapore	1019	134.381	0.164	−0.581	−2.684(0.299)	0.102(0.01)
	Rotterdam	453	81.193	−0.003	−1.115	−2.494(0.375)	1.640(0.01)

According to Table 2, the average number of ships arriving in the port of Singapore is greater than that of the Shanghai port at any time scale. Therefore, although Shanghai's container port ranks first in the world list, the port of Singapore is still the busiest container port in the world. This is mainly due to its unique geographical location and its role as the largest transit port in the Asia-Pacific region.

Besides, the standard deviation of ship flow sequences in port of Shanghai is larger than that of Singapore and Rotterdam ports, regardless of the time scale. This shows that the numbers of arriving ships in the ports of Rotterdam and Singapore are relatively stable compared to Shanghai.

The results from skewness and kurtosis analysis demonstrated that the different time scales significantly affect the temporal structure of the ship flow sequence. Indeed, skewness reflects the degree of symmetry in the distribution. Table 2 shows that the skewness in almost all the ship flow sequences are negative, except for the daily ship flow sequence of Rotterdam (0.263) and monthly ship flow sequence of Singapore (0.164). This indicates that all of the ship flow sequences are leftward distributions, except for the daily ship flow sequence of Rotterdam and monthly ship flow sequence of Singapore.

On the other hand, the kurtosis reflects the sharpness of the image. The higher the kurtosis, the sharper the center point on the image; in this sense, the kurtosis measures the degree of data aggregation in the center. In this study, the traditional kurtosis is replaced by the "super kurtosis" calculation method, which is to subtract the kurtosis 3 of the normal distribution from the original kurtosis so that the comparison benchmark is zero. Table 2 shows that the super kurtosis values of daily ship flow sequences are greater than 0, indicating that the daily ship flow sequence distributions are more concentrated and have a longer tail than the normal distribution. In the weekly ship flow sequences, only the super kurtosis of Singapore is greater than 0. The super kurtosis of all the monthly ship flow sequences are less than 0, indicating that these sequences are scattered and have a shorter tail than the normal distribution. Therefore, only the daily ship flow sequence of three ports and the weekly ship flow sequence of Singapore exhibited the characteristic of "sharp peak or fat tail". Further, the ship flow sequences of three ports with different time scales deviated from the normal distribution. The results are further validated by the frequency and probability density distribution of ship flow at the port in Figures 2–4.

In addition, we investigated the stationarity of the above ship flow series with different time scales. In general, the ADF and KPSS tests are complementarily used to evaluate the stationarity in time series models. Therefore, both methods were used in this study to test the stationarity of the ship flow time series with the intention of obtaining more precise results. The null hypothesis of the ADF test was the presence of a unit root, indicating the nonstationarity; the null hypothesis of the KPSS test was the absence of unit root, indicating stability. Table 1 showed that the p values of the ADF test of all the ship flow sequences were greater than 0.05; so, the null hypothesis is accepted—that is, the sequences are nonstationary time series. Similarly, the KPSS test results indicated a p value less than 0.05, rejecting the null hypothesis. Therefore, all the ship flow sequences of three ports are nonstationary time series.

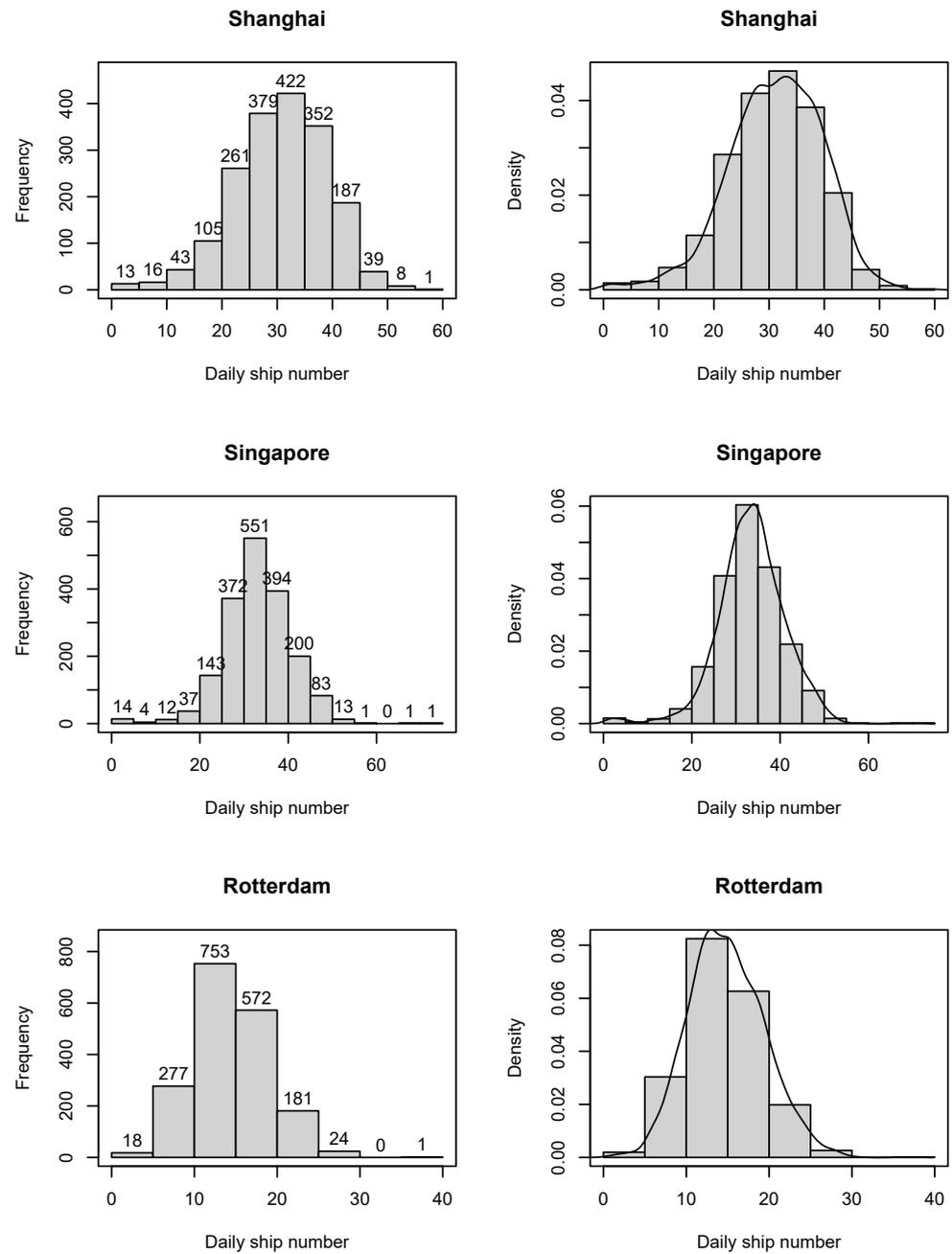


Figure 2. Frequency and probability density distribution of daily ship flow sequences.

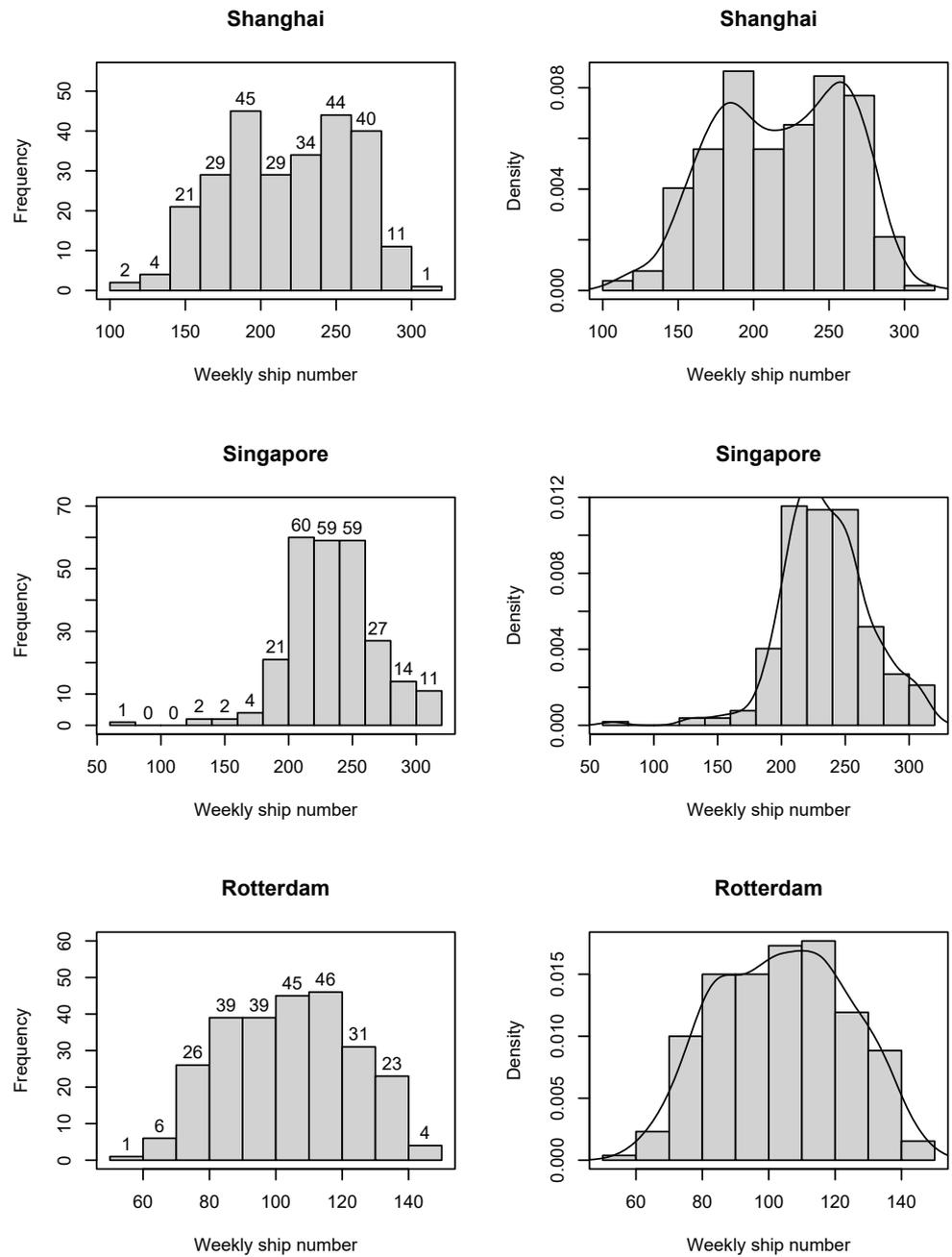


Figure 3. Frequency and probability density distribution of weekly ship flow sequences.

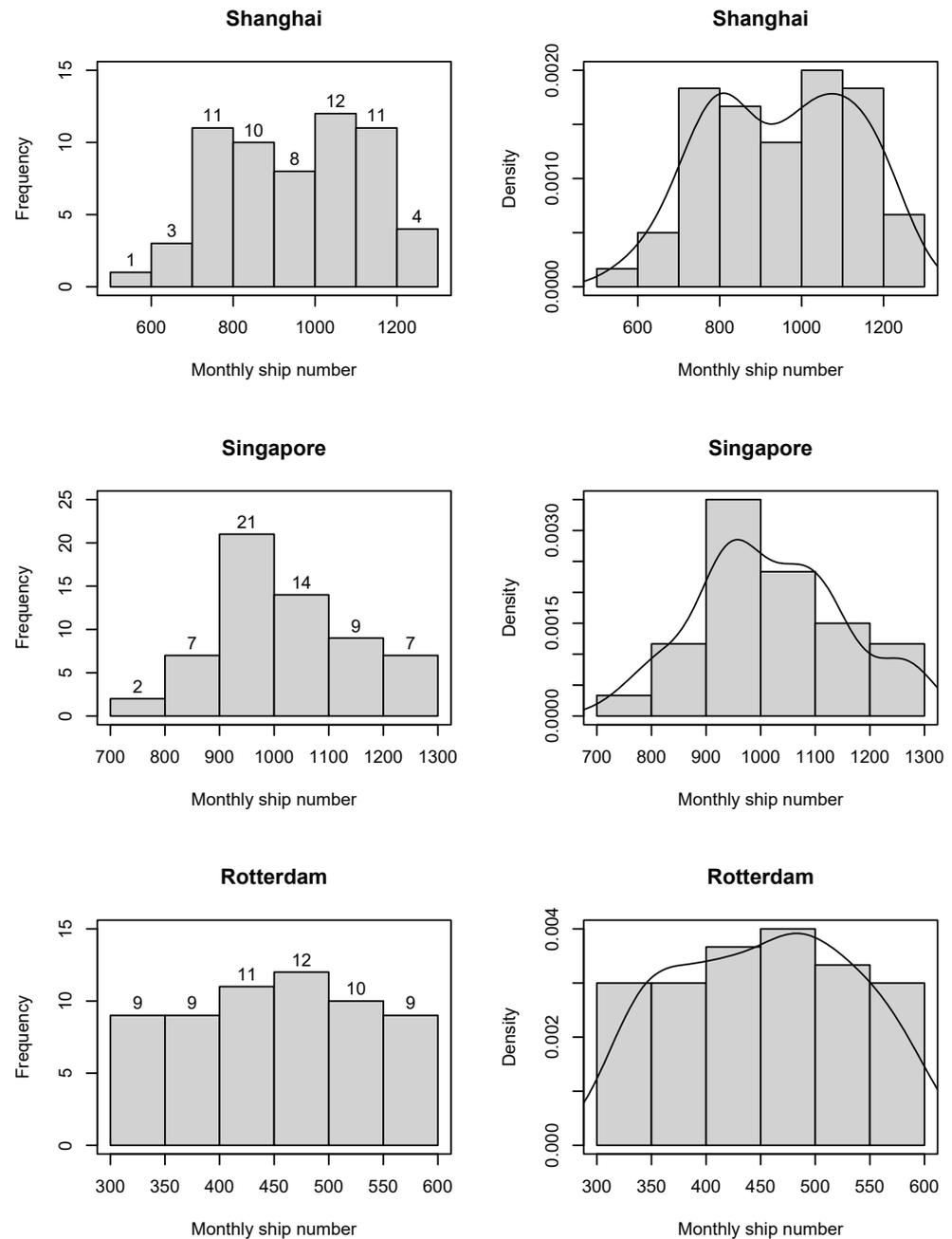


Figure 4. Frequency and probability density distribution of monthly ship flow sequences.

5.2. Long-Range Dependence and Multifractality

Figure 5 depicts the fluctuation function F versus scale for the ship flow sequences of three ports with different time scales in log-coordinates and the OLS linear regression for these curves when q is equal to two. We calculated the generalized Hurst exponents from the slopes of these straight lines. The Hurst exponent of all the ship flow sequences was greater than 0.5, which indicated a positive long-range dependence and persistence in these ship flow sequences.

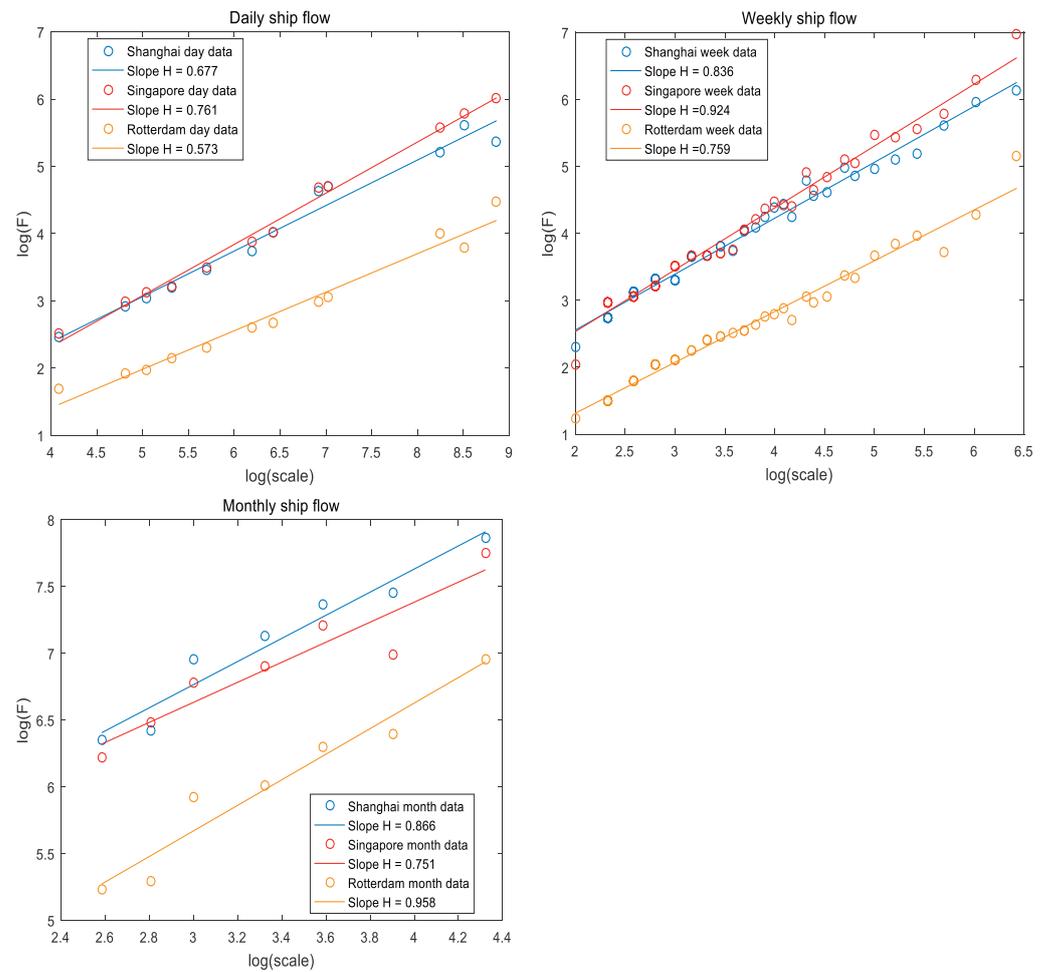


Figure 5. Generalized Hurst exponent ($q = 2$) for three ports with different time scales.

Table 3 presents the results for long-range dependence in ship flow sequences of three ports when q is equal to two. Empirical results suggested high Hurst exponents for ship flow sequences. An important feature of these results implied that, in general, Hurst exponents becomes larger as the time scale increases. The port of Singapore has the highest Hurst exponent in weekly ship flow sequence, which is close to 1. However, in the monthly ship flow sequences, Singapore obtains the lowest Hurst exponent, which is just 0.751. Therefore, ship flow sequences in the port of Singapore have a higher Hurst exponent when the time scale is small. On the contrary, the ship flow sequences in the ports of Shanghai and Rotterdam have a higher Hurst exponent when the time scale is larger. Moreover, the differences in Hurst exponents under different time scales seem to be higher for the port of Singapore.

Table 3. Generalized Hurst exponents ($q = 2$) for ship flow sequences with different time scales.

Ship Flow Sequences	Time Scale		
	Daily	Weekly	Monthly
Shanghai ($m = 2$)	0.677	0.836	0.866
Singapore ($m = 2$)	0.761	0.924	0.751
Rotterdam ($m = 2$)	0.573	0.759	0.958

Figure 6 presents the results for the MF-DFA methodology for $q = [-10 : 10]$. Qualitative results showed that the generalized Hurst exponent for all ship flow sequences decreased with the increase in q ; the generalized Hurst exponent of these ship flow sequences is dependent on the selection of q . However, for the monthly ship flow sequence

in the ports of Shanghai and Rotterdam, this dependence on q was not significant. The differences in generalized Hurst exponents of daily and weekly series for three ports seem to be smaller as q increases. The generalized Hurst exponents for port of Singapore decreased faster than those for Shanghai and Rotterdam with the increase in q in all time scales. In particular, there was no substantial change in generalized Hurst exponent of the monthly ship flow series for Shanghai and Rotterdam; $H(q)$ remained between 0.8 and 1.2 with the increase in q . For the monthly sequence, when q is less than zero, the generalized Hurst exponents of Singapore and Rotterdam are larger than that of Shanghai; meanwhile, when q is larger than zero, the situation may change to the opposite, and the generalized Hurst exponent of ship flow for Shanghai becomes the largest. These results suggest that these differences are not spurious or due to error measures. An important additional comment is that the degrees of multifractality of ship flow sequences in Shanghai and Singapore container ports are much higher than the ones found for ship flow in Rotterdam port.

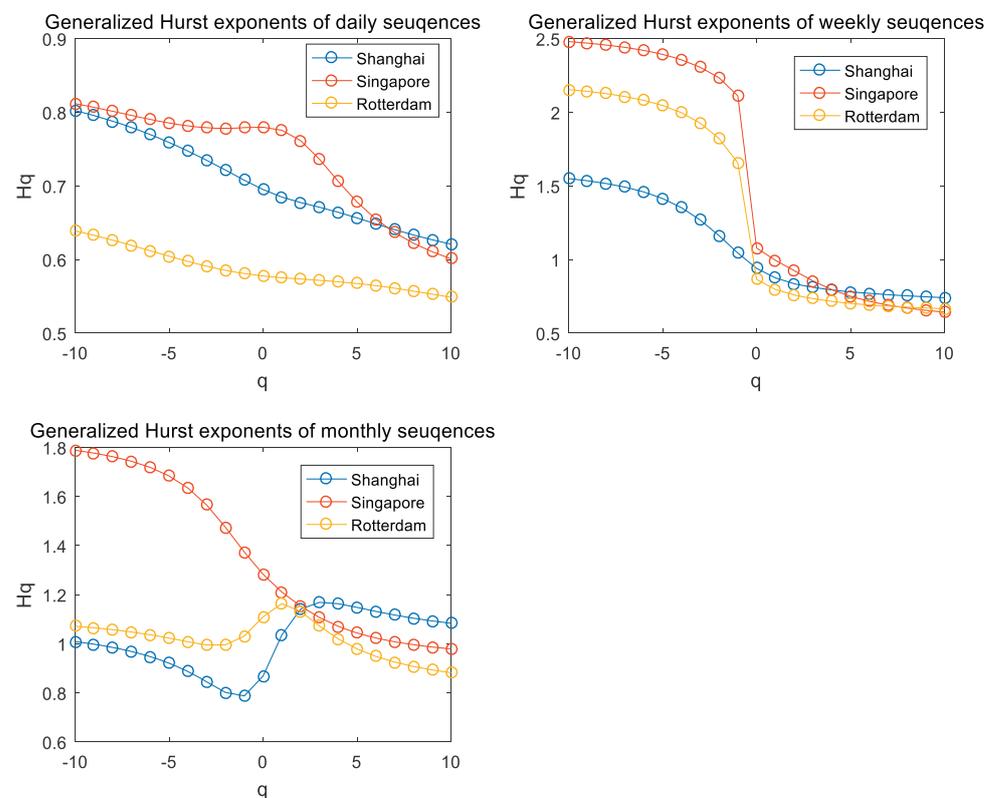


Figure 6. Hurst exponents of ship flow sequences calculated by Multifractal Detrended Fluctuation Analysis (MF-DFA) for $q = [-10 : 10]$.

The leveling of q -order Hurst exponent reflects that the q -order root-mean-square (RMS) is insensitive to the magnitude of local fluctuations. The multifractal spectrum will have a long left tail when the time series have a multifractal structure that is insensitive to the local fluctuations with small magnitudes. In contrary, the multifractal spectrum will have a long right tail when the time series have a multifractal structure that is insensitive to the local fluctuations with small magnitudes [37].

Figure 7 depicts the multifractal spectrum for ship flow sequences of three ports with different time scales. According to Figure 7, the multifractal spectrum of ship flow sequences for all ports can be divided into two sections. However, the spans of the multifractal singularity are different, implying that they have different multifractality strengths. For the daily ship flow sequences, the port of Rotterdam has the lowest multifractal strength, while the port of Singapore has the highest multifractal strength. Furthermore, the shape of multifractal spectrum of daily ship flow for Singapore shifts to the right and the spectrum is

slightly right-skewed, indicating that the scaling behavior of small fluctuations dominates the fluctuation of the daily ship flow for Singapore port.

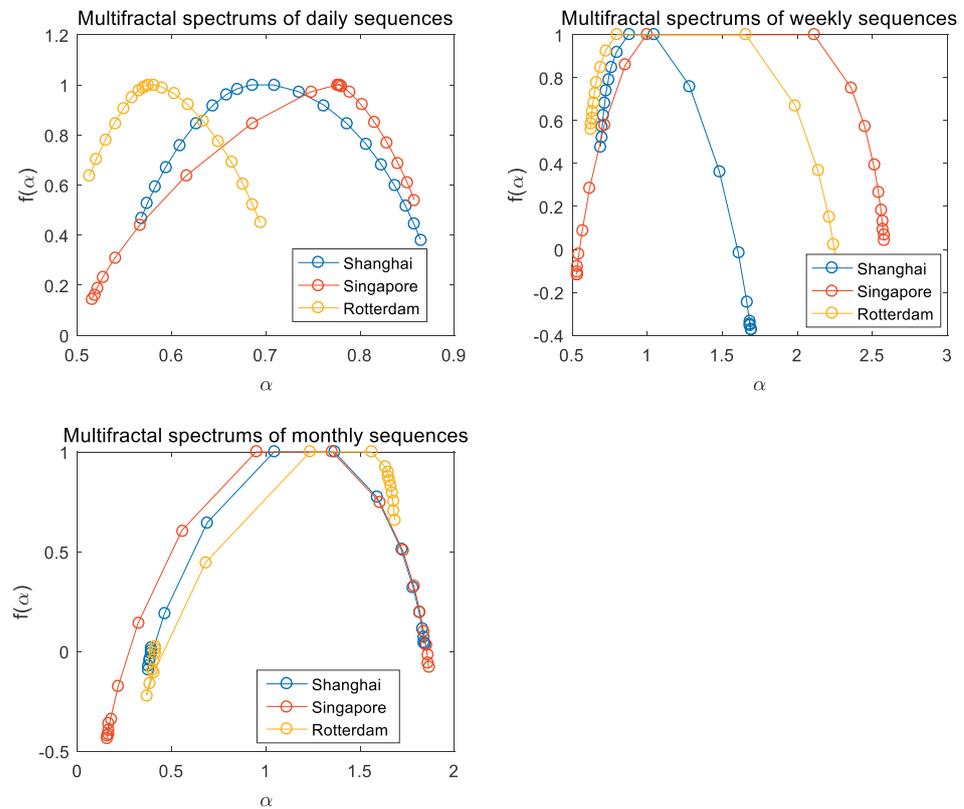


Figure 7. Multifractal spectra for ship flow sequences of three ports with different time scales.

Compared with the daily ship flow time series, the multifractal spectra of Shanghai and Rotterdam obtained from weekly ship flow sequences shifted to the left and the spectrum is slightly left-skewed. This indicates that the scaling behavior of large fluctuations dominates the fluctuation of the weekly ship flow for Shanghai and Rotterdam port.

According to the multifractal theory, the strength of multifractality can be characterized by the span of the multifractal singularity strength function in Equation (6). The bigger the $\Delta\alpha$ is, the stronger the degree of multifractality becomes.

Table 4 presents the quantitative strength of multifractality of all the ship flow sequences of the three ports. It can be seen that the degree of multifractality of weekly series is the strongest for all three ports, followed by the monthly time series, and the daily time series has the weakest multifractality. The monthly ship flow sequence of Singapore port has the highest degree of multifractality. An interpretation for this result is that the weekly ship flow sequence of Singapore is very sensitive to the changes of various influencing factors, and it is very hard to predict. Therefore, compared with the forecast of the ship traffic flows of days and months, the weekly forecast is the most difficult for the ship flow at the container port. This shows that the multifractal method is essential for the analysis of ship flow sequences.

Table 4. The strength of multifractality for ship flow sequences.

Ship Flow Sequences	Time Scale		
	Daily	Weekly	Monthly
Shanghai	0.296	0.999	0.459
Singapore	0.342	2.004	0.984
Rotterdam	0.182	1.643	0.383

5.3. Type of Multifractality

Another contribution of this study is to identify the type of multifractality presented in the ship traffic flow data. We performed the same analysis on the randomly shuffled series of the original ship traffic flow sequences. The randomly shuffled sequences were obtained by shuffling the original ship flow sequences. The shuffled sequences remained with the same fluctuation distributions, though it destroyed any temporal correlations in the original data.

The process of shuffling can be depicted as the three steps presented by [39]. Firstly, (p, q) pairs are generated from random integer numbers with $p, q \leq N$, where N is the length of the original time series. Secondly, p and q entries are swapped with each other. Finally, the above two steps are repeated $N = 20$ times to ensure that the original series is fully shuffled.

We shuffled the ship flow sequences and calculated $H_{\text{shuffle}}(q)$. As seen in Figure 8, the $H_{\text{shuffle}}(q)$ is approximately 0.5 for most of the ship flow sequences except for the weekly ship flow sequences of Singapore. The shuffled ship flow sequences with $H_{\text{shuffle}}(q)$ of about 0.5, indicating the multifractality of these ship flow sequences, are caused by different fluctuations in correlations of small and large scales. However, the multifractality of the weekly ship flow sequences of Singapore port is caused by a broadening of the probability density function. This result is consistent with the results shown in Figure 3. Therefore, the multifractality presented in the ship flow sequences of container ports are due to the correlation properties as well as the probability density function of the ship flow series.

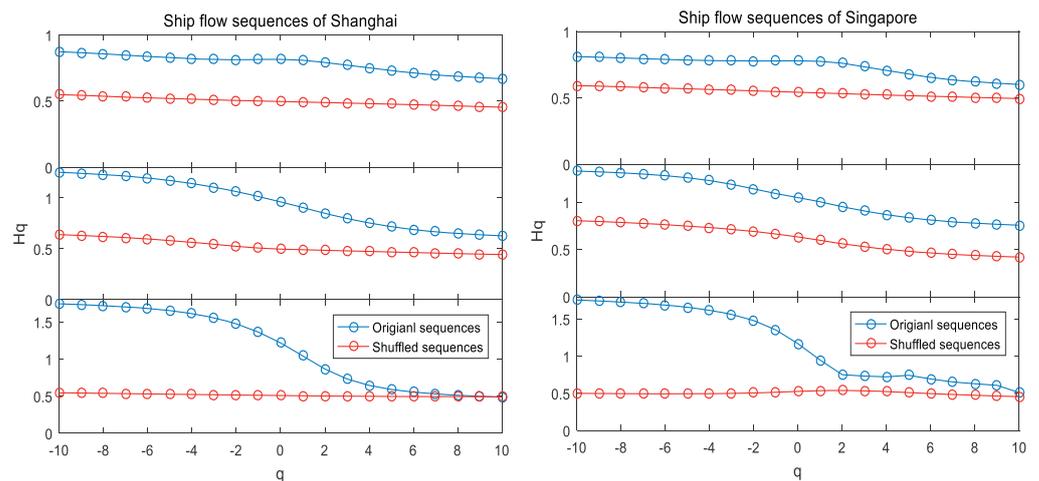


Figure 8. Cont.

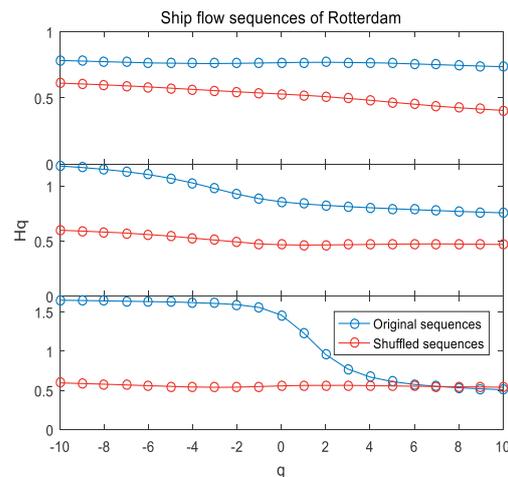


Figure 8. Hurst exponents of the shuffled ship flow sequences for $q = [-10 : 10]$.

6. Conclusions

In this study, we investigated the statistical properties of container ship flow time series and made a detailed investigation on long-range behaviors and fractal characteristics of ship flow sequences for three representative container ports in the world—Shanghai, Singapore, and Rotterdam. This study concludes three main findings.

Firstly, the empirical evidence given in this study emphasize the significance of long-range dependence behaviors and multifractal property in all ship flow sequences at different time scales for three container ports.

Secondly, the empirical evidence from comparisons among these ship flow sequences at different time scales implies that the long-range dependence becomes larger for each port as the time scale increases, except for the port of Singapore. Shanghai and Rotterdam were identified as the ports with the highest degree of long-range dependence in monthly ship flow sequences, while Singapore was identified as having the highest degree of long-range dependence in weekly ship flow sequence.

Finally, the empirical evidence confirmed the multifractal property as an impact factor on the ship flow sequences of container ports. The analysis on the shuffled data indicated that the presence of multifractality in the ship flow sequences of container ports is due to the correlation properties as well as to the probability density function of the ship flow series.

The findings of this study provide some interesting implications. First, the existence of long-range dependence and multifractality in container ship flow could be exploitable and helpful for shipping companies and policy makers. In other words, the presence of chaotic structure such as long-range dependence and multifractality in container ship flow sequences implies that the volume and direction of container ship flow may demand certain rules. Therefore, the shipping company can carry out short-term capacity allocation and adjustment according to container ship flow prediction. Second, the presence of long-range dependence and multifractality in the data suggests that container ship flow forecasting models should account for existing nonlinearities in the data, otherwise, their results may be biased and highly misleading.

Port groups can use these findings in forecasting the expected volatility in the number of arriving container ships, and thereby, in developing and carrying out the layout planning of the port infrastructure, shipping date planning, and even port expansion investment. Moreover, some advanced modeling approaches can be employed for ship flow sequence forecasting, such as statistical modeling [40,41] and machine learning methods [42,43].

In addition, we regarded the trends in the ship traffic flow time series in this study as caused by external conditions. We identified and filtered out these trends in MF-DFA analysis. However, these trends may not be completely caused by external conditions, and some trends may carry endogenous power from data. At this time, when we explore the long-range power-law dependence of the time series, whether to filter out the trend

needs further discussion, as pointed out by Hu et al. [38]. The effect of trends on detrended fluctuation analysis for ship traffic flow time series of ports can be further studied in the future.

Author Contributions: C.-J.L.: conceptualization, methodology, software, validation, formal analysis, investigation, visualization; J.W.: writing—review and editing; H.L.J.: writing—review and editing; Z.-H.H.: resources, data curation, writing—original draft preparation, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), under grant number CE140100049, the National Natural Science Foundation of China (No. 71871136 and 71801150), and the 2021 General project of Shanghai Philosophy and Social Science Planning (2021BJB006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Simões, P.; Marques, R.C. Influence of congestion efficiency on the European seaports performance: Does it matter? *Transp. Rev.* **2010**, *30*, 517–539. [\[CrossRef\]](#)
2. Ke, G.Y.; Li, K.W.; Hipel, K.W. An integrated multiple criteria preference ranking approach to the Canadian west coast port congestion conflict. *Expert Syst. Appl.* **2012**, *39*, 9181–9190. [\[CrossRef\]](#)
3. Peng, C.K.; Buldyrev, S.V.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* **1994**, *49*, 1685. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Audit, B.; Vaillant, C.; Arneodo, A.; d'Aubenton Carafa, Y.; Thermes, C. Long-range correlations between DNA bending sites: Relation to the structure and dynamics of nucleosomes. *J. Mol. Biol.* **2002**, *316*, 903–918. [\[CrossRef\]](#)
5. Leite, A.; Rocha, A.; Silva, M.; Gouveia, S.; Carvalho, J.; Costa, O. Long-range dependence in heart rate variability data: ARFIMA modelling vs. detrended fluctuation analysis. In Proceedings of the 2007 Computers in Cardiology, Durham, NC, USA, 30 September–3 October 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 21–24.
6. Cajueiro, D.O.; Tabak, B.M. Evidence of long range dependence in Asian equity markets: The role of liquidity and market restrictions. *Phys. A Stat. Mech. Appl.* **2004**, *342*, 656–664. [\[CrossRef\]](#)
7. Mishra, R.K.; Sehgal, S.; Bhanumurthy, N. A search for long-range dependence and chaotic structure in Indian stock market. *Rev. Financ. Econ.* **2011**, *20*, 96–104. [\[CrossRef\]](#)
8. Pérez-Rodríguez, J.V.; Andrada-Félix, J.; Rachinger, H. Testing the forward volatility unbiasedness hypothesis in exchange rates under long-range dependence. *N. Am. J. Econ. Finance* **2021**, *57*, 101438. [\[CrossRef\]](#)
9. Rego, C.; Frota, H.; Gusmão, M. Multifractality of Brazilian rivers. *J. Hydrol.* **2013**, *495*, 208–215. [\[CrossRef\]](#)
10. Tyralis, H.; Dimitriadis, P.; Koutsoyiannis, D.; O'Connell, P.E.; Tzouka, K.; Iliopoulou, T. On the long-range dependence properties of annual precipitation using a global network of instrumental measurements. *Adv. Water Resour.* **2018**, *111*, 301–318. [\[CrossRef\]](#)
11. Iliopoulou, T.; Papalexiou, S.M.; Markonis, Y.; Koutsoyiannis, D. Revisiting long-range dependence in annual precipitation. *J. Hydrol.* **2018**, *556*, 891–900. [\[CrossRef\]](#)
12. Pal, S.; Dutta, S.; Nasrin, T.; Chattopadhyay, S. Hurst exponent approach through rescaled range analysis to study the time series of summer monsoon rainfall over northeast India. *Theor. Appl. Climatol.* **2020**, *142*, 581–587. [\[CrossRef\]](#)
13. Karmakar, S.; Goswami, S.; Chattopadhyay, S. Exploring the pre-and summer-monsoon surface air temperature over eastern India using Shannon entropy and temporal Hurst exponents through rescaled range analysis. *Atmos. Res.* **2019**, *217*, 57–62. [\[CrossRef\]](#)
14. Chakraborty, S.; Chattopadhyay, S. Exploring the Indian summer monsoon rainfall through multifractal detrended fluctuation analysis and the principle of entropy maximization. *Earth Sci. Inform.* **2021**, *14*, 1571–1577. [\[CrossRef\]](#)
15. Chand, S.; Aouad, G.; Dixit, V.V. Long-Range Dependence of Traffic Flow and Speed of a Motorway: Dynamics and Correlation with Historical Incidents. *Transp. Res. Rec.* **2017**, *2616*, 49–57. [\[CrossRef\]](#)
16. Wang, J.L.; Wei, H.R.; Ye, C.; Ding, Y. Fractal behavior of traffic volume on urban expressway through adaptive fractal analysis. *Phys. A Stat. Mech. Appl.* **2016**, *443*, 518–525.
17. Hu, Z.H.; Liu, C.J.; Chen, W.; Wang, Y.G.; Wei, C. Maritime convection and fluctuation between Vietnam and China: A data-driven study. *Res. Transp. Bus. Manag.* **2020**, *34*, 100414. [\[CrossRef\]](#)
18. Hu, Z.H.; Liu, C.J.; Tae-Woo Lee, P. Analyzing interactions between Japanese ports and the Maritime Silk Road based on complex networks. *Complexity* **2020**, *2020*, 3769307. [\[CrossRef\]](#)

19. Ghorbani, M.; Jonckheere, E.A.; Bogdan, P. Gene expression is not random: Scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks. *Front. Physiol.* **2018**, *9*, 1446. [[CrossRef](#)] [[PubMed](#)]
20. Gaudio, R.; De Bartolo, S.G.; Primavera, L.; Gabriele, S.; Veltri, M. Lithologic control on the multifractal spectrum of river networks. *J. Hydrol.* **2006**, *327*, 365–375. [[CrossRef](#)]
21. Franzke, C. Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *J. Clim.* **2012**, *25*, 4172–4183. [[CrossRef](#)]
22. Hurst, H.E. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Eng.* **1951**, *116*, 770–799. [[CrossRef](#)]
23. Wang, Z.; Yan, Y.; Chen, X. Long-range correlation and market segmentation in bond market. *Phys. A Stat. Mech. Appl.* **2017**, *482*, 477–485. [[CrossRef](#)]
24. Wang, Y.; Wei, Y.; Wu, C. Analysis of the efficiency and multifractality of gold markets based on multifractal detrended fluctuation analysis. *Phys. A Stat. Mech. Appl.* **2011**, *390*, 817–827. [[CrossRef](#)]
25. Abry, P.; Veitch, D. Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Inf. Theory* **1998**, *44*, 2–15. [[CrossRef](#)]
26. Coronado, A.V.; Carpena, P. Size effects on correlation measures. *J. Biol. Phys.* **2005**, *31*, 121–133. [[CrossRef](#)] [[PubMed](#)]
27. Wang, F.; Liao, G.P.; Zhou, X.Y.; Shi, W. Multifractal detrended cross-correlation analysis for power markets. *Nonlinear Dyn.* **2013**, *72*, 353–363. [[CrossRef](#)]
28. Mensi, W.; Hamdi, A.; Yoon, S.M. Modelling multifractality and efficiency of GCC stock markets using the MF-DFA approach: A comparative analysis of global, regional and Islamic markets. *Phys. A Stat. Mech. Appl.* **2018**, *503*, 1107–1116. [[CrossRef](#)]
29. Ali, S.; Shahzad, S.J.H.; Raza, N.; Al-Yahyaee, K.H. Stock market efficiency: A comparative analysis of Islamic and conventional stock markets. *Phys. A Stat. Mech. Appl.* **2018**, *503*, 139–153. [[CrossRef](#)]
30. Shang, P.; Lu, Y.; Kamae, S. Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis. *Chaos Solitons Fractals* **2008**, *36*, 82–90. [[CrossRef](#)]
31. Doménech-Carbó, A. Rise and fall of historic tram networks: Logistic approximation and discontinuous events. *Phys. A Stat. Mech. Appl.* **2019**, *522*, 315–323. [[CrossRef](#)]
32. Feng, S.; Wang, X.; Sun, H.; Zhang, Y.; Li, L. A better understanding of long-range temporal dependence of traffic flow time series. *Phys. A Stat. Mech. Appl.* **2018**, *492*, 639–650. [[CrossRef](#)]
33. Wang, Y.; Vormer, F.; Hu, M.; Duong, V. Empirical analysis of air traffic controller dynamics. *Transp. Res. Part C Emerg. Technol.* **2013**, *33*, 203–213. [[CrossRef](#)]
34. Yuan, P.; Lin, X. How long will the traffic flow time series keep efficacious to forecast the future? *Phys. A Stat. Mech. Appl.* **2017**, *467*, 419–431. [[CrossRef](#)]
35. Gómez-Extremera, M.; Carpena, P.; Ivanov, P.C.; Bernaola-Galván, P.A. Magnitude and sign of long-range correlated time series: Decomposition and surrogate signal generation. *Phys. Rev. E* **2016**, *93*, 042201. [[CrossRef](#)]
36. Kantelhardt, J.W.; Zschiegner, S.A.; Koscielny-Bunde, E.; Havlin, S.; Bunde, A.; Stanley, H.E. Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A Stat. Mech. Appl.* **2002**, *316*, 87–114. [[CrossRef](#)]
37. Ihlen, E.A.F. Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. *Front. Physiol.* **2012**, *3*, 141. [[CrossRef](#)]
38. Hu, K.; Ivanov, P.C.; Chen, Z.; Carpena, P.; Stanley, H.E. Effect of trends on detrended fluctuation analysis. *Phys. Rev. E* **2001**, *64*, 011114. [[CrossRef](#)]
39. Lu, W.Z.; Xue, Y. Detrended fluctuation analysis of particle number concentrations on roadsides in Hong Kong. *Build. Environ.* **2014**, *82*, 580–587. [[CrossRef](#)]
40. Xu, X.; McGrory, C.A.; Wang, Y.G.; Wu, J. Influential factors on Chinese airlines' profitability and forecasting methods. *J. Air Transp. Manag.* **2021**, *91*, 101969. [[CrossRef](#)]
41. Zhang, S.; Wu, J.; Jia, Y.; Wang, Y.G.; Zhang, Y.; Duan, Q. A temporal LASSO regression model for the emergency forecasting of the suspended sediment concentrations in coastal oceans: Accuracy and interpretability. *Eng. Appl. Artif. Intell.* **2021**, *100*, 104206. [[CrossRef](#)]
42. Zhu, M.; Wu, J.; Wang, Y.G. Multi-horizon accommodation demand forecasting: A New Zealand case study. *Int. J. Tour. Res.* **2021**, *23*, 442–453. [[CrossRef](#)]
43. Wu, J.; Wang, Y.G.; Tian, Y.C.; Burrage, K.; Cao, T. Support vector regression with asymmetric loss for optimal electric load forecasting. *Energy* **2021**, *223*, 119969. [[CrossRef](#)]

Article

Forecasting Taxi Demands Using Generative Adversarial Networks with Multi-Source Data

Hasan A. H. Naji, Qingji Xue *, Huijun Zhu and Tianfeng Li

School of Digital Media, Nanyang Institute of Technology, Chang Jiang Road No. 80, Nanyang 473004, China; hasanye1985@gmail.com (H.A.H.N.); zhuhj1201@163.com (H.Z.); tianfengli@163.com (T.L.)

* Correspondence: xue_qj@sina.com

Abstract: As a popular transportation mode in urban regions, taxis play an essential role in providing comfortable and convenient services for travelers. For the sake of tackling the imbalance between supply and demand, taxi demand forecasting can help drivers plan their routes and reduce waiting time and oil pollution. This paper proposes a deep learning-based model for taxi demand forecasting with multi-source data using Generative Adversarial Networks. Firstly, main features were extracted from multi-source data, including GPS taxi data, road network data, weather data, and points of interest. Secondly, Generative Adversarial Network, comprised of the recurrent network model and the conventional network model, is adopted for fine-grained taxi demand forecasting. A comprehensive experiment is conducted based on a real-world dataset of the city of Wuhan, China. The experimental results showed that our model outperforms state-of-the-art prediction methods and validates the usefulness of our model. This paper provides insights into the temporal, spatial, and external factors in taxi demand-supply equilibrium based on the results. The findings can help policymakers alter the taxi supply and the taxi lease rents for periods and increase taxi profit.

Keywords: taxi demand; forecasting; multi-source data; generative adversarial networks

Citation: Naji, H.A.H.; Xue, Q.; Zhu, H.; Li, T. Forecasting Taxi Demands Using Generative Adversarial Networks with Multi-Source Data. *Appl. Sci.* **2021**, *11*, 9675. <https://doi.org/10.3390/app11209675>

Academic Editors:

Giovanni Randazzo,
Anselme Muzirafuti and
Dimitrios S. Paraforos

Received: 23 August 2021
Accepted: 6 October 2021
Published: 17 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Taxis play a vital role in the modern urban transportation system as they comfortably and conveniently serve many urban passengers. According to the annual report of urban passenger transport operation [1], more than 216 million passengers used a taxi service in 2020 in Wuhan city. However, a critical challenge emerged that there is a significant mismatch between the supply of taxis and passengers' demands. For instance, passengers may not find taxis for an extended period in an area at a specific time. In contrast, taxi drivers may cruise roads without getting passengers in another area at the same time. Therefore, this may lead to several problems, such as increasing passengers' wait time and oil consumption and decreasing taxi incomes. To this end, it becomes significant to accurately predict fine-grained taxi demands in advance to guide taxi drivers to areas with high demands [2].

There is a need for a deep understanding of the temporally varying taxi-passenger demand over various spatial areas that can guide and motivate drivers to be in the areas with a high potential of passenger demands and thus enhance the taxis' utilization rate [3].

Recently, there has been much research investigating the correlation between taxi demands and related dependencies. However, taxi demand forecasting is still an open problem, which is mainly affected by several kinds of factors [4–6]:

- (1) Temporal factors include weekdays, weekends, rush hours, holidays.
- (2) Spatial factors such as neighboring areas.
- (3) External factors including POI data, weather, traffic condition, etc.

In this study, we try to study taxi demands and related factors extracted from multi-source data. The taxi demands, represented by taxi pick-up events, were analyzed using

a generative adversarial networks (GAN) to forecast taxi-passenger demand by considering significant factors, including temporal, spatial, and external factors. A novel deep learning-based approach is proposed for feature extraction and fine-grained taxi demands forecasting in urban areas.

The main contributions of this article include the following:

- (1) To provide a method for extracting features related to taxi demands from multi-source data following a detailed analytic for the relationship between taxi demands and significant factors.
- (2) The generative adversarial networks (GAN) model is adopted to perform taxi demand forecasting. In the GAN, the long short-term memory (LSTM) structure is selected in the generative network, whereas the Conventional network is used in the discriminator network to enable the GAN model to deal with fundamental factors and related patterns naturally.
- (3) The proposed forecasting model was evaluated using a real multi-source data set collected of Wuhan city, China. The experiments proved that the proposed model showed better results and outperformed other state-of-the-art methods.

The rest of the paper is structured as follows: Section 2 presents a related review of the recent taxi demand prediction methods. Section 3 introduced the mathematical definition of taxi demand forecasting. Section 4 introduces the proposed methodology, followed by explaining the conducted experiment in detail in Section 5. The experimental results are discussed in Section 6. Section 7 concludes the implications and value of taxi demand forecasting findings and introduces the future work.

2. Related Work

Taxi demand forecasting has attracted researchers' and taxi service companies' attention due to the massive number of GPS trajectories and the huge spatio-temporal information produced every day by GPS sensors. In general, taxi demands forecasting methods can be categorized into three main classes: traditional methods, machine, and deep learning-based methods.

2.1. Traditional Methods

Traditional methods include statistical and time-series analysis-based methods. Statistical models have been used to study the predicting taxi demands. For instance, Tang et al. [7] developed a probabilistic-based model to predict vehicle trip routes using Hidden Markov Model (HMM). Moreira et al. proposed a data-driven method to predict passengers' spatial distribution in short-term periods [8]. Liu et al. developed three predictive methods for detecting high hotspots and predicting taxi demands [9]. Chang et al. [10] investigated historical trajectory data to forecast the spatio-temporal patterns of taxi demands.

Time-series analysis-based methods are also considered in traffic data prediction. For example, in [11], the Automatic ARIMA Model is adopted for forecasting the passenger's hotspot regions using spatio-temporal data. In [12], the authors modeled taxi demand prediction as a time series issue, and an improved ARIMA method is proposed to predict taxi demands by using temporal dependencies. Tong et al. [13] introduced a linear regression-based model along with high-dimensional features to forecasting taxi demands in urban regions.

2.2. Machine and Deep Learning-Based Methods

Markou et al. [14] utilized the information extracted by unstructured data in taxi GPS data and adopted machine learning techniques to forecast taxi demands. In [15], a backpropagation neural network (BPNN) with an extreme gradient boosting (XGB) based method is proposed to predict taxi-hailing demand. In [16], the authors introduced a machine learning-based approach for identifying and predicting the short-term demand

for on-demand ride-hailing services. The predicting methodology studied factors related to traffic, trip fare, and weather conditions.

Recently, deep learning-based methods have been popularly adopted in forecasting traffic flow problems. Multi-layer perceptron (MLP), Convolution Neural Network (CNN), Recurrent Neural Network (RNN), and their variation networks have achieved superior achievements in taxi demand prediction.

CNN has been used to forecast traffic flow. For example, Ma et al. [17] split a city into several small grids, transformed city traffic speed into images, and adopted CNN for forecasting traffic speed. Zhang et al. [18] applied CNN by modeling temporal and spatial factors for predicting the traffic flow of bikes in the short-future periods.

The success of RNNs and their enhanced models, such as long short-term memory (LSTM), and gated recurrent units (GRU), led researchers to adopt these methods for predicting traffic flow [5,19,20]. Xu et al. utilized a sequence learning method for predicting future taxi demands in city regions, considering current taxi requests and relevant factors. Mixture density-based recurrent neural networks are developed to investigate historical taxi demand distribution and taxi demand predictions. Rossi in [19] obtained the sequences of historical pick-up and drop-offs and then employed the LSTM network to extract the sequential features for taxi request forecasting. Zhao et al. [20] used a cascaded-based LSTM combined with an origin–destination correlation matrix to capture spatial-temporal patterns. In [10], an LSTM neural network-based method is adopted for forecasting a future pick-demand of a given taxi stand by analyzing the spatial demand of a particular taxi stand and neighboring stands.

To make full use of the spatio-temporal correlation with taxi demands, many researchers combined both CNN and RNN for forecasting traffic flow.

A convolutional-recurrent network-based model is developed for forecasting fine-grained taxi requests [21]. The authors considered various factors related to taxi demands, including the spatial correlations between neighboring areas and function-similar areas, long and short-term periods, and external factors. A context-based attention method combines regions' predictions to improve the prediction results [22]. Niu et al. [6] used LSTM with CNN in a real-time prediction system by streaming taxi-passenger data. The CNN network is adopted to extract spatial features, whereas LSTM obtained temporal dimensions. Zhang et al. [23] proposed a deep-learning-based model to predict the in-flow and the out-flow of crowds in a city.

Previous works on taxi demand prediction led us to consider historical trip information to forecast future taxi requests. One of the differences between our work and the preceding works is that our model can fill this gap by integrating taxi GPS data with other related historical data sources, including weather conditions, temporal data, POI data, to fully understand taxi supply patterns and consider the significant factors affecting external dependencies of taxi demand simultaneously.

In this research, we employed the generative adversarial networks (GAN) model to forecast taxi demands. In the generative network part, the long short-term memory (LSTM) structure is selected, whereas the Conventional network is used in the discriminator network. Thus, the GAN model (GAN_LSTM_CNN) can deal with the factors included in the Taxi dataset and related patterns perfectly. In addition, and by considering successive iteration and adversarial learning of the discriminator and the generator, we can obtain better prediction results.

3. Problem Definition

In this section, firstly we define several key concepts, and then formally formulate the taxi demand forecasting problem.

Definition 1 (taxi trip). A taxi trip is defined as a tuple (ID, pickup_{time}, pick_{location}, drop_{time}, drop_{location}, duration, distance, fare), where ID represents the trip identification, pickup_{time} is the pick-up time (in hours and minutes), pick_{location} is the pick-up location (longitude and latitude), drop_{time} is the drop-off time, drop_{location} represents the

drop-off location, duration, distance, and fare are the calculated duration, distance and, fare of a trip.

Definition 2 (Road Network). A road network of a city is composed of a set of road segments. Each road segment is associated with two terminal points (i.e., intersections of crossroads), and connects with other road segments by sharing the same terminals. All road segments compose the road network in the format of a graph.

Definition 3 (Temporal Data). For the sake of a detailed description of the spatial and temporal dimensions of fine-grained taxi demands, the time is discretized into time slots t (date, hour, and minutes).

Definition 4 (Point of Interest POI). POI represented a place (like a restaurant) in an urban area r . Each POI p_i associated with a location $p_{i,l}$ and a POI category $p_{i,c} \in C$, where C is a set of categories.

Definition 5 (Weather). Weather presents several parameters related to weather status in a determined time slot (date, hour, and minutes) of an urban area r . Each weather may consider several associated parameters including temperature, humidity, weather status (rain, sunny), wind, etc.

Here, we can provide a formal definition of fine-grained taxi demands as follows.

Definition 6 (Taxi Demands). We use Y_i to represent the amount of taxi pick-ups (demand of taxis) in an area $r \in R$ at a time slot $t \in T$. Hence, taxi demands in a time slot t are defined as $Y_{r,t} = [Y_{0,t}, Y_{1,t}, \dots, Y_{n,t}]$.

To forecast taxi demands in future, we first dig fine-grained taxi demands in past time slots from historical taxi trips data defined as in Definition 1. As Taxi Demands have a strong correlation with other temporal, spatial and external factors, we collect significant factors related to taxi demands (as in Definitions 2,3,4 and 5) in the determined area r , at time slot t .

Now, we can define the problem of forecasting the fine-grained taxi demands in a determined time slot t as follows.

Definition 7 (Taxi Demand Forecasting Problem). Consider an urban area is divided into disjoint regions R by the road network.

Given fine-grained taxi demands in the past time slots $\{Y_t | t = 0, 1, \dots, t\}$ extracted from historical taxi trips data (pick-ups), we try to predict fine-grained taxi demands at a determined time slot. The taxi demand forecasting is denoted as $\hat{Y}_{t+1} = [\hat{Y}_{1,t+1}, \dots, \hat{Y}_{n,t+1}]$.

4. Methodology

This section introduces the proposed taxi demand forecasting model in detail, including raw data, features extraction and analysis, features pre-processing, and the GAN model. Figure 1 depicts the framework of the proposed model.

4.1. Raw Data

4.1.1. Taxi GPS Data

The dataset used in this paper contains temporally ordered location records collected from 9124 GPS-enabled taxis during 91 days (historical GPS dataset of September, October, and November in 2013) of the city of Wuhan, China.

The temporal resolution of the dataset is 25 s; therefore, approximately 3500 GPS records for a taxi per a day (24 h) would be collected, and the total volume of the dataset approached 320 million records. The GPS meter triggered information to the taxi data center and contains seven taxi attributes, i.e., taxi (driver) ID, timestamp, longitude, latitude, speed, direction, and taxi status. The detailed description of the taxi attributes is shown in Table 1.

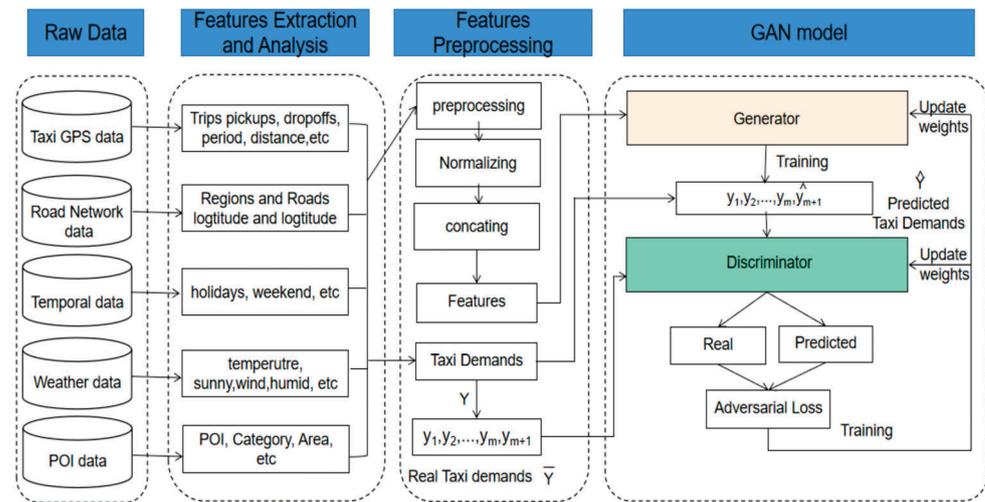


Figure 1. Framework of the proposed model.

Table 1. Description of the attributes of a record in the Taxi GPS dataset.

Attribute	Description	Example
Taxi (Driver) ID	8~10 digit number as taxi identifier.	28784651
Time	14-digit number, format yyyyMMddHHmmss	20131112181211
Longitude	2~3 digit number with 6 decimals, in the degree unit.	33.2260002
Latitude	2~3 digit number with 6 decimals, in the degree unit.	114.232253
Speed	Taxi speed in miles/hour unit.	39.57
Taxi Status	0: Taxi device is invalid, 1: Taxi not occupied; 2: temporarily stopped, 3: occupied.	0,1,2,3
Direction	2~3 digit number with 2 decimals, in the degree unit.	141.71

4.1.2. Road Network Data

A dataset of the city of Wuhan’s road network obtained from the OpenStreetMap website [24] is used in this research, containing 94,214 intersections and 95,781 road segments. Each road intersection is vectorial, located by related latitude and longitude. The whole area of the city is divided into eight significant regions and 623 disjoint small regions. Figure 2 shows the road network and the central areas in the city of Wuhan, China.

4.1.3. Temporal Data

Besides the taxi GPS data, our study considered temporal data and related factors, including weekday, weekend, holiday, etc., significantly influencing traffic flow and taxi demand distribution.

4.1.4. POI Data

The information of points of interests (POI) may provide useful details about the functions of urban regions relevant to taxi demands. We obtained POI dataset of the city of Wuhan by web data crawler tool (using selenium and PhantomJS in python) from guihuayun website [25] which resulted in collecting approximately 425476 POIs of Wuhan city, grouped into 16 different categories as shown in Table 2.

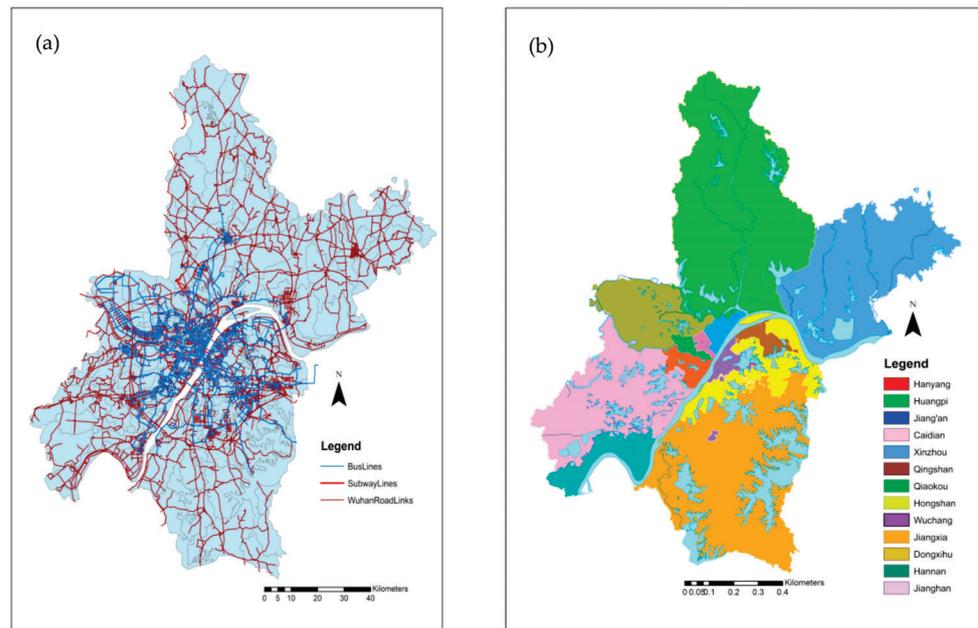


Figure 2. (a) Road network, (b) Main regions of the city of Wuhan.

Table 2. Description of POIs categories in the city of Wuhan.

Category	Total Number	Description
C1	20034	Transportation locations and related services
C2	8358	Accommodation and related services such as hotels.
C3	9902	Sports
C4	3402	Public facilities and public toilets
C5	44160	Business and Companies
C6	10846	Health and beauty
C7	12013	Government and Community
C8	17450	Business residence; residential district
C9	19432	Education
C10	63608	Life activities and Entertainment
C11	121239	Shopping
C12	1357	Scenic Attraction and Travel
C13	62922	Food and Dining

4.1.5. Weather Data

Besides the above external factors, weather conditions have a notable influence on taxi demands [2]. A weather dataset from 1 September 2013 to 30 November 2013 is also obtained by web data crawler tool from the timeanddate website [26]. For ease and simplification, weather categories are converted into categories defined as shown in Figure 3.

Moreover, the weather data includes other variables, including temperature, wind speed, humidity information, etc. To enhance weather data accuracy, we adopted updated data for every hour unless there is a change in the existing conditions.

4.2. Features Extraction and Analysis

To understand the detailed information behind taxi demands and significant correlated factors which play a vital role in taxi demands forecasting, we need to extract and analyze features from the raw multi-source data introduced in Section 4.1.

As shown in Figure 4, small green circles form an occupied trip with a giant green circle (represents a taxi pick up) and end up with a big red circle (represents a taxi drop off). In contrast, small red circles together are considered a vacant trip. In this study, taxi demand prediction focuses on occupied trips. Table 3 shows variables related to occupied trips.

Table 3. Attributes of an occupied trip.

Attribute	Description
Occupied trip ID	represents Taxi ID and trip number such as 28784651_10
Pick up time-stamp	Format 20131112181211 represents year, month, date and time.
Pick up coordinates	includes longitude, latitude of a pick up event.
Drop off time-stamp	Format 20131112181211 represents year, month, date and time.
Drop off coordinates	includes longitude, latitude of a drop off event.
Occupied trip distance	Distance by kilometers
Occupied trip period	Period by minutes
GPS Record number	The number of GPS records in an occupied trip.

For the sake of more accurate results, the trips with a distance less than 200 m or trip period less than 3 min or number of GPS records less than five have been removed and are not considered in this study. Finally, the dataset considered 29,984,635 trips, and this number is sufficient for performing taxi demand prediction. Figure 5 shows three selected study areas in the city of Wuhan, (a) Jiedokou, (b) GuangGu area, and (c) Wuhan railway station area.

4.2.2. Taxi Demands and Temporal Factors

- Taxi demands distribution in Weekdays and Weekends

Figure 6 shows taxi demands distribution in Weekdays and Weekends of three study areas in the city of Wuhan.

We can see that these areas have similar taxi demand distributions (especially on Fridays, indicating that people take taxis on the last weekday for reasons such as going to travel or for entertainment or coming back on Sundays) and similar changes in taxi demands. Therefore, we can leverage the weekdays and weekends data to help in forecasting taxi demands.

- Taxi demands distribution in Holidays

The collected dataset contains eight holidays (National days 1–7 October and 24 November Christmas day). Here we study the taxi demand distribution in the three study areas on the holidays, as shown in Figure 7.

In general, there is a noticeable pattern during holidays in the three areas. Days 1 and 2 October showed high taxi demands followed by a decrease in the other days and finally reached another high demand on 7 October. It can be explained that in the beginning holiday period, and 24 November Christmas day, people may take taxis to travel. In contrast, taxi demands increase again on 7 October when people may come back home.

- Taxi demands distribution in daily hours

Figure 8 reports taxi demand distribution during a day (24 h) in the areas in Wuhan city.

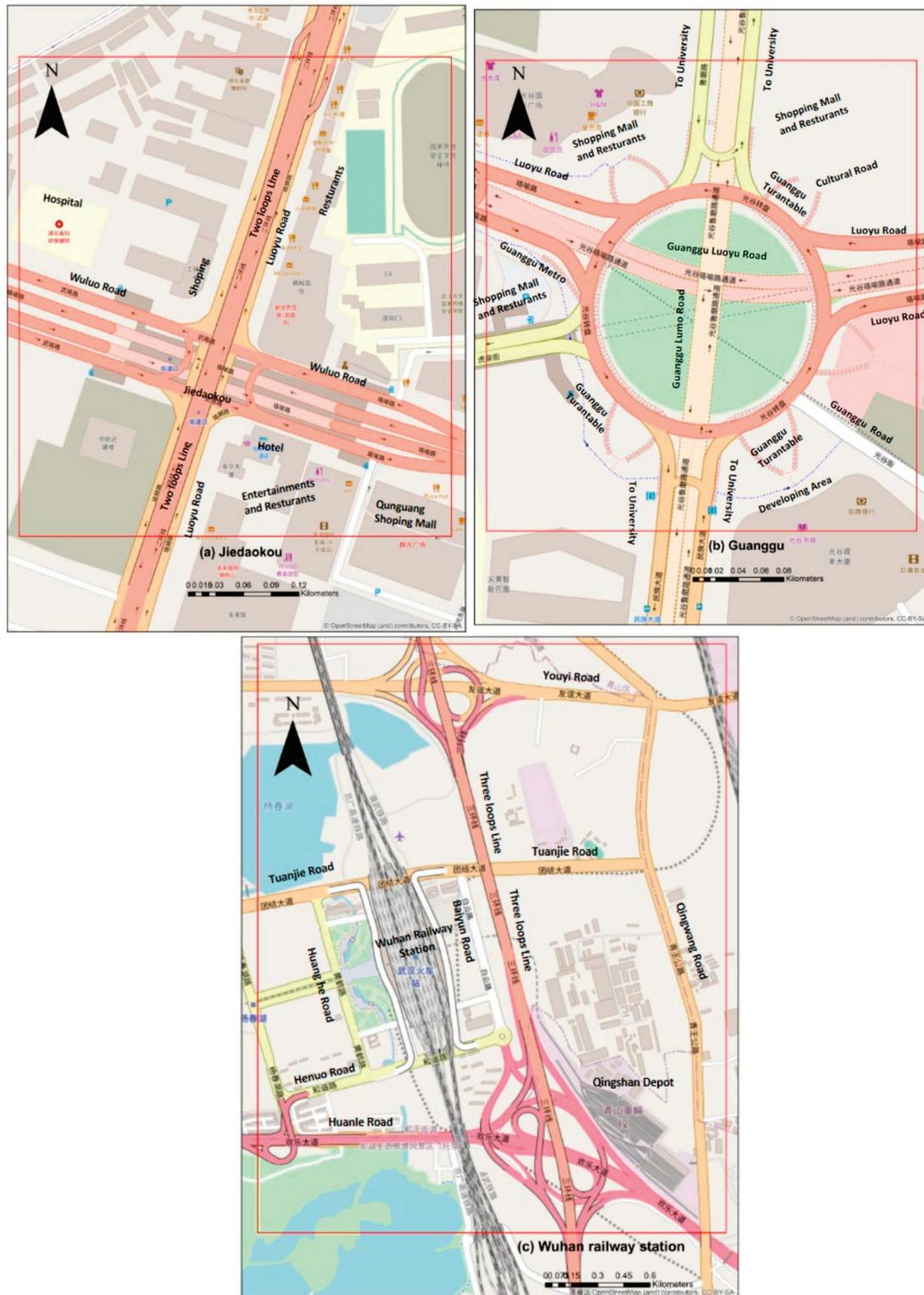


Figure 5. Three study areas in the city of Wuhan, (a) Jiedaokou area, (b) Guanggu area, (c) Wuhan railway station area.

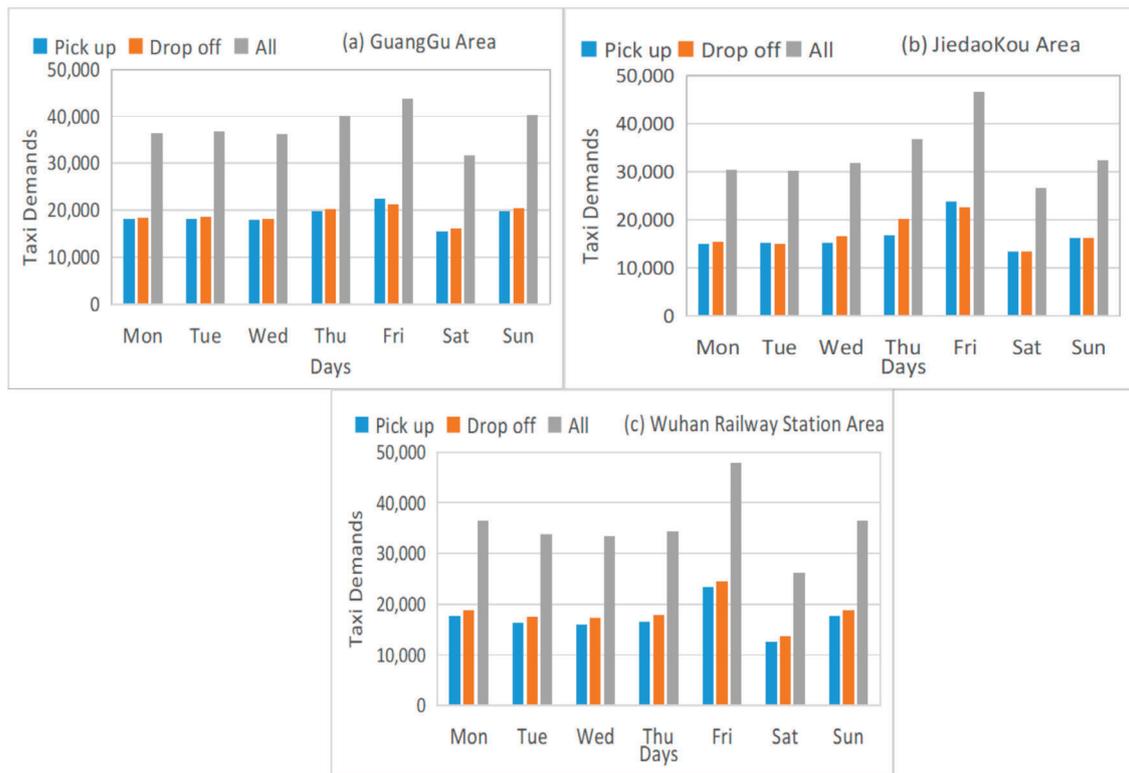


Figure 6. Taxi demands distribution in Weekdays and Weekends in the city of Wuhan. Weekday show high taxi demands as they related to working time, Sunday has a high number of Taxi demands as well.

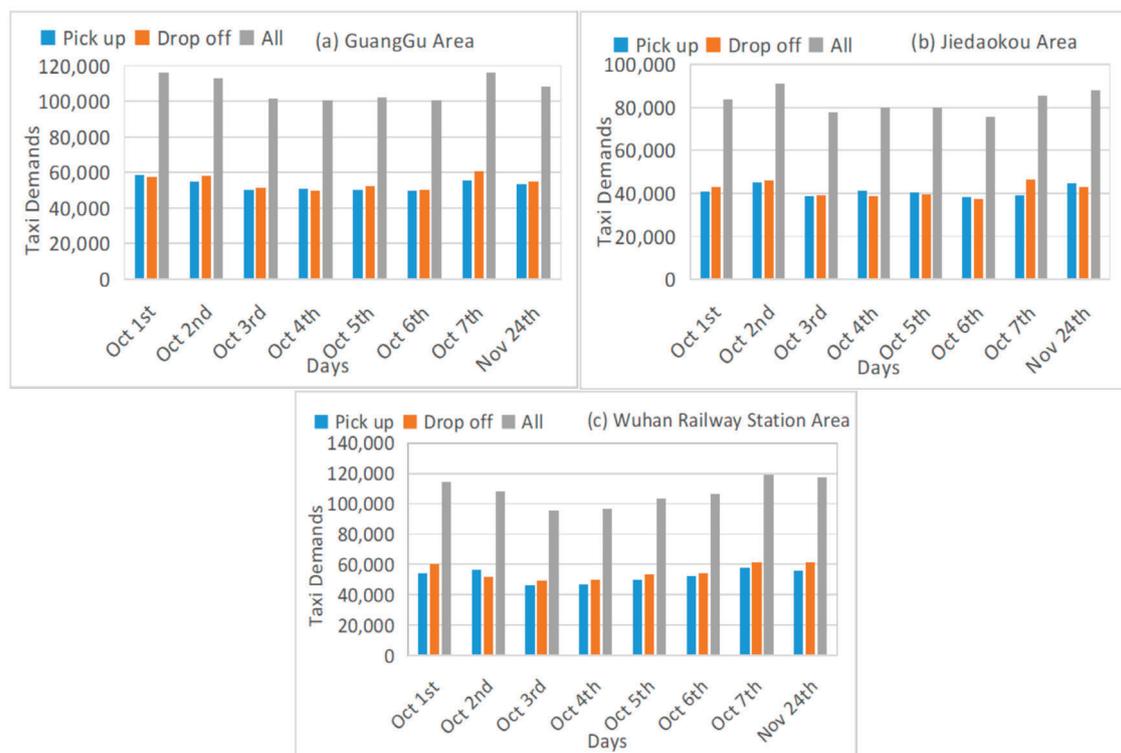


Figure 7. Taxi demands distribution in the city of Wuhan during holidays. Holidays show high taxi demands in the three areas.

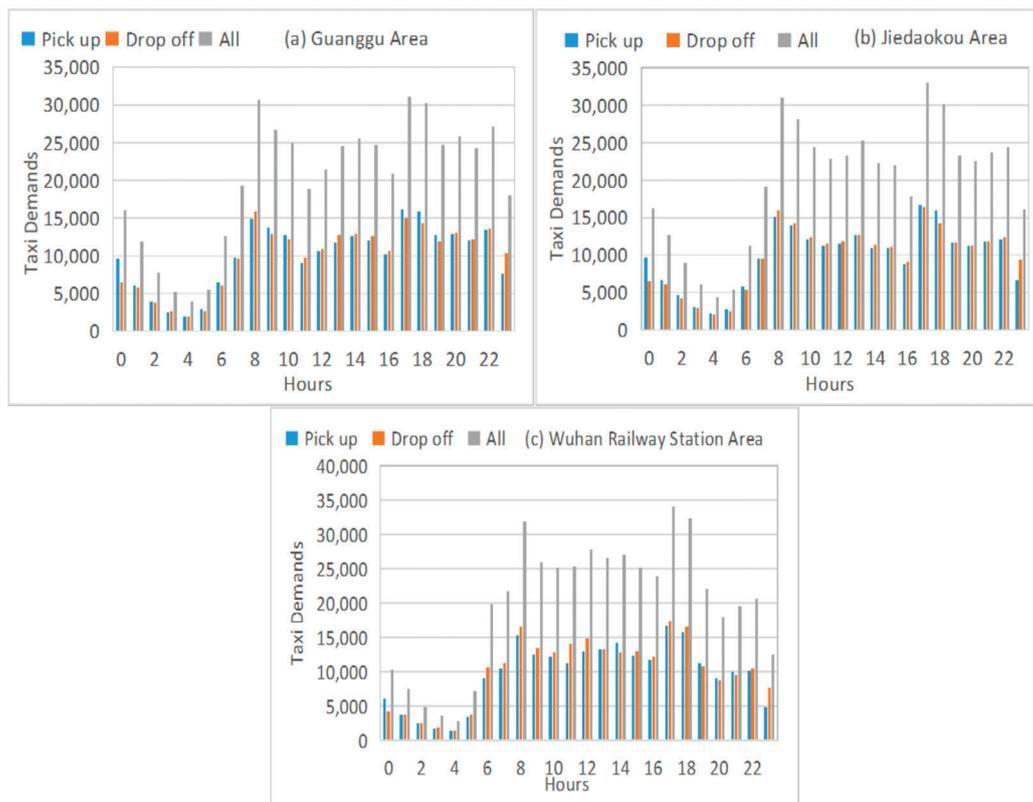


Figure 8. Taxi demands distribution in the city of Wuhan during daily hours.

As we can see in Figure 8, the three study areas have some shared patterns. For example, all of them showed the minimum rates at 4 am and peak-rush at 8 am and 5 pm.

The taxi demand shows high periodicity at long terms, due to the various changes of taxi demands on Thursdays and Fridays in many weeks are almost similar, in the three areas. The periodicity of short-term and long-term can assist in forecasting taxi demands in city areas in the future.

4.2.3. Taxi Demands and POI

In general, the number and type POI may have a significant influence on taxi demand numbers. Figure 9 shows the distribution of POIs and taxi demands in the three areas of Wuhan city.

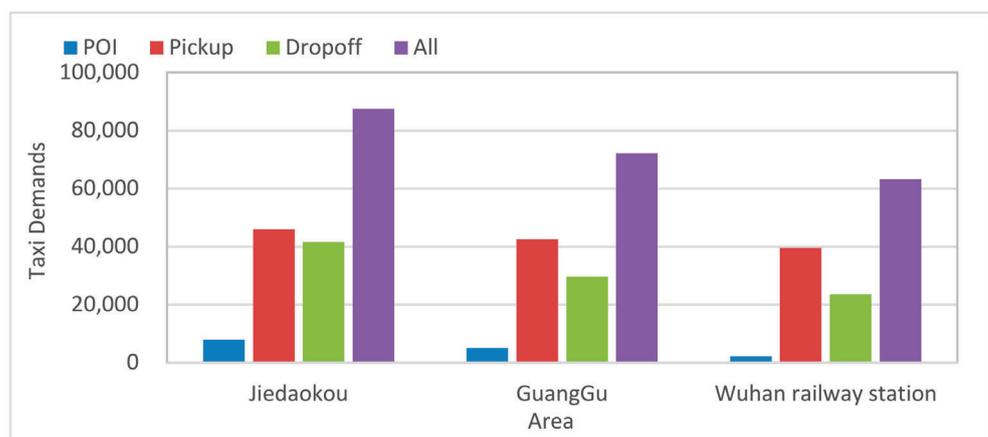


Figure 9. Distribution of POIs and taxi demands in different regions.

From Figure 9, we can find that the distribution of POIs and taxi demand is unbalanced over the three areas, and the more POIs there are in each area, the more taxi demands. For instance, POIs and taxi demands in Jiedaokou and Guanggu are at high numbers, while those in Wuhan railway are at lower numbers.

4.2.4. Taxi Demands and Weather Factors

Besides the above factors, taxi demands are also influenced by weather conditions [22]. This study considers various weather-related factors, including weather conditions (e.g., fog, clear, and rainy), temperature, wind speed, and humidity information, etc., which are reported each hour. Figure 10 shows the general distribution by percentages of different weather types during the study periods.

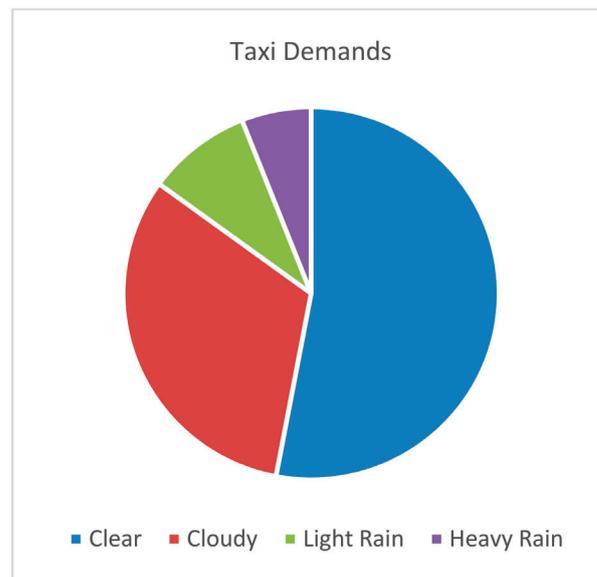


Figure 10. General distribution by percentages of different weather types.

Figure 10 shows the general taxi demands of the three-study area in a day with various weather-related factors.

From Figure 11, we can observe a distinctly remarked impact of weather on taxi demands. We can find that taxi demands significantly increased on a rainy day, and more people may prefer to take taxis when it rains during 7–9 am. This can assist in providing information of taxi traffic flow and human traveling behavior. In adaption, with the functions of each area (using POI) we can understand the flow of passengers.

4.3. Features Preparation

After extracting and analyzing features and taxi demands, there is a need to prepare the features to be ready for utilization in the proposed model. Here we perform three steps: features normalization, transformation, and concatenation.

4.3.1. Feature Transformation

For the sake of reducing the complexity of computing factors, we transformed factors into categorical values, including: Is weekend, Is holiday, region, weather, etc. These factors' values are converted to be numbers that begin with 0 and end with the summation number of the values. For instance, Is weekend factor's values would be represented by 1 for a weekend and 0 for a weekday; weather condition's value would be 0,1,2,3,4 for Clear, Cloudy, Rain, Light Rain, and Heavy Rain, respectively.

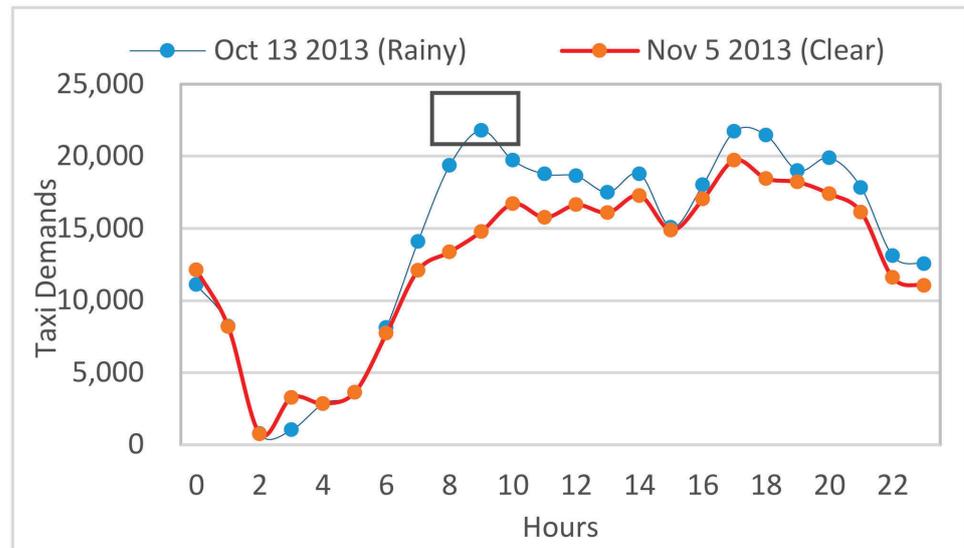


Figure 11. Weather Influence on Taxi demands.

4.3.2. Feature Normalization

We use the Min-Max normalization [0, 1] standard to reduce the absolute scale’s effects. The normalization process is performed on continuous values as follows:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x , x_{min} , x_{max} , x_{norm} are the original, minimum, maximum and normalized values from the dataset (training dataset), respectively.

4.3.3. Feature Concatenation

To facilitate the dataset training, we embed the features in Table 4 into a 1×11 vector according to a time-step t . Figure 12 depicts Feature Concatenation.

Table 4. Features in the dataset.

Factor	Symbol	Data-Type	Description
Output (y)			
Taxi demands	dems	continuous	Summation of Taxi pickups and dropoffs
Inputs (x)			
Date	dt	time series	Date Format month-date
Time	tm	time series	Time Format hour:mints
Is rush-time	rush	categorical	1. Yes 0. No
Is weekend	weekend	categorical	1. Yes 0. No
Is holiday	holiday	categorical	1. Yes 0. No
Region	reg	categorical	Determining region by longitude and latitude
Temperature	temp	continuous	Temperature
Weather	weath	categorical	0. Clear 1. Light Rain 2. Heavy Rain 3. Fog
Wind	wind	continuous	Wind (km\h)
Humidity	humid	continuous	Humidity by percentage
poi_counts	poi	continuous	POI number in region

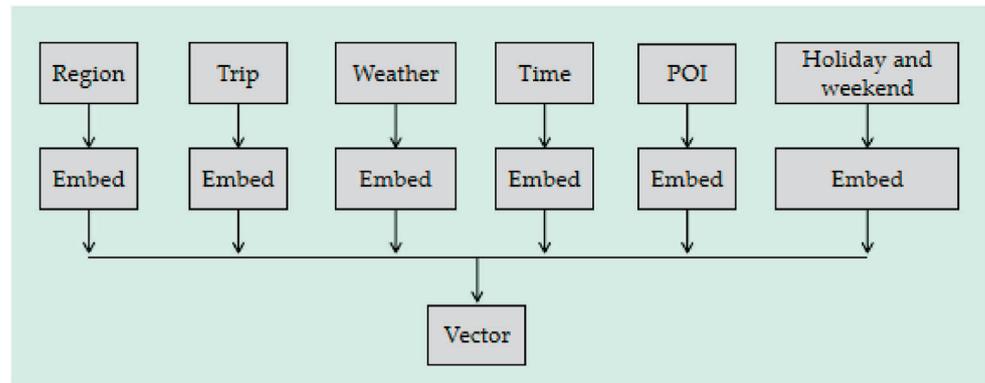


Figure 12. Feature Concatenation.

After performing features normalization, transformation, and concatenation, the dataset was ready for forecasting by prediction models. The features included in the dataset are shown in Table 4.

4.4. Generative Adversarial Network (GAN) Model

A generative adversarial network (GAN) contains two deep neural networks, a generator, and a discriminator. The generator network provides a (fake) dataset fed to the discriminator with the real data. The discriminator network determines and distinguishes the real data and the fake (generated) ones. During the model’s training process, both the generator and the discriminator’s weights would be updated using the related loss function. Once the discriminator becomes unable to differentiate the two types of data, it terminates the training process, and then the model becomes ready to be used; otherwise, in GANs, both components can be any deep neural network. In the following, we introduce LSTM and CNN and then illustrate the proposed GAN model used in this study.

4.4.1. LSTM and CNN

- LSTM

LSTM neural networks are emerged to add long-term memory function, which enhanced Recurrent Neural Networks’ ability to deal with more complicated issues, including prediction and classification [29]. In the LSTM network, the input vector sequence X would be mapped to an output vector sequence y by i iterations. Figure 13 depicts the architecture of an LSTM cell.

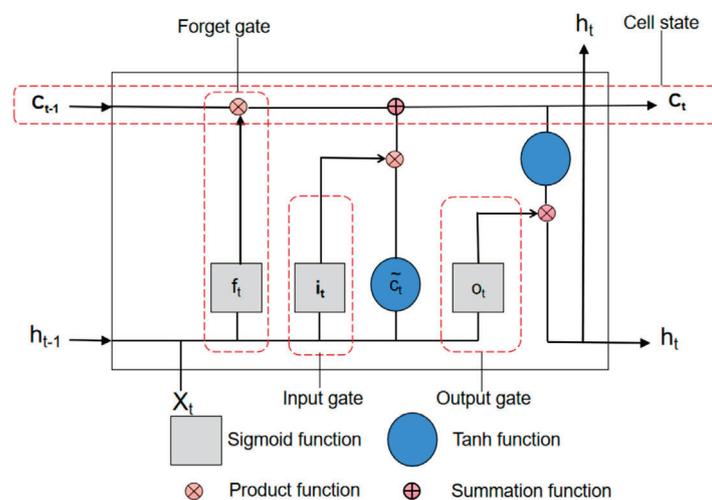


Figure 13. The architecture of a traditional LSTM network.

An LSTM cell contains three layers: an input layer, an output layer, and a memory block layer. The memory block includes three types of gates, including the input gate, the output gate o_t , and the forget gate f_t . Besides, c_t and c_t as memory cell vectors and the candidate value, respectively. The notation t represents a random time-step. During the training process, the related information can be updated as the following formula [30]:

$$i_t = \sigma(W_i[h_{t-1}, \{X\}_t] + b_i) \tag{2}$$

$$f_t = \sigma(W_f[h_{t-1}, \{X\}_t] + b_f) \tag{3}$$

$$c = \tanh(W_c[h_{t-1}, \{X\}_t] + b_c) \tag{4}$$

$$c_t = f_t \times c_{t-1} + i_t \times c_t \tag{5}$$

$$o_t = \sigma(W_o[h_{t-1}, \{X\}_t] + b_o) \tag{6}$$

$$h_t = o_t \times \tanh(c_t) \tag{7}$$

where function $\sigma(\cdot)$ is a sigmoid function can be computed as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{8}$$

W_i , W_f , and W_c are the weight and the of the input gate, input gate, and input gate, respectively, whereas the function $\tanh(\cdot)$ is a tangent function that is computed as follows:

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{9}$$

- CNN

In general, Convolutional Neural Networks (CNNs) are composed of a convolutional cell group, pooling layers, and a set of fully connected layers, as depicted in Figure 14.

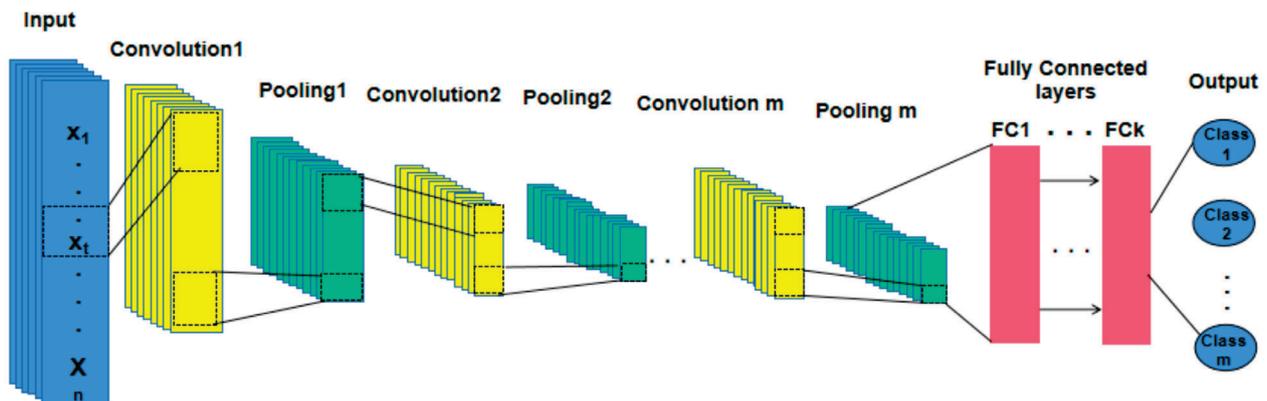


Figure 14. The architecture of traditional Convolutional Neural Network.

A convolutional layer i includes a group of filters $W_i = \in \mathbb{R}^{S \times D \times N}$ that is convolved with an input tensor, S denotes the filters' number, D is a filter's size, and N represents the input channels' number [31]. A pooling layer can pool the output of a convolutional layer. Both convolutional and pooling layers are adopted to obtain the temporal patterns and to correlate temporally distant features. A set of fully connected layers follows the last convolutional/pooling layer to classify the input time series. The network output showed the dataset's classification result and provided one of the two labels (real, predicted) for each time step.

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{10}$$

4.4.2. GAN_LSTM_CNN Model

In the proposed model, and due to their stability, the generator component adopted LSTM networks and CNN for the discriminator network. Figure 15 illustrates the architecture of the GAN_LSTM_CNN Model.

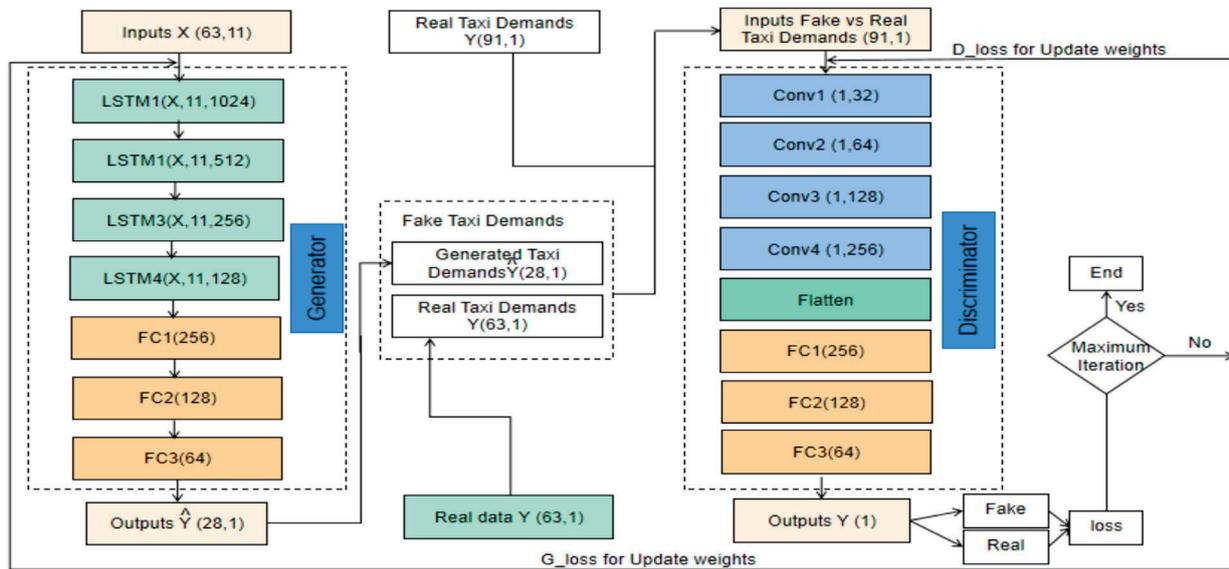


Figure 15. The architecture of GAN_LSTM_CNN model.

- The generator

In our GAN model, we set the LSTM network as the generator according to its stability. The dataset contains 12 features (for 24 h of 91 days) as listed in Table 4. For building up a robust generator network with good performance, we use four layers of LSTM; the numbers of the neuron are 1024, 512, 256, and 128, followed by three fully connected layers. The activation function used in LSTMs is ReLU, and the dropout neurons are 20 percentage, and the neuron number of the latest layer will be the same as the output step we are going to predict.

- The discriminator

The discriminator in the proposed GAN model is a Convolutional Neural Network that aimed to distinguish whether the input data of the discriminator is real or fake. The input for the discriminator will be from the real data or the fake data from the generator. The discriminator network contains four Convolution layers with 32, 64, 128, and 256 neurons separately. Besides, a flatten layer is combined to flatten the output of the convolutional layers to generate a single feature vector for classification, which is performed by the following four fully connected layers with 220, 220, 220, and a neuron, respectively. The Leaky Rectified Linear Unit (ReLU) has been set as the activation function on all layers, except the output layer (adopted Sigmoid function). The sigmoid function will produce a single scalar output, 0 or 1, which means real or fake as the final result.

4.4.3. Training Process

In the proposed model, the generator generates fake data and tries to fool the discriminator, whereas the discriminator tries to distinguish the real data from the predicted data. Thus, there is a need to perform dataset training on the generator and the discriminator. During the training process, the cross-entropy loss is utilized in our GAN model to minimize the difference between the two data distributions.

In training the discriminator, we aim to maximize its objective function, the probability of assigning the correct label to the samples, the loss function of the discriminator is defined as follows:

$$D_loss = -\frac{1}{m} \sum_{i=1}^m \log D(y^i) - \frac{1}{m} \sum_{i=1}^m (1 - \log D(G(x^i))) \quad (11)$$

where x is the features vector, y is taxi demands from the real data, $G(x^i)$ is the generated taxi demands produced by the generator. Then we train the generator to minimize the loss function which is obtained as follows:

$$G_loss = -\frac{1}{m} \sum_{i=1}^m (1 - \log D(G(x^i))) \quad (12)$$

Through the training process, it always needs to minimize the loss function to obtain better results. In the proposed model, we adopted cross-entropy to calculate our loss for both generator and discriminator. In the discriminator, we combined the generated taxi demands with the historical taxi demands of input steps as our input for the discriminator. This step enhances the data length and increases the accuracy for the discriminator to learn the classification. In addition, the batch size is 30, and the epochs' number is 400. We apply Leaky ReLU for fully connected layers as the activation Nesterov Accelerated Gradient (NAG) function. We utilized the NAG algorithm with a learning rate of 0.01 [31]. Besides, to avoid over-fitting, we implemented the dropout method with a probability of 0.2.

5. Experiment

This section introduces the experimental setting, baselines, evaluation metrics and shows the results of the proposed model and baselines in detail.

5.1. Experimental Settings

The main aim of this paper is to perform taxi demand prediction for the three-study areas in the city of Wuhan, including Jiedaokou, Guanggu, and Wuhan railway station, following 18 days with the data of the past 73 days (whole dataset 91 days in 24 h). During training the forecasting model, the input data contains not only the historical taxi demands but also 11 related features that might have a significant effect on the taxi demands. In the training process, the dataset will be split into a training dataset of 80% (73 days with 3264 observations) and a testing dataset of 20% (18 days with 649 observations).

Our model is implemented on a hardware environment of two GPUs, NVIDIA GeForce RTX 2070 with 32 Gigabyte memory, and executed by the conjunction of scikit-learn with the TensorFlow framework.

5.2. Model Comparison

To illustrate the effectiveness of our model, we compare the performance along with six mainstream baselines and tune the parameters for all methods. The models used for the comparison are as follows.

5.2.1. Auto-Regressive Integrated Moving Average (ARIMA)

In ARIMA (p,d,q) , the parameters p and q are related to the order of the autoregressive term and the moving average term, respectively, while the parameter d indicates the d th order different from the original data series, which points to remove the trend from the data series [11]. In this study, these parameters were optimized by the auto-optimal function in the forecast functions in the scikit-learn package in python.

5.2.2. Gradient Boosting Decision Tree (XGBoost)

Researchers widely use XGBoost to perform traffic flow prediction and provide state-of-the-art results [32].

5.2.3. Multiple Layer Perceptron (MLP)

Our model is compared with the MLP method, which contains four hidden layers. The structure of every layer includes 128, 128, 128, and 64 hidden units, respectively.

5.2.4. GAN_LSTM Model

In this model, both the generator and the discriminator have a similar structure as in the generative network G in the GAN_LSTM_CNN model, i.e., four LSTM layers and three fully connected layers. A flattening layer would be added as in the discriminator network in the GAN_LSTM_CNN model. Figure 16 shows the structure of the GAN_LSTM model.

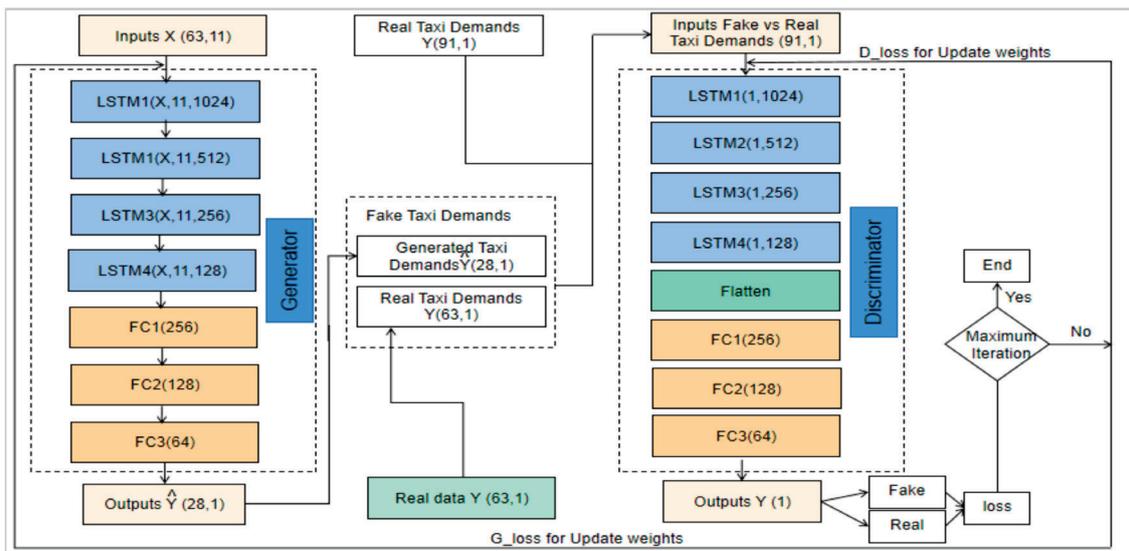


Figure 16. The structure of GAN_LSTM model.

5.2.5. GAN_CNN Model

In this model, the structure of the generator and discriminator is as same as the discriminator network D in the proposed GAN_LSTM_CNN model, i.e., four convolutional layers and three fully connected layers. Figure 17 represents the structure of the GAN_LSTM model.

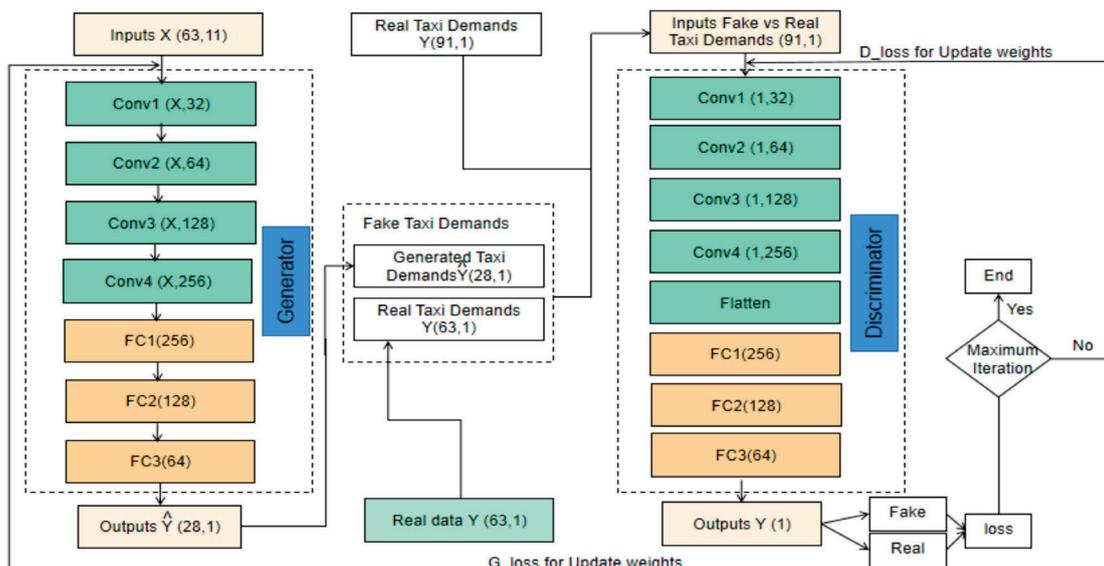


Figure 17. The structure of GAN_CNN model.

5.3. Experimental Metrics

For the sake of assessing the verification and validation of a predictive model, many measures for assessing the predictive accuracy have been used [14,22]. Such as MAE, root MSE (RMSE) r and r^2 are among the most commonly used or recommended measures [14,22,30]. Therefore, the commonly used measures, mean absolute error (MAE), root mean square error (RMSE), and the mean absolute percentage error, are considered in this study. The metrics are computed as the following formulas [29]:

$$MAE(i) = \frac{1}{T} \sum_{t=1}^T \left| \hat{Y}_{i,t} - Y_{i,t} \right| \tag{13}$$

$$RMSE(i) = \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_{i,t} - \hat{Y}_{i,t})^2} \tag{14}$$

$$MAPE(i) = \frac{1}{T} \sum_{t=1}^T \frac{\left| Y_{i,t} - \hat{Y}_{i,t} \right|}{Y_{i,t} + a} \tag{15}$$

where $Y_{i,t}$ is a real taxi demand in the area i at the time-step t , whereas $\hat{Y}_{i,t}$ is the predicted taxi demand. The constant number a in the Equation (15) is a small parameter ($a = 1$) to avoid division by zero situation when both $Y_{i,t}$ and $\hat{Y}_{i,t}$ are 0.

Thus, the forecasting performance by the three areas at time-step t can be defined as the followings:

$$sMAPE_{all} = \sum_{t=1}^T \frac{\left| Y_{i,t} - \hat{Y}_{i,t} \right|}{Y_{i,t} + \hat{Y}_{i,t} + a} \tag{16}$$

$$RMSE_{all} = \sqrt{\sum_{t=1}^T (Y_{i,t} - \hat{Y}_{i,t})^2} \tag{17}$$

$$MAE_{all} = \sum_{i=1}^T \frac{\left| \hat{Y}_{i,t} - Y_{i,t} \right|}{Y_{i,t}} \tag{18}$$

By considering the initial batch size was 36, and the epoch size was 500, it was 50 iterations per epoch and totally about 25,000 iterations.

6. Results

6.1. Comparison with Baselines

This section illustrates a detailed comparison among models by experiential metrics as shown in Section 5.3. Firstly, we report the performance on MAE, sMAPE, and RMSE over the three areas together. Secondly, we show the prediction performance at the all specific areas as time passes, separately.

6.1.1. Performance Comparison over the Three Areas Together

To assess the forecasting performance over the three study areas together (Jiedaokou area, Guanggu area, and Wuhan railway station area), we analyze the performance of the ARIMA, XBoost, MLP, GAN_LSTM_CNN model, GAN_LSTM, and GAN_CNN model in terms of MAE, RMSE, and sMAPE as defined in Equations (16)–(18). The comparison results are shown in Figures 18–20.

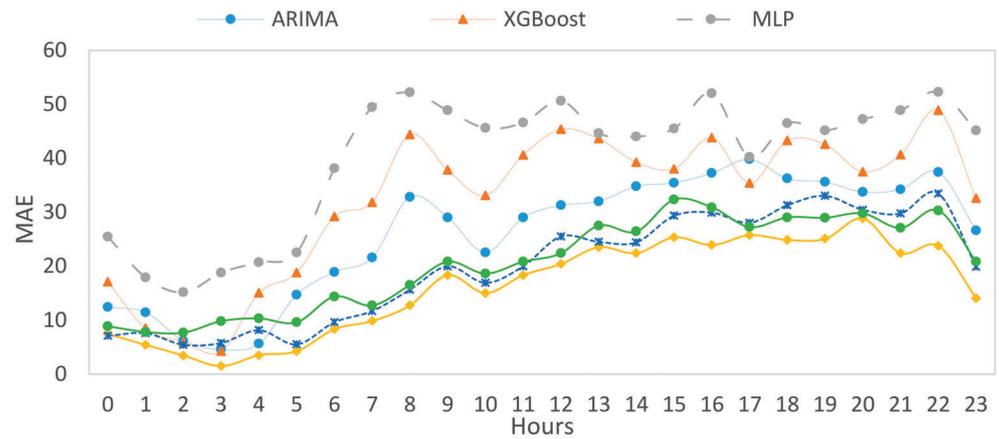


Figure 18. MAE Comparison Results with different time slots. MLP, XGBoost attained high MAE whereas GAN_LSTM_CNN achieved better MAE.

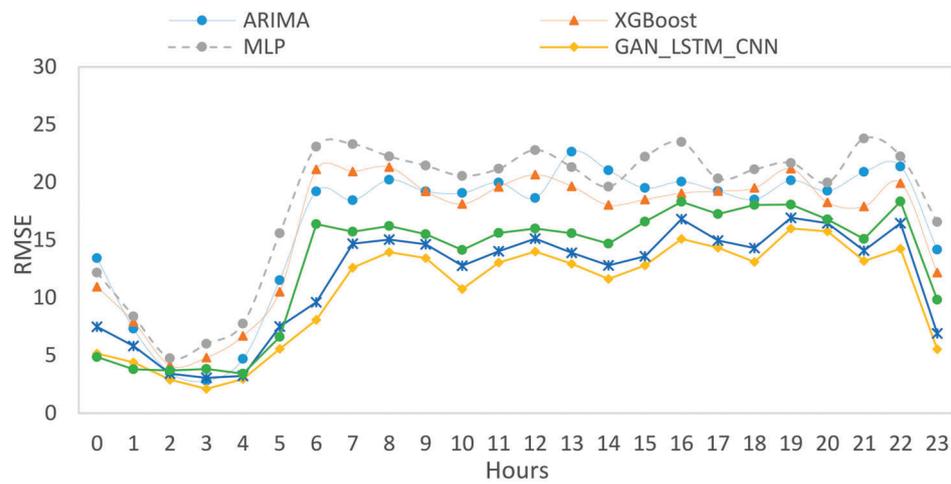


Figure 19. RMSE Comparison Results with hours ranged from 0–23 with increasing values from 6:00 to 22:00 and all deep learning methods achieved better results.

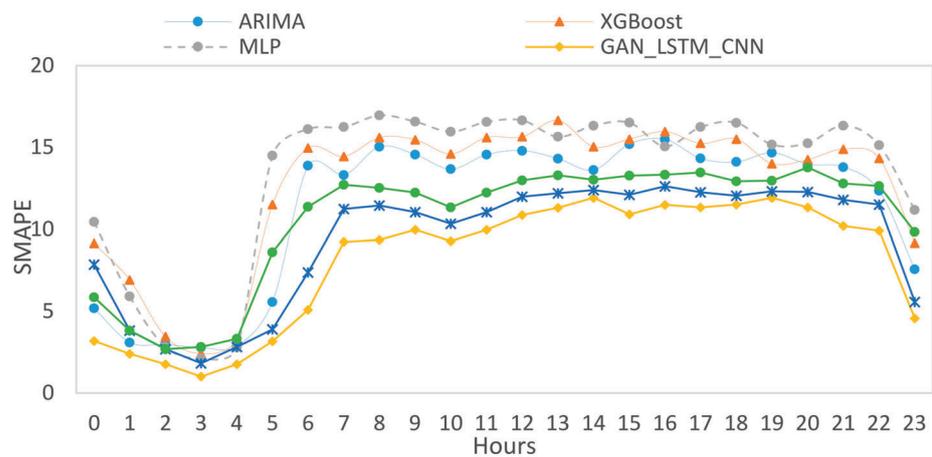


Figure 20. sMAPE Comparison Results during time slots. The achieved values showed deep learning methods (such as GAN_CNN, GAN_LSTM, GAN_LSTM_CNN) achieved better results comparing to ARIMA, XGBoost an MLP.

From the three figures above, we can notice that although they are different forecasting performance metrics, there are some common patterns. For example, all metrics reach the minimum rates at three hours 3, 10, and 23 as the with lowest number of pickups wrt their neighboring time slots have. Moreover, the GAN_LSTM_CNN model presents a better performance compared to other predictors.

Figure 18 plots the mean absolute error (MAE), which describes the prediction accuracy. There is a dramatic decrease at the early hours of a day until it reaches 1.4 (the lowest point) in the GAN_LSTM_CNN model at 3 am, following with a gradual increase until 22, and then a further sharp fell is found.

As shown in Figures 19 and 20, sMAPE presents the mean absolute error, whereas RMSE shows the root mean squared error between the predicted demands and the real demands. The values dropped from 3 am until 10 am, in which the value becomes lower again, then a noticeable increase has emerged again. GAN_LSTM_CNN model also provides the minimum MAPEs and RSMEs, which shows the highest accuracy between the predicted models adopted in this paper.

6.1.2. Performance Comparison over the Three Areas Separately

Our proposed model’s performance and the baselines for the three study areas, namely Jiedaokou, Wuhan Railway Station Area, and Guanggu, are shown in Tables 5–7.

Table 5. Performance Comparison Jiedaokou Area.

Method	MAPE	RMSE	MAE
ARIMA	11.32	16.46	25.97
XGBoost	12.66	16.23	32.40
MLP	13.46	18.40	40.16
GAN_LSTM	9.58	10.82	18.71
GAN_CNN	9.35	11.10	18.46
GAN_LSTM_CNN	8.06	10.57	16.19

Table 6. Performance Comparison Wuhan railway Station Area.

Method	MAPE	RMSE	MAE
ARIMA	13.83	15.16	22.73
XGBoost	14.16	17.12	31.85
MLP	15.25	18.83	39.75
GAN_LSTM	8.67	10.86	18.11
GAN_CNN	8.91	11.79	18.73
GAN_LSTM_CNN	7.87	10.44	16.87

Table 7. Performance Comparison Guanggu Area.

Method	MAPE	RMSE	MAE
ARIMA	11.32	16.46	25.67
XGBoost	12.66	16.22	32.85
MLP	13.46	18.23	40.75
GAN_LSTM	9.38	10.87	18.73
GAN_CNN	9.35	10.17	18.78
GAN_LSTM_CNN	8.94	9.78	16.75

As we can see from the three Tables above, we can notice that the GAN_LSTM_CNN model presents the lowest MAE, RMSE, and MAPE in the three areas, compared with the other baselines. Specifically, MLP and XGboost perform the poorest, whereas GAN_LSTM and GAN_CNN achieve good performance. Therefore, it proves that GAN-based models achieve better performance than other baselines. It confirms that deep neural networks can work efficiently in taxi demand forecasting. Compared with GAN_LSTM,

GAN_LSTM_CNN model achieves 15.86% (2.31%, 13.46%) lower MAPE (RMSE, MAE) on Jiedaokou area, 9.22% (3.86%, 6.84%) lower MAPE (RMSE, MAE) on Wuhan railway Station area, and 4.69% (10.02%, 6.84) lower MAPE (RMSE, MAE) and on Guanggu area, respectively. Compared with GAN_CNN, our model reduced MAPE (RMSE, MAE) by 13.79% (4.77%, 12.29%) on the Jiedaokou area, 11.67% (11.45%, 9.93%) on Wuhan railway Station Area, and 4.38% (3.83%, 10.80%) on Guanggu area, respectively.

6.2. Comparison of Training Loss Function

The optimizer used in our model is the Adam algorithm with a learning rate of 0.0001. The batch size is 128, and we trained the model for 400 epochs. Figure 21 shows the Comparison of Training Loss Function among GAN_LSTM and GAN_CNN and GAN_LSTM_CNN models.

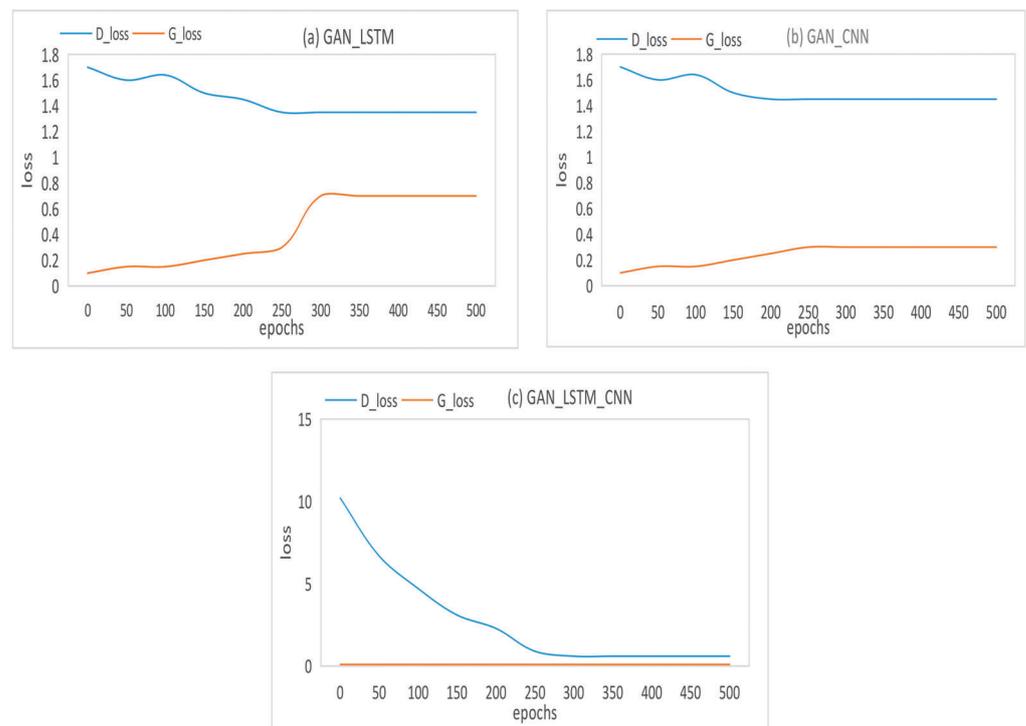


Figure 21. Training loss of the Generator(G_loss) and Discriminator(D_loss) of (a) GAN_LSTM, (b) GAN_CNN and (c) GAN_LSTM_CNN models. In GAN_LSTM and GAN_CNN, the loss of discriminator is higher than the loss of generator, and through the training process, both loss paths become flat. Compared with GAN_LSTM and GAN_CNN, the discriminator in GAN_LSTM_CNN learns better and faster.

Figure 21a,b showed the loss of the GAN_LSTM and GAN_CNN. The blue line represents the loss path of the discriminator, and the orange line is the generator's loss path. From the beginning, the loss of discriminator is higher than the loss of generator, and through the training process, both loss paths become flat.

In Figure 21c, the discriminator loss of GAN_LSTM_CNN decreases dramatically towards 0. Compared with GAN_LSTM and GAN_CNN, the discriminator in GAN_LSTM_CNN learns better and faster.

6.3. Comparison of Time Efficiency

We estimate the time efficiency reached by the GAN_LSTM_CNN model and GAN-based baselines (i.e., GAN_LSTM and GAN_CNN) in terms of the time consumption of training and testing.

As shown in Table 8, time spent for training and testing by GAN_LSTM is significantly larger than GAN_CNN and GAN_LSTM_CNN, even though its number of trainable parameters is the fewest. Moreover, we can find that our proposed GAN_LSTM_CNN achieves higher time consumption during training and testing processes than GAN_CNN. This is because the trainable parameters in GAN_LSTM_CNN are almost 22 times as much as in GAN_CNN.

Table 8. Comparison of Time Efficiency.

Method	Parameters	Training (for Each Epoch)	Testing (for Each Batch)
GAN_LSTM	8320	233.3 s	1.5 s
GAN_CNN	443,003	50.9 s	1.2 s

7. Discussions

In this study, we try to investigate taxi demands and related factors extracted from multi-source data. A generative adversarial network (GAN) is adopted to forecast taxi-passenger demand by considering significant factors, including temporal, spatial, and external factors. Using GANs, the experiments showed that overall performance is very good compared to similar methods.

Comparing this study to the study by [2,5], we highlight several similarities and dissimilarities. Regarding their methodology and results, in particular, this study has several differences and innovations.

Firstly, a comprehensive experiment was conducted by GANs using multi-source data collected for the city of Wuhan in China, which covers all variables associated with taxi demand. The obtained variables include time, Is rush-time, Is weekend, Is holiday, Region, temperate, Weather, Wind, Humidity, and POI counts. In contrast, key variables were not included in the methods of Kuang Li et al.'s study [2] and Jun Xu et al.'s study [5], such as the POI counts, Is rush-time, Is weekend, Is holiday, Region, temperate, etc. These variables undoubtedly play a vital role in forecasting the taxi-passenger demand, and some of these variables were significant according to the results of this study. Not surprising that these factors are considered significant in forecasting taxi-passenger demand. It conforms to the findings of a previous study [16,18,29].

Secondly, Kuang Li et al.'s study [2] adopted the Convolutional Neural Network (CNN) to study contributing factors to taxi demand, but this study adopted a generative adversarial networks model. Three kinds of GANs have been adopted, namely the GAN_LSTM_CNN model, GAN_LSTM model, and GAN_CNN model. The results comparison indicated that the GAN_LSTM_CNN model outperformed the GAN_LSTM model and the GAN_CNN model in MAE, RMSE, and sMAPE introduced in Figures 18–20. It shows that the GAN_LSTM_CNN model can present a more comprehensive view of the impact of significant factors on taxi demands.

Moreover, this study presents an analysis of the Training Loss Function among the GAN_LSTM_CNN model, GAN_LSTM model, and GAN_CNN model. It is important for considering computing consumed time during the Training process. The discriminator in GAN_LSTM_CNN shows that the model may learn better and faster, which can improve the efficiency and effectiveness of taxi demand prediction.

Finally, although the training time consumed by GAN_LSTM is considerably higher than those in GAN_CNN and GAN_LSTM_CNN, even though the trainable parameters are the lowest number. In addition, GAN_LSTM_CNN results in higher time consumption during training and testing processes than in GAN_CNN. This is reasonable as the training parameters in GAN_LSTM_CNN are 22 times bigger than those in GAN_CNN.

The findings of this study indicate that our hypothesis is correct. The results highlighted the vital role which deep learning-based methods can play in improving the accuracy of taxi demand forecasting. And this can help policy-makers regulating the taxi industry to enhance the temporal taxi supply during specific periods, which in turn can increase taxi profit.

Although the used dataset contains various vital variables, there are three limitations found during this study. The first is that more datasets can be collected for a more extended period, such as a year or two years. It may enhance the results and may provide a more accurate understating of taxi demands. The second limitation is the dataset used in this study is a historical dataset, not a real-time dataset. A real-time dataset may help in real-time traffic situations, where some areas have great demands and drivers compete for getting passengers in another area of the city. Thirdly, as the dataset of the whole city are huge, and beyond the capability of computers' resources, we investigated three specific areas. If we have a considerable computer resource for dealing with such big data, a more comprehensive understanding of taxi demands would be obtained, which may help enhance taxi demands prediction.

8. Limitations

Although the used dataset contains various vital variables, there are three limitations found during this study. The first is that more datasets can be collected for a longer period such as a year or two years. This may enhance the results and may provide a more accurate understating of taxi demands. The second limitation is the dataset used in this study is a historical dataset, not a real-time dataset. Real-time dataset may help in real-time traffic situations, where some areas have large demands and drivers are competing with each other for getting passengers in another area of the city. Thirdly, as the dataset of the whole city are huge and beyond the capability of computers' resources, we investigated three specific areas. If we have a huge computer resource for dealing with such big data, a more comprehensive understanding of taxi demands would be obtained which in turn may help in taxi demands prediction.

9. Conclusions

Accurate taxi demand forecasting can solve the traffic congestion problem caused by the supply-demand imbalance. Although many methods have been successfully employed to address the taxi demand prediction problem, most existing methods may not comprehensively consider various factors that influence the forecasting results. To fill the gap, we propose a deep learning-based model for forecasting taxi demands in the urban area by considering multi-source data.

Various factors have been considered, including trips factors, temporal factors, spatial factors, weather conditions, and POI. Firstly, in the proposed model, significant factors are extracted from raw data and then analyzed to understand the influences of these factors on taxi demands. Pick-up locations of Taxi trips are derived from taxi GPS trajectory, combined with temporal factors, weather conditions, POI data, and road network data. All information is then integrated to explore the travel pattern of taxi demand and its related influences.

Secondly, the extracted factors are prepared for use in the forecasting model. Normalization, transformation, and concatenation were employed. Thirdly, the Generative Adversarial Networks (GANs) structure is introduced, followed by a training process setting. The convolutional recurrent neural network model. In the proposed model, the LSTM network is adopted as the generator according to its stability. The convolutional neural network (CNN) is employed to distinguish whether the input data is real or generated by the generator. Finally, comprehensive experiments are performed on real-world datasets. The proposed model can automatically learn various characteristics to understand spatiotemporal patterns and enhance forecasting performance. For proving the xxx of predictive accuracy, the proposed model is validated and compared with several benchmark algorithms, including the ARIMA, XBoost, MLP, GAN_LSTM, and GAN_CNN on the real-world data of Wuhan city.

The results show that our model outperforms the other prediction approaches in the measurements of MAPE, RMSE, MAE, and time-consuming.

The evidence from findings proves that considering a deep learning-based approach and considering spatial-temporal, weather, road network correlations in models can significantly improve predictive accuracy. The results can assist policymakers in regulating the taxi industry to enhance the temporal taxi supply and the taxi lease rents (which vary by shift) for specific periods, which may improve passengers' degree of satisfaction and improve the transportation capacities of the taxi industry cities.

For future work, there is a need to investigate the impact of other similar types of traffic demands such as online car-hailing services (for instance, didi and ober), which lead the taxi demands to be reduced. Moreover, the spatial correlation between city areas would be considered and then forecast taxi demand by graph-based deep neural networks.

Author Contributions: H.A.H.N. designed, developed the methodology, collected and analyzed the data; Q.X. supervised the work and provided analysis tools; H.A.H.N. wrote the paper; H.Z. performed data curation and visualization; T.L. project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by Foundation of Nanyang Institute of Technology (grant no.: 510102), Projects of Henan Provincial Department of Science and Technology (no. 212102310297).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bureau of Wuhan City's Transportation. Analysis of Urban Passenger Transport Operation in Wuhan. 2021. Available online: http://jtj.wuhan.gov.cn/zwgk/zfxgkml/tjsj/202102/t20210202_1624093.shtml (accessed on 21 August 2021).
2. Kuang, L.; Yan, X.; Tan, X.; Li, S.; Yang, X. Predicting Taxi Demand Based on 3D Convolutional Neural Network and Multi-task Learning. *Remote Sens.* **2019**, *11*, 1265. [CrossRef]
3. Qian, X.; Ukkusuri, S.V. Time-of-day pricing in taxi markets. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1610–1622. [CrossRef]
4. Le Quy, T.; Nejdil, W.; Spiliopoulou, M.; Ntoutsis, E. In A neighborhood-augmented LSTM model for taxi-passenger demand prediction. In *International Workshop on Multiple-Aspect Analysis of Semantic Trajectories*; Springer: Cham, Switzerland, 2019; pp. 100–116.
5. Xu, J.; Rahmatizadeh, R.; Boloni, L.; Turgut, D. Real-Time Prediction of Taxi Demand Using Recurrent Neural Networks. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 1–10. [CrossRef]
6. Niu, K.; Cheng, C.; Chang, J.; Zhang, H.; Zhou, T. Real-Time Taxi-Passenger Prediction With L-CNN. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4122–4129. [CrossRef]
7. Tang, J.; Liang, J.; Zhang, S.; Huang, H.; Liu, F. Inferring driving trajectories based on probabilistic model from large scale taxi GPS data. *Phys. A Stat. Mech. Appl.* **2018**, *506*, 566–577. [CrossRef]
8. Moreira-Matias, L.; Gama, J.; Ferreira, M.; Damas, L. A predictive model for the passenger demand on a taxi network. In *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems*, Anchorage, AK, USA, 16–19 September 2012; pp. 1014–1019.
9. Liu, Z.; Chen, H.; Li, Y.; Zhang, Q. Taxi Demand Prediction Based on a Combination Forecasting Model in Hotspots. *J. Adv. Transp.* **2020**, *2020*, 1–13. [CrossRef]
10. Chang, H.-W.; Tai, Y.-C.; Hsu, J.Y.-J. Context-aware taxi demand hotspots prediction. *Int. J. Bus. Intell. Data Min.* **2010**, *5*, 3–18. [CrossRef]
11. Jamil, M.S.; Akbar, S. Taxi passenger hotspot prediction using automatic ARIMA model. In *Proceedings of the 2017 3rd International Conference on Science in Information Technology (ICSITech)*, Bandung, Indonesia, 25–26 October 2017; pp. 23–28.
12. Li, X.; Pan, G.; Wu, Z.; Qi, G.; Li, S.; Zhang, D.; Zhang, W.; Wang, Z. Prediction of urban human mobility using large-scale taxi traces and its applications. *Front. Comput. Sci.* **2012**, *6*, 111–121.
13. Zhang, L.; Chen, C.; Wang, Y.; Guan, X. Exploiting taxi demand hotspots based on vehicular big data analytics. In *Proceedings of the 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Montreal, QC, Canada, 18–21 September 2016; pp. 1–5.
14. Markou, I.; Rodrigues, F.; Pereira, F.C. Real-Time Taxi Demand Prediction using data from the web. In *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, USA, 4–7 November 2018; pp. 1664–1671.
15. Liu, Z.; Chen, H.; Sun, X.; Chen, H. Data-driven real-time online taxi-hailing demand forecasting based on machine learning method. *Appl. Sci.* **2020**, *10*, 6681. [CrossRef]
16. Saadi, I.; Wong, M.; Farooq, B.; Teller, J.; Cools, M. An investigation into machine learning approaches for forecasting spatio-temporal demand in ride-hailing service. *arXiv* **2017**, arXiv:1703.02433.
17. Ma, X.; Yu, H.; Wang, Y.; Wang, Y. Large-scale transportation network congestion evolution prediction using deep learning theory. *PLoS ONE* **2015**, *10*, e0119044. [CrossRef] [PubMed]

18. Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X. DNN-based prediction model for spatio-temporal data. In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame CA, USA, 31 October–3 November 2016; pp. 1–4.
19. Rossi, A.; Barlacchi, G.; Bianchini, M.; Lepri, B. Modelling taxi drivers' behaviour for the next destination prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 2980–2989. [[CrossRef](#)]
20. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IEEE Trans. Intell. Transp. Syst.* **2017**, *11*, 68–75. [[CrossRef](#)]
21. Ke, J.; Zheng, H.; Yang, H.; Chen, X. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transp. Res. Part C Emerg. Technol.* **2017**, *85*, 591–608. [[CrossRef](#)]
22. Liu, T.; Wu, W.; Zhu, Y.; Tong, W. Predicting taxi demands via an attention-based convolutional recurrent neural network. *J. Knowl.-Based Syst.* **2020**, *206*, 106294. [[CrossRef](#)]
23. Zhang, J.; Zheng, Y.; Qi, D. Deep spatio-temporal residual networks for citywide crowd flows prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
24. OpenStreetMap. Available online: <https://www.openstreetmap.org/> (accessed on 17 March 2021).
25. Guihuayun. Available online: <http://www.guihuayun.com/> (accessed on 5 November 2020).
26. Timeanddate. Available online: <https://www.timeanddate.com/> (accessed on 10 November 2020).
27. Hasan, N.; Wu, C.; Hui, Z. Understanding the Impact of Human Mobility Patterns on Taxi Drivers' Profitability Using Clustering Techniques: A Case Study in Wuhan, China. *Information* **2017**, *8*, 67.
28. Naji, H.; Wu, C.; Hui, Z.; Li, L. Towards understanding the impact of human mobility patterns on taxi drivers' income based on GPS data: A case study in Wuhan—China. In Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS), Banff, AB, Canada, 8–10 August 2017.
29. Yu, H.; Chen, X.; Li, Z.; Zhang, G.; Liu, P.; Yang, J.; Yang, Y. Taxi-based mobility demand formulation and prediction using conditional generative adversarial network-driven learning approaches. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3888–3899. [[CrossRef](#)]
30. Bogaerts, T.; Masegosa, A.D.; Angarita-Zapata, J.S.; Onieva, E.; Hellinckx, P. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transp. Res. Part C Emerg. Technol.* **2020**, *112*, 62–77. [[CrossRef](#)]
31. Lv, J.; Li, Q.; Sun, Q.; Wang, X. T-CONV: A Convolutional Neural Network for Multi-scale Taxi Trajectory Prediction. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 15–18 January 2018.
32. Vanichrujee, U.; Horanont, T.; Pattara-Atikom, W.; Theeramunkong, T.; Shinozaki, T. Taxi Demand Prediction using Ensemble Model Based on RNNs and XGBOOST. In Proceedings of the 2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES), Khon Kaen, Thailand, 7–9 May 2018; pp. 1–6.

Article

An Improved Multiple Features and Machine Learning-Based Approach for Detecting Clickbait News on Social Networks

Mohammed Al-Sarem ^{1,*}, Faisal Saeed ^{1,2,*}, Zeyad Ghaleb Al-Mekhlafi ^{3,*}, Badiea Abdulkarem Mohammed ³, Mohammed Hadwan ^{4,5}, Tawfik Al-Hadhrami ⁶, Mohammad T. Alshammari ³, Abdulrahman Alreshidi ³ and Talal Sarheed Alshammari ³

- ¹ College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia
 - ² Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, City Campus, Pengkalan Chepa, Kota Bharu 16100, Kelantan, Malaysia
 - ³ College of Computer Science and Engineering, University of Ha'il, Ha'il 81481, Saudi Arabia; b.alshaibani@uoh.edu.sa (B.A.M.); md.alshammari@uoh.edu.sa (M.T.A.); ab.alreshidi@uoh.edu.sa (A.A.); talal.alshammari@uoh.edu.sa (T.S.A.)
 - ⁴ Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia; m.hadwan@qu.edu.sa
 - ⁵ Department of Computer Science, College of Applied Sciences, Taiz University, Taiz 6803, Yemen
 - ⁶ School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK; tawfik.al-hadhrami@ntu.ac.uk
- * Correspondence: msarem@taibahu.edu.sa (M.A.-S.); fsaeed@taibahu.edu.sa (F.S.); ziadgh2003@hotmail.com (Z.G.A.-M.)

Citation: Al-Sarem, M.; Saeed, F.; Al-Mekhlafi, Z.G.; Mohammed, B.A.; Hadwan, M.; Al-Hadhrami, T.; Alshammari, M.T.; Alreshidi, A.; Alshammari, T.S. An Improved Multiple Features and Machine Learning-Based Approach for Detecting Clickbait News on Social Networks. *Appl. Sci.* **2021**, *11*, 9487. <https://doi.org/10.3390/app11209487>

Academic Editors:
Giovanni Randazzo,
Anselme Muzirafuti
and Dimitrios S. Paraforos

Received: 10 September 2021
Accepted: 6 October 2021
Published: 13 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The widespread usage of social media has led to the increasing popularity of online advertisements, which have been accompanied by a disturbing spread of clickbait headlines. Clickbait dissatisfies users because the article content does not match their expectation. Detecting clickbait posts in online social networks is an important task to fight this issue. Clickbait posts use phrases that are mainly posted to attract a user's attention in order to click onto a specific fake link/website. That means clickbait headlines utilize misleading titles, which could carry hidden important information from the target website. It is very difficult to recognize these clickbait headlines manually. Therefore, there is a need for an intelligent method to detect clickbait and fake advertisements on social networks. Several machine learning methods have been applied for this detection purpose. However, the obtained performance (accuracy) only reached 87% and still needs to be improved. In addition, most of the existing studies were conducted on English headlines and contents. Few studies focused specifically on detecting clickbait headlines in Arabic. Therefore, this study constructed the first Arabic clickbait headline news dataset and presents an improved multiple feature-based approach for detecting clickbait news on social networks in Arabic language. The proposed approach includes three main phases: data collection, data preparation, and machine learning model training and testing phases. The collected dataset included 54,893 Arabic news items from Twitter (after pre-processing). Among these news items, 23,981 were clickbait news (43.69%) and 30,912 were legitimate news (56.31%). This dataset was pre-processed and then the most important features were selected using the ANOVA F-test. Several machine learning (ML) methods were then applied with hyperparameter tuning methods to ensure finding the optimal settings. Finally, the ML models were evaluated, and the overall performance is reported in this paper. The experimental results show that the Support Vector Machine (SVM) with the top 10% of ANOVA F-test features (user-based features (UFs) and content-based features (CFs)) obtained the best performance and achieved 92.16% of detection accuracy.

Keywords: ANOVA-test; clickbait news; feature selection; social network

1. Introduction

Currently, social networks have become the main environment for communicating, sharing, and posting news on the Internet. Twitter, Facebook, and Instagram are the main social networks that are used to share our opinions and news. With this development, a huge amount of textual data are posted on these media, which increasingly become difficult to process manually. Although the social networks provide an easy way to express our opinions, this platform also can be used to share misinformation in the form of news and advertisements. This is a very serious issue, because this misinformation has the power to influence individuals and sway their opinions. Therefore, finding a way to protect users of social networks from the spread of this misinformation and develop a reliable mechanism to detect it is very important. This misinformation can take the form of clickbait, which aims at enticing the users into clicking a link to news items or advertisements, whose titles (headlines) do not completely reflect the inside contents. According to Chen et al. [1], clickbait is defined as “Content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page”.

The automatic detection of clickbait headlines from the huge volume of news on social networks has become a difficult research issue in the field of data science. Some previous efforts have utilized machine learning to detect clickbait headlines automatically. For instance, Biyani et al. [2] applied Gradient Boosted Decision Trees (GBDT) on a dataset drawn from news sites such as Huffington Post, New York Times, CBS, Associated Press and Forbes. The dataset contains 1349 clickbait and 2724 non-clickbait webpages. The best results achieved were an F1-score of 61.9% with five-fold cross-validation for the clickbait class and an F1-score of 84.6% for the non-clickbait category. Potthast et al. [3] applied linear regression, Naïve Bayes, and random forest methods on a dataset gathered from Twitter. The dataset contained 2992 data points. The results recorded were relatively close, with an approximate precision of 75%.

Chakraborty et al. [4] built a browser extension that used support vector machine (SVM), decision tree, and random forest to automatically detect the clickbait headlines. For training purpose, they collected a well-balanced dataset which contains 30,000 headlines (clickbait and non-clickbait) from ViralStories, Upworthy, BuzzFeed, Wikinews, Scoopwhoop, and ViralNova. In addition, for each data point in the dataset, they extracted sentence structure, clickbait language, word patterns, and n-gram features. The results they achieved are as follows: SVM: an accuracy rate of 93% with 95% precision, 90% recall, 93% F1-score, and 97% ROC-AUC values; Decision Tree: 90% accuracy rate with 91% precision, 89% recall, 90% F1-score, and 90% ROC-AUC values; Random Forest: 92% accuracy rate, 94% precision, 91% recall, 92% F1-score, and finally; ROC-AUC values of 97% using a combination of all extracted features.

Khater et al. [5] proposed the use of logistic regression and linear SVM. They extracted 28 features from a dataset provided by Bauhaus-Universität Weimar at the time of a clickbait detection challenge. The most commonly extracted features were Bag of Words (BOW), noun extraction, similarity, readability, and formality. The best results achieved were 79% and 78% precision for logistic regression and linear SVM respectively. Since the methods of the first category require extracting and labeling each feature before feeding the data into the machine learning tool, researchers have found that deep learning techniques are useful to overcome the feature engineering phase. For instance, López-Sánchez et al. [6] combined metric learning with a CNN deep learning algorithm by integrating them with case-based reasoning methodology. For feature selection, they used TF-IDF, n-gram, and 300 dimensional Word2Vect using the dataset provided by [4]. The proposed approach achieved average areas of 99.4%, 95%, and 90% under the ROC curve using Word2vec, TF-IDF, and n-gram count. Agrawal [7] also used a CNN model to classify a manually constructed news corpus obtained from Reddit, Facebook, and Twitter social networks into clickbait and non-clickbait. As feature selection methods, they used Click-Word2vec and Click-scratch. The highest results that they achieved were 89% accuracy with 87% ROC-AUC score for Click-scratch features and 90% when the Click-Word2vec was used.

Kaur et al. [8] also proposed a hybrid model where a CNN model is combined with LSTM. They found that the CNN-LSTM model when implemented with pre-trained GloVe embedding yields the best results, based on accuracy, recall, precision, and F1-score performance metrics. They also identify eight other types of clickbait headlines: reaction, reasoning, revealing, number, hypothesis/guess, questionable, forward referencing, and shocking/unbelievable. They also found that shocking/unbelievable, hypothesis/guess, and reaction clickbait types to be the most frequently occurring types of clickbait headlines published online.

Although several machine learning approaches have been proposed to detect clickbait headlines, most of these recent methods are not very robust. The previous studies used hybrid categorization techniques such as Gradient Boosted Decision Trees, linear regression, Naïve Bayes and random forest methods, SVM, decision tree, logistic regression, and convolutional neural network deep learning. Most of these studies used datasets with headlines written in English. However, this paper uses an Arabic language dataset and proposes a comprehensive approach that includes three main phases: data collection, data preparation, and machine learning model training and testing phases. This dataset was pre-processed and then the most important features were selected using ANOVA F-test. Several machine learning methods were then applied which include random forest (RF), stochastic gradient descent (SGD), Support Vector Machine (SVM), logistic regression (LR), multinomial Naïve Bayes (NB), and k-nearest neighbor (k-NN). Hyper-parameter tuning methods were applied to ensure finding the optimal settings. Finally, the ML models were evaluated and the overall performance is reported here. The key contributions of this paper are as follows:

- We constructed the first Arabic clickbait headline news dataset. The raw dataset is available publicly for research purpose.
- We extracted a set of user-based features and content-based features for the constructed Arabic clickbait dataset.
- We implemented six machine learning-based classifiers, including Random Forest (RF), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Logistic Regression (LR), Multinomial Naïve Bayes (NB), and k-Nearest Neighbor (k-NN).
- We proposed an effective approach for enhancing the detection process using a feature selection technique, namely a one-way ANOVA F-test.
- We conducted extensive experiments, and the results show that the proposed model enhances the performance of some classifiers in terms of accuracy, precision, and recall.

2. Related Works

2.1. Characteristics of Clickbait News

Biyani et al. [2] define eight types of clickbait, which include exaggeration, teasing, inflammatory, formatting, graphic, bait-and-switch, ambiguous, and wrong. In exaggeration, the title overdraws the content on the target page. Teasing means hiding the details from the title to build more suspense. In the inflammatory type, inappropriate or vulgar words are phrased. Formatting means overusing the capitalization/punctuation in the headlines, for instance ALL CAPS or exclamation points are used. In graphic types, the subject matter is disturbing or unbelievable. Bait-and-switch means the news included in the title is not found at the target page. Ambiguous means the title is unclear or confusing, while wrong means using a plainly incorrect article. Kaur et al. [8] also identify eight other types of clickbait headlines: reaction, reasoning, revealing, number, hypothesis/guess, questionable, forward referencing, and shocking/unbelievable. They also found that shocking/unbelievable, hypothesis/guess, and reaction clickbait types to be the most frequently occurring types of clickbait headlines published online.

According to Zheng et al. [9], different ways of attracting users' attention are used by the headlines of different article types, which means the characteristics of clickbait vary between article types. This is different from traditional text-analysis issues. For instance, the headlines of forums or blogs are more colloquial than the headlines of other traditional

news. The main difference between these two types of headlines is the use of functional linguistic characteristics such as wondering, exaggerating, and questioning. In [9], two types of characteristics were used: general clickbait, and the type-related characteristics, while the main characteristics used by Naeem et al. [10] for detection of clickbait were sensationalism, mystery, notions of curiosity, and shock.

In another approach, Potthast et al. [3] used three types of features for clickbait headlines, which are: the teaser message, the linked web page, and meta information. The first type includes basic text statistics and dictionary features, while the second type analyses the web pages linked from a tweet, and the third type includes meta information about the tweet’s sender, medium, and time.

Bazaco, Redondo, and Sánchez-García [11] describe the characteristics of clickbait using six variables under two categories: presentation variables and content variables. The first category includes incomplete information, appealing expressions, repetition and serialisation, and exaggeration, while the second type includes the use of soft news and sensationalist content and striking audiovisual elements. According to [1], the characteristics of curiosity used in clickbait are: its intensity, tendency to disappoint, transience, and association with impulsivity. These lead to a knowledge gap that are exploited by the clickbait headlines to encourage readers to click through to read the whole article.

2.2. Machine Learning and Deep Learning Methods for Clickbait Detection

Several machine learning and deep learning methods have been applied to detect clickbait headlines from different social networks, including Twitter, Facebook, Instagram, Reddit, and others. Table 1 summarizes recent studies on clickbait detection methods. The results in the table show that the performance of machine learning methods still needs to be improved. In the best cases, the highest accuracy obtained reached 0.87 by [12]. In contrast, the use of deep learning showed a good improvement in performance, where the accuracy obtained by [13] reached 0.97. Most of the existing studies used headlines written in English or other languages. Only a few studies focused on clickbait headlines in Arabic. Although Arabic and English scripts have some similarities, there are a number of characteristics that specify the uniqueness of Arabic script. These include: the direction of Arabic, which is written from right to left, and the fact that neither upper nor lower cases exist in Arabic, which is written cursorily. In Arabic, all letters are connected from both sides, except six letters that can be connected from the right side only. Each of the 28 letters of Arabic script has different shapes, depending on its position in the word, and some letters are very similar, differing only in the number and/or the position of dots [14,15]. In addition, there are other special features which are unique to Arabic script such as elongation, morphological characteristics, word meters, and morphemes [16].

Table 1. Summary of recent studies on clickbait detection methods.

Study	Dataset	Classification Method(s)	Accuracy of the Model(s)	Issues/Future Directions
[2]	The dataset includes 1349 clickbait and 2724 non-clickbait websites from different news websites whose pages surfaced on the Yahoo homepage.	Gradient Boosted Decision Trees (GBDT)	0.76	(1) Include the non-textual features (example: images and videos) and the comments of users on articles. (2) Find the most effective types of clickbait that can attract clicks and propose methods to block them. (3) Deep learning is proposed to be applied to obtain more indicators for clickbaits. The obtained performance needs to be improved.

Table 1. Cont.

Study	Dataset	Classification Method(s)	Accuracy of the Model(s)	Issues/Future Directions
[3]	The dataset includes 2992 tweets from Twitter, 767 of which are clickbait.	Logistic regression, naive Bayes, and random forest	0.79	The first evaluation corpus was proposed with baseline detection methods. However, this task needs more investigation to detect clickbait between different social media, and improving the performance of detection. The obtained performance needs to be improved Future works can be: (1) Extract more features; (2) apply other machine learning methods; (3) collect more high-quality data. The obtained performance needs to be improved.
[17]	Clickbait Challenge 2017 Dataset includes over 21,000 headlines.	Random Forest Regression	0.82	Future works: Develop the model as a website or mobile application for Twitter and Instagram. The obtained performance needs to be improved.
[12]	CLDI dataset from Instagram includes 7769 instances and WCC dataset from Twitter includes 19538 instances.	KNN, LR, SVM, GNB, XGB, MLP,	0.87	The maximum length of the headline is limited. If the headlines are long, this might cause information loss. This needs more investigation to solve information-loss problem and including user-behavior analysis. The obtained performance needs to be improved.
[9]	The dataset contains 14,922 headlines, where half of them are clickbait. These headlines are taken from four famous Chinese news websites	Clickbait convolutional neural network (CBCNN)	0.80	The good accuracy was obtained due to the loop back approach that was employed by the LSTM that allows for a better understanding of the context and then better classification of headlines. Future works include (1) Find the most important features needed for learning process.
[10]	Dataset of head-lines from Reddit. The datasets includes 16,000 legitimate news and 16,000 clickbait samples.	LSTM using word2vec word embedding	0.94	(2) Gather more data to develop better models (3) Develop web application that can utilize this model and can alert the user to the clickbait websites.
[6]	The dataset was collected from Reddit, Facebook and Twitter. It includes 814 clickbait samples and 1574 nonclickbait samples.	Convolutional neural network	0.90	
[13]	The dataset includes 32,000 headlines that includes 16,000 clickbait and 16,000 non-clickbait titles.	Recurrent Convolutional Neural Network (RCNN) + Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU)	0.97	A larger dataset can be used.

Table 1. Cont.

Study	Dataset	Classification Method(s)	Accuracy of the Model(s)	Issues/Future Directions
[18]	The three datasets (A, B and C) from Clickbait Challenge 2017 were used. It includes 2495, 80,012 and 19,538 respectively. Clickbait Challenge datasets include 20,000 pairs of training and validation posts.	Self-attentive RNN	0.86	The obtained performance needs to be improved.
[19]	FNC dataset includes 49,972 pairs of training and validation posts.	Deep Semantic Similarity Model (DSSM)	0.86	The other features like image information were not considered in this work. Also, the obtained performance needs to be improved.

To address the lack of study of clickbait detection in Arabic texts, this paper focuses on improving the performance of machine learning methods for detecting clickbait headlines on social networks in the Arabic language.

2.3. Problem Formulation for Clickbait Detection

The clickbait detection problem is a subset of natural language processing that can be represented as a binary classification as follows:

Given a set of shared posts via social networking platforms (tweets) $T = \{t_1, t_2, \dots, t_n\}$, let $t \in T$ denote a post that is classified into a class $\mathbb{C} = \{\mathbb{C}+, \mathbb{C}-\}$ where $\mathbb{C}+$ is a class of the tweets $t_i \in T$ that are considered as legitimate news, and $\mathbb{C}-$ is the class of the clickbait news $t_j \notin \mathbb{C}+$.

To solve the problem, let D be a dataset of all posts $D = \{V1, V2, \mathbb{C}\}$ where $V1 = \{v_1^1, v_2^1, v_3^1, \dots, v_n^1\}$ a vector of extracted features from user portfolio (user-based features (UFs)) and $V2 = \{v_1^2, v_2^2, v_3^2, \dots, v_n^2\}$ is a vector of extracted features from the post/tweet content (content-based features (CFs)). Let also v_i^1 and v_i^2 be the points of a specific feature I and $v_i^1 \in V1$ and $v_i^2 \in V2$.

Let D' be a training set and D'' be a testing set, where D' and $D'' \in D$. Let ξ be a function that generates I from D' and D'' based on the feature space $V : \xi : T \times V \rightarrow I$. As the vector space can be high-dimensional, the clickbait detection problem is now formulated as follows:

Let χ be a function that maps post $t_i \in T$ to $\mathbb{C} = \{\mathbb{C}+, \mathbb{C}-\}$, $\mathbb{C} : \chi : T \rightarrow \mathbb{C}$, where $\mathbb{C} = \langle \mathbb{C}, r \rangle$ and r is a binary relation which takes value 1 if a post $t_i \in T$ is a legitimate post and $t_i \in \mathbb{C}+$, and 0 otherwise.

The function χ can now be set as an optimization problem as follows:

optimize $f_\chi(V1, V2)$ subject to $c(V1, V2)$ where c is a constraint set on the search space.

3. Materials and Methods

The proposed multiple-feature-based approach for detecting clickbait news is presented in this section. Since the difference between clickbait and normal news can be distinguished directly by analysis of the linguistic character of news content [20], the proposed approach takes into consideration both the headlines and the content of the news features (CFs). In addition, to overcome the limitations of such approach, they are combined with news content features.

Figure 1 presents the methodology followed in this study, which consists of the following phases: data collection, data preparation, and machine learning model training and testing. For detecting clickbait news on social networks, both of the investigated news and profile of the user who shared the post are collected. We first constructed a baseline dataset from the raw dataset by labelling the news as clickbait or legitimate. Since the

amount of collected data was huge and for building a sufficiently satisfactory dataset, we used a pseudo labelling learning (PLL) technique [21]. In the next phase, both of the news headlines and contents are pre-processed, including text cleansing, normalization, stemming, stop word removal, and tokenization. These steps are necessary to enhance the overall performance of the ML-based model. We concatenated the processed text with user-based features and then applied the feature reduction using a one-way ANOVA test. The selected features were fed to the ML model. A set of ML models was tested, and their hyper-parameters were tuned to ensure finding the optimal settings. Finally, the ML model was evaluated, and the overall performance reported.

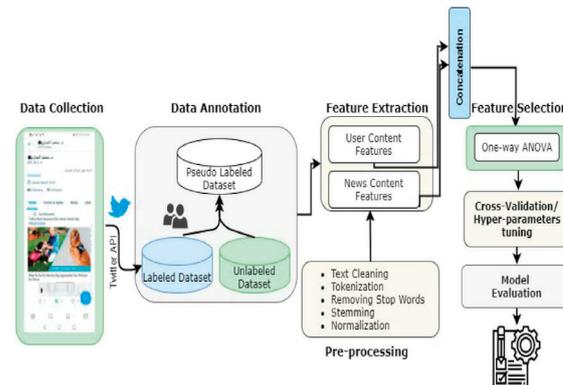


Figure 1. The proposed multiple feature approach for detection of clickbait news.

3.1. Data Collection

We collected 72,321 Arabic news items from Twitter. The dataset can be obtained from github.com (<https://github.com/Moh-Sarem/Clickbait-Headlines#clickbait-headlines>) (accessed on 1 October 2021). For this purpose, we implemented a special crawler that can access breaking news on social networks by feeding the name of the public breaking news agencies. Often, Twitter APIs return tweets in JSON format. However, because many features are not helpful for the proposed model, the used crawler filters out and saves all the collected information from user profile and shared content in comma-separated values (CSV) format. The details of the collection process through multiple feature analysis are shown in Algorithm 1. In addition, the full description of the features used is presented in Tables 2 and 3.

Algorithm 1 Pseudocode of dataset collection process for extracting UFs and CFs

Input: A list of public Twitter breaking news agencies' profiles N
Output: Unlabelled dataset with UFs and CFs
For each profile $p \in N$ **do:**
 Access public page of p
 Retrieve all shared tweets t^p
 Pull out using Twitter APIs tweet's features (USs)
If t^p contains an external URL **Then:**
 Visit the external webpage p^e
For all html tags in p^e **do:**
 Find html tag that contains news full text (CFs) Compute similarity score between t^p and p^e
End
End if
 Store the extracted features in csv format
End

Table 2. User-based features.

# Feature	Feature Name	Description
UF ¹	User ID	Every user has one unique ID.
UF ²	Name	The name of the user who post news on Twitter
UF ³	Screen name	The screen name that this user identifies themselves with.
UF ⁴	Date of join	The creation time of the user account
UF ⁵	#Url	Number of URL provided by the user in association with their profile
UF ⁶	Profile description	A text that shows how the user describes his/her account
UF ⁷	Verified	A boolean indicator shows whether the user has a verified account or not
UF ⁸	Count of followers	Total number of followers
UF ⁹	Count of friends	Numeric value indicates how many friends that the user has
UF ¹⁰	Count of favorites' accounts	Numeric value indicates how many tweets this user has liked
UF ¹¹	Count of public lists	Total number of public lists that this user is a member of.
UF ¹²	Location	The geographical location
UF ¹³	Hashtage	The associated hashtag with the post
UF ¹⁴	Lang	The post language
UF ¹⁵	Number of post shared	Total number of content shared by the user

Table 3. Content-based features.

# Feature	Feature Name	Description
CF ¹	#Url	Number of external URLs provided by the user in association with news
CF ²	Source	Name of the source of the news article.
CF ³	Headline	The headline of the news article for catching the reader's attention
CF ⁴	Tweet Text	The body of the tweet news
CF ⁵	Body Text	The full news in readable format, often with external content
CF ⁶	Retweet count	Total number of times this tweet has been retweeted.
CF ⁷	media	Boolean value indicates whether there are associated images or videos
CF ⁸	Similarity score	The score for similarity between headline text and body text.
CF ⁹	Creation date	The posted date of the news content

3.2. Data Annotation

Once we obtained the final dataset by using the implemented crawler, we prepared a baseline dataset from the retrieved dataset. Every shared tweet was labelled as a clickbait or legitimate by asking three media professionals to volunteer in judging 12,321 tweets and their associated news. They were asked to access the external webpage by following the URL link provided with tweet and comparing the tweet's body and headline with the full text in the destination webpage. To facilitate this job, we provided them with examples showing what clickbait news looks like. Table 4 shows a guideline for how to classify the content of the shared tweets.

Table 4. Example of clickbait news.

Type	Definition	Example of Arabic Clickbait News	Translation of the Arabic Clickbait News
Ambiguous	Title unclear or confusing to spur curiosity.	هذا الأمر لم يحدث في المملكة؟	This matter did not happen in the kingdom.
Exaggeration	The title exaggerates the content of the landing page.	الرجل ده كريم أوى يا بابا.. زيون يأكل بـ20 دولار ويترك 1400 بقشيش بأمريكا	This man is kind father. In America, the customer eats for \$20 and leaves 1400 tips
Inflamm-atory of inappropriate/vulgar words.	Either phrasing or use of inappropriate/vulgar words.	اطباء تحت مسمى الطب ” الاطباء ” المجرمين	Doctors under the name of medicine “criminal doctors”
Teasing	Omission of details from title to build suspense: teasing.	بين ليلة وضحاها... أمريكي يربح مليار دولار	Overnight... an American wins a billion dollars
Formatt-ing	Excessive use of punctuation or exclamation marks.	كيف أنت عمياء ومصورة؟! هذه “ العبارة السلبية كانت انطلاقة ”المطيري	“How are you blind and a photographer”?!.. This negative phrase was the launch of “Al-Mutairi”
Wrong	Just plain incorrect article: factually wrong. The thing promised/im-	أمور يقوم بها الأغنياء ولا تقوم بها 10 إنفسك	10 things rich people do that you don’t do yourself!
URL redirection	lied from the title is not on the landing page: it requires additional clicks or just missing.	كندا: ينمو الناتج المحلي الإجمالي الحقيقي بنسبة 0.7% في نوفمبر مقابل 0.4% المتوقعة	Canada: Real GDP grows 0.7% in November vs. 0.4% expected
Incomplete	The title is incomplete	عاجل: تطور في أرامكو و مدينة - ...صناعية	Urgent: An improvement in Aramco and an industrial city

As shown in Table 4, there are seven categories that the volunteers could use to label each post as clickbait news. In case of unclearness or doubt about which class the post belongs to, the post is labelled as “incomplete”. Every content text in the baseline dataset has three labels, one provided by each annotator. To assign the final class label, we applied the majority voting algorithm and labelled the content as clickbait or legitimate news. Table 5 shows the details of the baseline dataset, which includes 4325 items of clickbait news and 6743 legitimate items. The news items that are labelled as incomplete were later removed from the dataset. The remaining baseline dataset contained 11,068.

Table 5. Details of baseline dataset.

Parameter	# of Treated Data
Total news in dataset	12,321
Remaining baseline dataset	11,068
% of treated news in respect to the whole dataset	17%
Clickbait news items, %	4325, 35.1%
Legitimate news items, %	6743, 54.72%
Incomplete posts, %	1253, 10.16%
Number of external URLs	4862
Number of breaking news sources	7

As the size of our final baseline dataset was quite small (17% of the original dataset), we decided to apply a pseudo-labelling learning technique to enhance the performance of the ML model. PLL is an efficient semi-supervised technique that can be applied to utilize unlabeled data while training ML models. As shown in Figure 1, the ML model is trained first on the labeled data (in this case: the baseline dataset). The model then predicts the labels of unlabeled data. The predicted pseudo-labels are assigned as target classes for unlabeled data and combined with the original baseline dataset (labeled data). Finally, the produced new dataset is then used to train the proposed ML-models. After applying PLL

technique, the size of the labeled dataset was increased to around 54893 instances. Table 6 shows the details of the final dataset after applying the PLL technique on 71.54% of the remaining unlabeled data.

Table 6. Final dataset after applying PLL technique.

Parameter	# of Treated Data
Total no. of news items in the dataset,	54,893
% of treated news items with respect to the whole dataset	75.90%
No. of clickbait items, % of the total news items	23,981, 43.69%
No. of legitimate news items, % of the total items	30,912, 56.31%
Number of external URLs	14,518
Number of breaking news sources	22

3.3. Pre-Processing and Numeric Representation

Beside the UFs and CFs described above in Tables 2 and 3, the “headline” CF^3 , “tweet text” CF^4 , and “body text” CF^5 features from CFs required additional treatment.

3.3.1. Pre-Processing

For many text classification systems, pre-processing is considered as an essential step to improve the quality of data as well as the efficiency and accuracy of ML models [22,23]. The common pre-processing steps include text cleansing, tokenization, removing stop words, stemming, and normalization. Since the obtained data is pulled out from Twitter and by accessing the external web pages following the URL links associated with the body of the tweets, additional pre-processing steps were performed, such as deletion of unnecessary, insignificant items from texts (e.g., digits, punctuation marks, URLs, special characters, non-Arabic characters, diacritics), and removal of emojis and hashtags.

3.3.2. Numeric Representation

By numeric representation, we mean converting the textual content into a form that could be fed into the ML model in treatable format. In this work, the term frequency-inverse document frequency (TF-IDF) is used as a numeric representation. Mathematically, the TF-IDF can be calculated as in Equations (1)–(3):

$$tf_idf_{t,D} = TF_{t,D} \times IDF_t \tag{1}$$

where

$$TF_{t,D} = \frac{\text{Number Of Repetitions of Term } t \text{ In a Document } D}{\# \text{ Of terms In a Document}} \tag{2}$$

and

$$IDF_t = \log \frac{\text{Number Of Documents}}{\text{Number Of Documents Containing The term } t} \tag{3}$$

After applying the TF-IDF technique on the final dataset, the training time of ML models was long because of high dimensionality, where the number of extracted features reached 10,230.

3.4. Feature Selection

Feature selection (FS) is an effective way to reduce large data [23]. The main purpose of FS is to delete irrelevant and noisy data. It also enables a representative subset of all data to be chosen to minimize the complexity of the classification process. Several FS techniques can be found in the literature. These include: Mutual Information (MI), Information Gain (IG), improved Chi-square, and the one-way ANOVA F-test [24] (referred to, hereafter as FV-ANOVA). This paper proposes to use FV-ANOVA as a feature selection method that is used to statistically select the important features according to the F-values. The features are sorted in ascending order, where the most relevant features appear on the top. Finding the

best cut-point value is a challenge. Thus, we divided the features into a set of groups based on a given percentile ($p\%$) of the original number of features. This step allows us to find the top-scoring features. Later, only the $p\%$ top-scoring features were used to train the ML classifiers. The process of selecting features for FV-ANOVA is presented in Algorithm 2.

Algorithm 2 Pseudocode for selecting features-based FV-ANOVA method.

```

Input:  $D$ -dataset,  $V$  features extracted as numeric representation by TF-IDF,  $C$ -class label and  $p\%$  percentile.
Output:  $D_{FS}$  subset of top-scoring features based on the given  $p\%$ 
 $k \leftarrow$  number of classes in  $D$ 
 $N \leftarrow$  number of features in  $D$ 
For each pair  $f_j \in (V, C)$  do:
  Count number of samples per class
  Compute (mean, standard deviation, standard error) of each  $f_j$  with respect to  $C_i$ 
  Compute degree of freedom between/within classes ( $SS_B, SS_w$ )
  Compute sum of square of ( $SS_B, SS_w$ )
  Find mean square  $MS_B$  between groups as  $MS_B = SS_B / (k - 1)$ 
  Find mean square  $MS_W$  between groups as  $MS_W = SS_W / (N - k)$ 
End for
 $F\_value \leftarrow \frac{MS_B}{MS_W}$ 
Sort  $F\_value$  in ascending order
 $D_{FS} \leftarrow$  Select the top-scoring features based on  $p\%$ 
Return  $D_{FS}$ 

```

3.5. Feature Selection

Six ML classifiers were implemented: Random Forest (RF), Logistic Regression with Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Logistic Regression (LR), Multinomial Naïve Bayes (NB), and k-Nearest Neighbor (k-NN). To explore the effectiveness of the proposed feature selection method, we conducted different experiments and employed these classifiers on different subsets of features based on F-values.

For tuning hyper-parameters of the used ML classifiers, the grid search algorithm with k-fold cross-validation is used. Subsequently, the values of hyper-parameters that yield the highest performance measure are set to be the final values of hyper-parameters for each classifier. The set of values of hyper-parameters used in this work is presented in Table 7.

Table 7. List of optimized hyper-parameters of each classifier.

ML Classifier	Hyper-Parameters Used for Tuning the Model	Best Values of Hyper-Parameters
RF	Criterion = [entropy, gini] max_depth = [10–1200] + [None] min_samples_leaf = [3–13] min_samples_split = [5–10] n_estimators = [150–1200]	Criterion = gini max_depth = 142 min_samples_leaf = 3 min_samples_split = 7 n_estimators: 300
SGD	alpha = [1×10^{-5} , 1×10^{-4} , 1×10^{-3} , 1×10^{-2} , 0.1, 1] loss = [log, hinge] max_iter= [10–1000] Penalty= [l2', 'l1', 'elasticnet']	alpha= 1×10^{-4} loss = log max_iter = 1000 Penalty = l2
SVM	C = [0.1, 1, 10, 1×10^2 , 1×10^3] Gamma = [1×10^{-4} , 1×10^{-3} , 1×10^{-2} , 0.1, 1, 10, 1×10^2] Kernel = [sigmoid, linear, rbf]	C = 10 Gamma = 1×10^{-3} Kernel = rbf
LR	C = [1×10^{-3} , 1×10^{-2} , 0.1, 1, 10, 100], fit_intercept = [True, False]	C = 1×10^{-3} fit_intercept = True
NB	alpha = [1×10^{-5} , 1×10^{-4} , 1×10^{-3} , 1×10^{-2} , 0.1, 1]fit_prior = [True, False]	alpha = 0.1 fit_prior = True
K-NN	n_neighbors = [1, 40]	Number of neighbours = 7

3.6. Model Evaluation

To evaluate the performance of classifiers, we computed the accuracy (Acc), recall (R), precision (P), and f1-score ($F1$) metric of each classifier with those features that were selected by the proposed F-values of the one-way ANOVA test. The descriptions of these metrics are shown in Equations (4)–(7) respectively.

$$Acc. = \frac{TP + TN}{D} \quad (4)$$

where $(TP + TN)$ is the accurately predicted content either clickbait or not, D is the total number of samples in the dataset.

$$P = \frac{TP}{TP + FP} \quad (5)$$

where $(TP + FP)$ is the total number of predicted clickbait content.

$$R = \frac{TP}{TP + FN} \quad (6)$$

where $(TP + FN)$ is the total number of actual clickbait content.

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (7)$$

4. Experimental Design

The experiments in this study were performed on Python 3.8 with Windows 10 operating system. We used numerous Python packages including sklearn 0.22.2 for implementing the classifiers, nltk 3.6.2 for pre-processing Arabic text and Beautiful soup 4.9.0 for scraping data from external web pages. The user-based features and content-based features were fed into classifiers separately. Later, we merged both types and measured the performance of ML classifiers based on top-scoring features $p\%$ that were selected based on f-values of one-way ANOVA. For ensuring fair comparison between classifiers, the same pre-processing steps and the same set of features were used for each classifier. In addition, we considered four experimental scenarios per feature type, as illustrated in Table 8.

Table 8. Number of features per each experiment.

#	Type of Experiment	Number of Features
UCFs	Baseline:	15
	F_values_5%: 5% of features	4
	F_values_10%: 10% of features	7
	F_values_15%: 15% of features	9
NCFs	Baseline: Including (TF-IDF) extracted features	10,236
	F_values_5%: 5% of features	732
	F_values_10%: 10% of features	2187
	F_values_15%: 15% of features	5867
UCFs + NCFs	Baseline:	10,251
	F_values_5%: 5% of the extracted features	736
	F_values_10%: 10% of the extracted features	2194
	F_values_15%: 15% of the extracted features	5876

5. Results and Findings

This section describes and discusses the results for each experiment shown in Table 8. First, we present the findings that were obtained when only the user-based features were used. The accuracy of each classifier is presented in Table 9. The second type of features,

content-based features, were then investigated, as shown in Table 10. Finally, we combined both types of features and the performance of classifiers is presented in Table 11.

Table 9. Accuracy of different experiments with user-based features only.

ML Classifier	Experiment			
	Baseline	F_Values_5%	F_Values_10%	F_Values_15%
RF	61.24	61.73	63.93	62.34
SGD	59.04	52.65	51.86	57.62
SVM	64.40	62.76	66.54	61.08
LR	61.87	61.87	61.87	61.87
NB	61.75	62.03	60.12	61.49
k-NN	60.57	46.83	42.72	41.33

Table 10. Accuracy of different experiments with content-based features only.

ML Classifier	Experiment			
	Baseline	F_Values_5%	F_Values_10%	F_Values_15%
RF	77.97	86.83	90.21	83.67
SGD	74.72	85.72	84.59	79.19
SVM	75.89	89.31	91.83	90.37
LR	77.65	75.43	75.76	75.09
NB	74.65	87.35	90.24	89.46
k-NN	76.99	73.77	65.08	65.08

Table 11. Accuracy of different experiments with combination of UFs and CFs.

ML Classifier	Experiment			
	Baseline	F_Values_5%	F_Values_10%	F_Values_15%
RF	77.18	86.94	88.13	85.02
SGD	74.83	82.52	87.02	85.39
SVM	75.00	88.92	92.16	90.65
LR	76.77	76.73	75.00	75.87
NB	75.30	89.27	90.74	89.62
k-NN	77.05	74.41	71.02	71.61

Based on the results presented in Table 9–11, the following findings are observed and can be summarized as follows:

- When the content-based features were used, the classifiers performed well and SVM, NB, and RF achieved notable results using 10% of top-scoring features compared to their results in the baseline experiment. Among these methods, SVM obtained the best accuracy (91.83%) for content-based features.
- In most cases of experiments with content-based features, all classifiers showed good results when the one-way ANOVA method was used as feature selection, except k-NN and LR. It is notable that k-NN had the worst performance when the number of selected features increased to 10% and 15%.
- Increasing the percentage of the top-scoring features to more than 10% leads to a reduction in the performance of the ML classifiers.
- RF and SVM benefited more when the user-based features were used, compared to their results in the baseline experiment.
- The result for LR remained constant, and no change was observed when user-based features were fed into the classifier.
- The k-NN and SGD do not benefit from the ANOVA-based feature selection at all for user-based features.

- Combining user-based features and content-based features shows an improvement in the performance of ML classifiers and only LR and k-NN classifiers did not show any improvement.
- The SVM outperforms all other classifiers and benefited more when the proposed feature selection method was used for the combination of user-based features and content-based features. The highest accuracy achieved was 92.16%.
- As the total number of features for the combination of user-based and content-based features is 10,251, selecting the top 10% of these features (2194) was more suitable for SVM, which performed well with low dimensionality data.
- As shown in the results, using the user-based features achieved lower performance than using the content-based features for all ML methods. Therefore, the proposed model relies more on the content-based features and the combined ones.

6. Conclusions

This paper has proposed a comprehensive approach that includes three main phases: data collection, data preparation, and machine learning modeling phases. After collecting the dataset, which is considered the first Arabic clickbait headline news dataset, the pre-processing tasks were performed, which included text cleansing, normalization, stemming, stop words removal, and tokenization. The features of the processed text (content-based features) were then combined with the user-based features and the feature selection was then applied using one-way ANOVA test. Finally, the ML models were applied, which included Random Forest (RF), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Logistic Regression (LR), Multinomial Naïve Bayes (NB), and K-nearest Neighbor (k-NN). Hyper-parameter tuning methods were applied to ensure finding the optimal settings. The experimental results showed a great enhancement when the CFs were used and also when a combination of UFs and CFs was used. The accuracy achieved reached 92.12% using 10% of the top-scoring features, which is better than that reported in many previous studies (discussed in the related works). This enhancement is particularly interesting, as we are dealing with Arabic contents. Future work will investigate the application of several deep learning methods on this Arabic dataset in order to enhance the detection performance. Moreover, collecting more Arabic content to add to the dataset will be a beneficial addition to conducting the analysis.

Author Contributions: Conceptualization, M.A.-S., F.S., T.A.-H.; methodology, M.A.-S., F.S.; software, M.A.-S.; validation, T.A.-H., A.A.; formal analysis, F.S., M.H.; investigation, M.T.A., T.S.A.; resources, Z.G.A.-M., B.A.M., M.H., A.A., T.S.A.; data curation, B.A.M.; writing—original draft preparation, M.A.-S., F.S.; writing—review and editing, M.A.-S., F.S.; visualization, F.S.; supervision, M.A.-S., F.S.; project administration, M.A.-S., Z.G.A.-M.; funding acquisition, Z.G.A.-M., B.A.M., M.T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by the Scientific Research Deanship at the University of Ha'il, Saudi Arabia, through project number RG-20 023.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset can be obtained from <https://github.com/Moh-Sarem/Clickbait-Headlines#clickbait-headlines> (accessed on 1 October 2021).

Acknowledgments: We would like to acknowledge the Scientific Research Deanship at the University of Ha'il, Saudi Arabia, for funding this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chen, Y.; Conroy, N.J.; Rubin, V.L. Misleading Online Content: Recognizing Clickbait as “False News”. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, Washington, DC, USA, 13 November 2015; pp. 15–19.
- Biyani, P.; Tsioutsoulouklis, K.; Blackmer, J. “8 Amazing Secrets for Getting More Clicks”: Detecting Clickbaits in News Streams Using Article Informality. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
- Potthast, M.; Kopsel, S.; Stein, B.; Hagen, M. Clickbait Detection. In *Advances in Information Retrieval. 38th European Conference on IR Research (ECIR 16)*; Springer: Cham, Switzerland, 2016; pp. 810–817.
- Chakraborty, A.; Paranjape, B.; Kakarla, S.; Ganguly, N. Stop clickbait: Detecting and preventing clickbaits in online news media. In Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), Davis, CA, USA, 18–21 August 2016; pp. 9–16.
- Khater, S.R.; Al-sahlee, O.H.; Daoud, D.M.; El-Seoud, M. Clickbait detection. In Proceedings of the 7th International Conference on Software and Information Engineering, Cairo, Egypt, 4–6 May 2018; pp. 111–115.
- López-Sánchez, D.; Herrero, J.R.; Arrieta, A.G.; Corchado, J.M. Hybridizing metric learning and case-based reasoning for adaptable clickbait detection. *Appl. Intell.* **2017**, *48*, 2967–2982. [[CrossRef](#)]
- Agrawal, A. Clickbait Detection Using Deep Learning. In Proceedings of the 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Dehradun, India, 14–16 October 2016; pp. 268–272.
- Kaur, S.; Kumar, P.; Kumaraguru, P. Detecting clickbaits using two-phase hybrid CNN-LSTM biterm model. *Expert Syst. Appl.* **2020**, *151*, 113350. [[CrossRef](#)]
- Zheng, H.T.; Chen, J.Y.; Yao, X.; Sangaiah, A.K.; Jiang, Y.; Zhao, C.Z. Clickbait convolutional neural network. *Symmetry* **2018**, *10*, 138. [[CrossRef](#)]
- Naeem, B.; Khan, A.; Beg, M.O.; Mujtaba, H. A deep learning framework for clickbait detection on social area network using natural language cues. *J. Comput. Soc. Sci.* **2020**, *3*, 231–243. [[CrossRef](#)]
- Bazaco, A.; Redondo, M.; Sánchez-García, P. Clickbait as a strategy of viral journalism: Conceptualisation and methodst. *Rev. Lat. de Comun. Soc.* **2019**, *74*, 94–115.
- Jain, M.; Mowar, P.; Goel, R.; Vishwakarma, D.K. Clickbait in Social Media: Detection and Analysis of the Bait. In Proceedings of the 2021 55th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 24–26 March 2021; pp. 1–6.
- Chawda, S.; Patil, A.; Singh, A.; Save, A. A Novel Approach for Clickbait Detection. In Proceedings of the 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 23–25 April 2019; pp. 1318–1321.
- Aburas, A.A.; Gumah, M.E. Arabic Handwriting Recognition: Challenges and Solutions. In Proceedings of the 2008 International Symposium on Information Technology, Kuala Lumpur, Malaysia, 26–28 August 2008.
- Al-Nuzaili, Q.A.; Hashim, S.Z.M.; Saeed, F.; Khalil, M.S.; Mohamad, D.B. Pixel distribution-based features for offline Arabic handwritten word recognition. *Int. J. Comput. Vis. Robot.* **2017**, *7*, 99–122. [[CrossRef](#)]
- Al-Sarem, M.; Cherif, W.; Wahab, A.A.; Emara, A.H.; Kissi, M. Combination of stylo-based features and frequency-based features for identifying the author of short Arabic text. In Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications, Rabat, Morocco, 24–25 October 2018; pp. 1–6.
- Cao, X.; Le, T. Machine learning based detection of clickbait posts in social media. *arXiv* **2017**, arXiv:1710.01977.
- Zhou, Y. Clickbait detection in tweets using self-attentive network. *arXiv* **2017**, arXiv:1710.05364.
- Dong, M.; Yao, L.; Wang, X.; Benatallah, B.; Huang, C. Similarity-aware deep attentive model for clickbait detection. In *Advances in Knowledge Discovery and Data Mining*; Springer: Cham, Switzerland, 2009; Volume 11440, pp. 56–69.
- Sahoo, S.R.; Gupta, B.B. Multiple features based approach for automatic fake news detection on social networks using deep learning. *Appl. Soft Comput.* **2021**, *100*, 106983. [[CrossRef](#)]
- Lee, D.-H. Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. In *Workshop on Challenges in Representation Learning*; ICML: Atlanta, GA, USA, 2013; Volume 3, p. 2.
- Al-Sarem, M.; Saeed, F.; Alsaeedi, A.; Boulila, W.; Al-Hadhrani, T. Ensemble methods for instance-based Arabic language authorship attribution. *IEEE Access* **2020**, *8*, 17331–17345. [[CrossRef](#)]
- Bahassine, S.; Madani, A.; Al-Sarem, M.; Kissi, M. Feature selection using an improved Chi-square for Arabic text classification. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *32*, 225–231. [[CrossRef](#)]
- Elssied, N.O.F.; Ibrahim, O.; Osman, A.H. A novel feature selection based on one-way ANOVA F-test for e-mail spam classification. *Res. J. Appl. Sci. Eng. Technol.* **2014**, *7*, 625–638. [[CrossRef](#)]

Review

An Investigation of the Policies and Crucial Sectors of Smart Cities Based on IoT Application

Armin Razmjoo ^{1,*}, Amirhossein Gandomi ^{2,*}, Maral Mahlooji ³, Davide Astiaso Garcia ⁴, Seyedali Mirjalili ^{5,6}, Alireza Rezvani ⁷, Sahar Ahmadzadeh ⁸ and Saim Memon ^{9,10}

- ¹ Escola Técnica Superior d'Enginyeria Industrial de Barcelona (ETSEIB), Universitat Politècnica de Catalunya (UPC), Av. Diagonal, 647, 08028 Barcelona, Spain
 - ² Faculty of Engineering & Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia
 - ³ Center for Environmental Policy, Imperial College London, London SE1 0AA, UK; maral.mahlooji@gmail.com
 - ⁴ Department of Planning, Design, and Technology of Architecture, Sapienza University of Rome, Via Gramsci 53, 00197 Rome, Italy; davide.astiasogarcia@uniroma1.it
 - ⁵ Centre for Artificial Intelligence Research and Optimisation, Torrens University Australia, Brisbane, QLD 4006, Australia; ali.mirjalili@gmail.com
 - ⁶ Yonsei Frontier Lab, Yonsei University, Seoul 03722, Korea
 - ⁷ Department of Electrical Engineering, Saveh Branch, Islamic Azad University, Saveh 3919715179, Iran; alireza.rezvani.saveh@gmail.com
 - ⁸ Research Fellow in IoT and Big Data, Business and Management Research Institute, University of Bedfordshire, Luton LU1 3JU, UK; sahar.ahmadzadeh2017@gmail.com
 - ⁹ Centre for Efficiency and Performance Engineering (CEPE), Department of Engineering & Technology, School of Computing and Engineering, University of Huddersfield, Queensgate, Huddersfield, West Yorkshire HD1 3DH, UK; s.memon@hud.ac.uk
 - ¹⁰ Solar Thermal Vacuum Engineering Research Group, London Centre for Energy Engineering, School of Engineering, London South Bank University, 103 Borough Road, London SE1 0AA, UK
- * Correspondence: arminupc1983@gmail.com (A.R.); gandomi@uts.edu.au (A.G.)

Citation: Razmjoo, A.; Gandomi, A.; Mahlooji, M.; Astiaso Garcia, D.; Mirjalili, S.; Rezvani, A.; Ahmadzadeh, S.; Memon, S. An Investigation of the Policies and Crucial Sectors of Smart Cities Based on IoT Application. *Appl. Sci.* **2022**, *12*, 2672. <https://doi.org/10.3390/app12052672>

Academic Editors: Dimitrios S. Paraforos, Anselme Muzirafuti, Giovanni Randazzo and Stefania Lanza

Received: 13 January 2022

Accepted: 1 March 2022

Published: 4 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: As smart cities (SCs) emerge, the Internet of Things (IoT) is able to simplify more sophisticated and ubiquitous applications employed within these cities. In this regard, we investigate seven predominant sectors including the environment, public transport, utilities, street lighting, waste management, public safety, and smart parking that have a great effect on SC development. Our findings show that for the environment sector, cleaner air and water systems connected to IoT-driven sensors are used to detect the amount of CO₂, sulfur oxides, and nitrogen to monitor air quality and to detect water leakage and pH levels. For public transport, IoT systems help traffic management and prevent train delays, for the utilities sector IoT systems are used for reducing overall bills and related costs as well as electricity consumption management. For the street-lighting sector, IoT systems are used for better control of streetlamps and saving energy associated with urban street lighting. For waste management, IoT systems for waste collection and gathering of data regarding the level of waste in the container are effective. In addition, for public safety these systems are important in order to prevent vehicle theft and smartphone loss and to enhance public safety. Finally, IoT systems are effective in reducing congestion in cities and helping drivers to find vacant parking spots using intelligent smart parking.

Keywords: smart cities; Internet of Things (IoT); strategy; monitoring

1. Introduction

While old, crowded cities are under pressure from many issues such as population explosion and improper infrastructure, the rise of smart cities (SCs) has provided a good solution for solving many of the existing problems and overcoming urban challenges [1]. Therefore, the rapid development of SCs in recent years reflects their importance [2]. In fact, SCs with greater opportunities for citizens and proper services are becoming an

attractive choice for people and communities, and they can be a place for fostering success in health and businesses across the world with the help of smart infrastructure [3]. On the other hand, achieving SC development requires detailed and specific planning and the proper implementation and establishment of policies. Thus, identifying obstacles can grant us a deeper understanding of how to determine the best solutions with less difficulty [4]. Since, in developing SCs, we are faced with numerous barriers and problems in various areas, such as roadways, environment, utilities, parking, public safety, waste management, and public transport, it is pertinent to enhance these sectors through accurate investigation and practical actions [5]. To overcome such barriers, SC governments must implement appropriate strategies and present proper solutions to mitigate or eliminate these barriers [6]. In this regard, the role of the Internet of Things (IoT) and the utilization of these systems is an essential and beneficial strategy to appropriately develop SCs [7]. In fact, with the appearance of new technologies such as the IoT, the concept of SCs has changed and continues to evolve for the better, subsequently improving and accelerating urban management across various sectors [8]. This means that the utilization of the IoT leads to the development of smart cities [9]. In recent years, many types of research have been conducted on SC development [10]. For instance, the importance of the IoT for SC development was also reviewed by Badis Hammi et al., who demonstrated that a higher level of interaction between SCs and IoT development is essential, as it can integrate electronic devices. However, the safety risks and privacy issues of participating individuals, companies, and organizations should be considered carefully in such cases [11]. Ejaz et al., investigated efficient energy management for the IoT in the context of SCs and observed that such management is a key paradigm to monitoring complex energy systems. They also showed that efficient energy management can support wireless energy transfer for IoT devices and energy-efficient planning in smart homes [12]. Tanveer et al. investigated the growth of the IoT markets across energy systems of SCs. Regarding the importance of smart grid technology innovations in supporting smart energy systems in SCs, the study showed that investment in these systems has increased in recent years. Based on the literature, the IoT of the global energy market exceeded USD 6.8 billion in 2015 and is anticipated to reach USD 26.5 billion by 2023, which portrays an annual growth rate of 15.5% between 2016 and 2023 [13]. Bresciani et al., investigated the IoT in terms of organizations, in order to innovate and implement it in everyday business activities. The results from the 43 IoT SC project alliances across Italian cities they investigated demonstrated that multinational enterprises are building alliances for exploring new technologies for cities as well as exploiting new IoT-based devices to gain economic profit. The study proved that for companies to achieve the desired results, they must integrate different types of knowledge to ensure efficient management and effective support [14]. Evertzen et al., analyzed the effects of smart governance on the quality of life in SCs in the three well-known cities of Palo Alto, Nice, and Stockholm. This research emphasized the importance of innovative approaches across SCs, which should be implemented based on the IoT, and consequently, many services should be promptly digitalized. Therefore, in order to achieve these goals and successfully implement an SC model, strong leadership, citizen involvement, and business collaboration are required [15]. With regard to the importance of transportation systems in SCs, the prospect of handling considerable information using sensor data from the environment for better monitoring of transport systems in SCs was examined by AlZubi et al., as the time data extracted from sensors is important; the researchers presented a responder-dependent add-on information fusion scheme concerning sensor data. This guided vehicle scheme can observe the responding sensor information in order to determine the success of the goal endorsed. This scheme, which is based on classification machine learning, can help us identify and subsequently reduce the errors caused by sensor information [16]. In light of the importance of the IoT in the development of smart cities, this article examines the problems and solutions of seven key sectors that have a significant impact on SC development, including the environment, public transportation, utilities, street lighting, waste management, public safety, and smart parking. We also considered certain important

cities in the EU (Paris, London, Copenhagen, Barcelona, Amsterdam, and Oslo) and in the United States (Boston, New York, and San Francisco) based on the relevance of the IoT.

2. Motivation and Objective of the Critical Review

Creating and developing SCs is an important objective for many countries [17] to enhance the life quality of their population through the optimal management of their resources [18]. In addition, SC development supports global mitigation strategies, especially across the environmental and energy sectors [19]. One of the most important factors for SC development is the IoT [20], which integrates different systems related to energy, transport, and waste and water management within SCs in order to enhance the inhabitants' quality of life [21]. Given that more of the global population resides in urban areas, therefore it can be said that cities are held accountable for the majority of the global energy consumption and greenhouse gas (GHG) emissions [22]. Thus, a reduction in energy use and the maximization of renewable energy use, when available, can support these objectives. The use of the IoT in SCs provides an opportunity to make incremental changes in efficiency by harnessing new technologies and automating processes in applications [23]. It is important to recognize that the innovation, advancement, and implementation of the IoT across SCs have a dynamic impact on many other intertwined systems, including the environment, economy, and transportation. Therefore, it is crucial to create an in-depth understanding of these independencies to ensure that negative impacts are not overlooked and positive impacts are enhanced and used to create an incentive to create changes across cities. The aims of this study include investigating the concept of SCs, identifying the IoT barriers across seven important sectors, and compiling appropriate solutions to tackle each barrier.

3. Methodology

To identify the potential barriers to IoT development in SC development and, based on the importance of the IoT, we conducted an exhaustive review of more than 400 relevant publications related to the IoT, and have searched in Internet the using established scientific databases, such as Google Scholar, Scopus, Web of Science and Journal sites (Taylor & Francis, Elsevier, MDPI, Springer, Willey, etc.).

In this regard, we searched, in the Internet, words such as smart cities, environment, road traffic, public transport, utilities, smart lighting, public safety, waste management, street lighting, and smart parking. In the first step, between 2019 and 2020, we investigated more than 200 review papers to understand the concepts of the IoT and smart cities. Then, we investigated, in 2021, more than 200 technical papers, and eventually selected 121 papers. After these steps, we categorized the most important papers which helped us to start writing this paper and we selected the methodology. Review articles helped us understand SC development and the IoT technologies that have come under the spotlight within a short period of time. Moreover, technical articles established a deeper understanding of effective policies in SC development relative to the IoT in order to obtain proper solutions to the barriers. Figure 1 shows the flowchart for the methodology of this study. After all the relative papers were collected, the articles were categorized into two groups—review papers and technical papers. We based the methodology on the best of these. In the last step, we determined recommended actions and policies to achieve the goal of the paper.

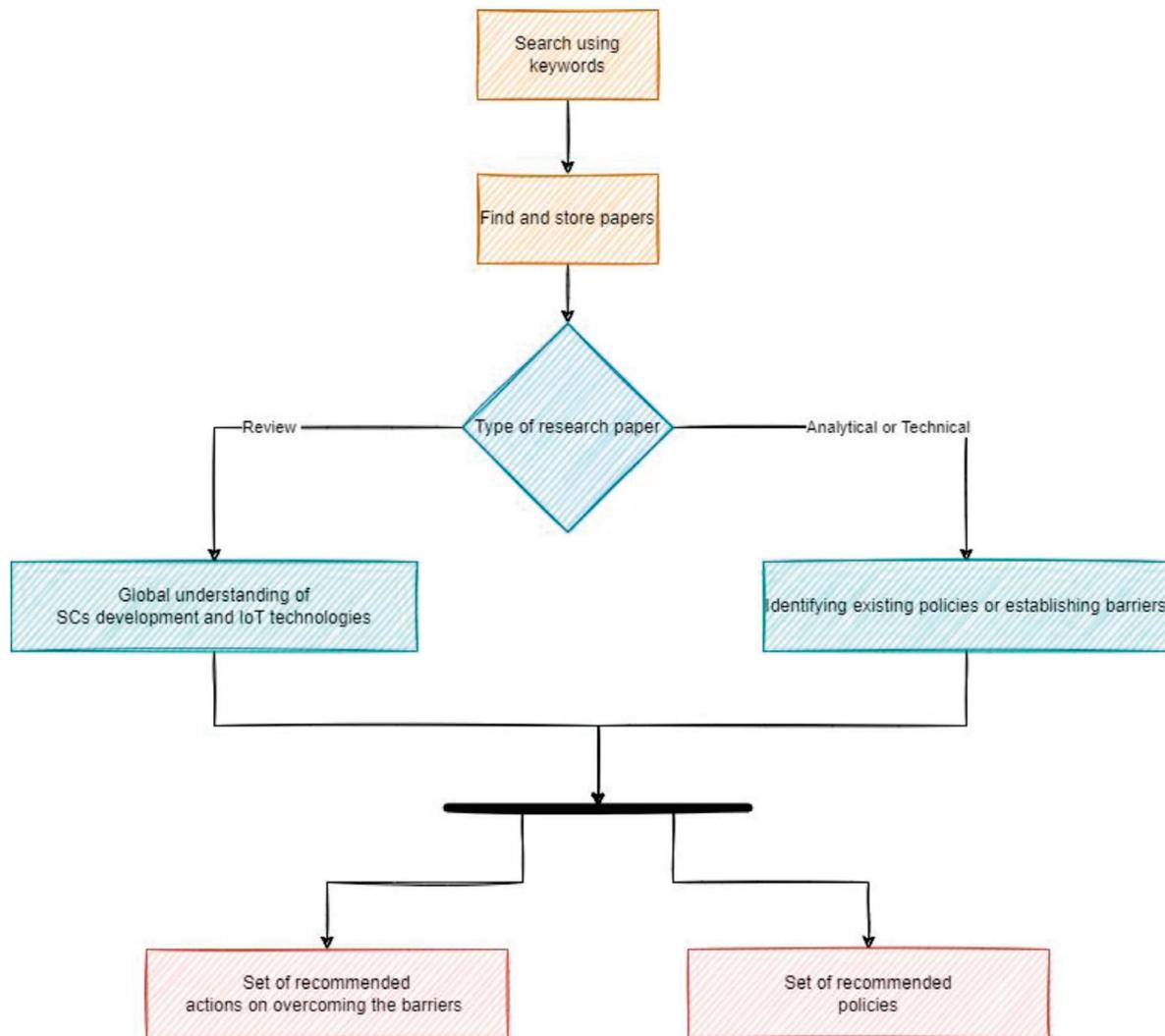


Figure 1. The methodology flowchart.

4. Results and Discussion

4.1. Recognizing the Existing Obstacles in the Development of SCs

As we are faced with various barriers and problems across seven specific sectors in SCs, i.e., environment, public transport, utilities, street lighting, waste management, public safety, and parking, we believe that the utilization and implementation of the IoT will be effective in mitigating or resolving the problems associated with these areas. In the following sections, we comprehensively discuss these problems and the solutions that we obtained from the review articles and scientific research.

4.1.1. Environment

Cleaner air and water systems are crucial elements of the environment [24]; for this, a network of sensors should be used to monitor air [25] and water quality [26]. Specifically, sensors can be used to detect the amount of CO₂, sulfur oxides, and nitrogen to monitor air quality and to detect water leakage, pH levels, and changes in the chemical composition of water. Therefore, sensors can be implanted along busy roads, around plants, and near houses, offices, and organizations [27]. Moreover, it is necessary to utilize sensors for detection and monitoring and to obtain data and results [28]. According to the McKinsey Global Institute, emissions can be reduced by 10–15% through applications that focus on building automation, mobility, and dynamic electricity pricing. Thus, SCs can support and

contribute to a cleaner and more sustainable environment [29]. Nowadays, sensors, as well as environmental sensors, have significantly affected lives, as individual environmental sensors obtain data about the environment and then transform that data into electrical signals to feed higher-level systems around the individual sensors. The advantages of these sensors are lower cost, smaller size, and reliability [30].

4.1.2. Public Transport

Considering the safety and efficiency of citizens of SCs is crucial, especially on roads [31]. Therefore, municipalities are attempting to implement smart traffic using IoT development solutions [32]. In this regard, the IoT will play a crucial role in traffic management. For instance, data from various types of sensors and GPS systems are sent by drivers' smartphones in order to determine the speed, number, and locations of vehicles on a particular road. Subsequently, smart traffic lights are immediately connected to a cloud-management platform and provide timing information to automatically and accurately monitor green lights, thereby preventing traffic congestion. Additionally, these methods can predict traffic in the future and offer prevention plans, with which the transport administration department is able to detect potentially dangerous situations in time and take required actions to prevent traffic congestion [33]. Therefore, considering the obvious importance of transportation systems in SCs, specific and accurate planning to control these systems is necessary [34]. According to [35], transport technological development with the IoT will have a big revolution between 2020 and 2030, that will have direct impact on toll operators and highways and provide safe and secure networks [35]. In addition, traffic data from multiple sources, such as traffic information and ticket sales, can be used to perform sophisticated analyses and achieve better results, and train operators can maximize the capacity of tracks and easily prevent train delays [36]. Fortunately, many countries around the world, especially developing countries, are now trying to make use of new systems connected to the IoT for controlling their transportation systems [37].

4.1.3. Utilities

IoT-equipped SCs give more control to citizens over their home utilities, reducing overall bills and related costs [38]. By utilizing IoT technologies and effective approaches, such as smart meters for billing, monitoring consumption patterns, and remote monitoring, municipalities can achieve cost-effective connectivity to utility companies' IT systems. This helps customers consume energy and water based on improved monitoring and, therefore, presents better management services to the citizens [39]. Precooling optimization using system data (IoT), while preserving the thermal comfort of the inhabitants, has a direct influence on expenses and energy consumption (electricity costs) for cooling of a building by up to 30% percent, according to an Australian study [40]. Also, other research shows that, in Arabian Gulf countries, a smart energy management system using the Internet of Things can reduce costs, especially for air conditioning, which accounts for up to 60% of electricity consumption, while still meeting energy demand [41]. On the other hand, use of the IoT in utilities has a good effect on attainment of efficiency (management of large-scale solar photovoltaic systems) [42] and conservation of resources [43].

4.1.4. Street Lighting

In SCs, the maintenance and control of streetlamps can be more cost-effective and straightforward through the use of the IoT [44]. In particular, IoT systems can be paired with sensors that connect to a cloud-management solution [45], providing confident monitoring of illuminated transport paths such as streets and the movement of people and vehicles. Measuring the environmental conditions can also allow for a more accurate analysis regarding the need to improve the lighting schedule and indicate if lights should be brighter or dimmer [46]. On the other hand, IoT systems have a remarkable effect on energy-saving associated with urban street lighting as using warmer lights and increasing light uniformity can result in a 30–50% energy saving on street lighting, and for medium-sized cities with

populations around 200,000–400,000 residents, energy savings on street lighting it can reach 8–23 MWh per annum [47].

4.1.5. Waste Management

Waste collection is one of the most important sectors of SCs [48]. In this regard, IoT can reduce a lot of problems in this regard [49]. To achieve this, a sensor will be placed on each waste container, which will gather data regarding the level of waste in the container; then, after the container is filled, a notification will be sent to truck drivers via a mobile app. By following this useful and effective plan, truck drivers will expend time and energy to only empty full containers instead of half-full ones [50]. A study in China of recycling and household waste segregation between 2018 and 2019 showed that integration of the Internet of Things (IoT) was effective in household waste management. During the study, collections of recyclable waste and biodegradable food waste were elevated by 431.8% and 88.8%, respectively, which had good environment effects and meant that this macro policy increased the recyclable waste collection by 431.8% in Shanghai [51].

4.1.6. Public Safety

Theft of motor vehicles throughout the world, coupled with a massive loss of cash, is a disaster for insurance companies. For instance, just in the USA, in 2019 about USD 6.4 billion was lost to motor vehicle theft [52]. Likewise, every year, worldwide, 70 million smartphones are lost or stolen [53]. In these regards, IoT-based SC technologies are vital for offering real-time monitoring, enhancing public safety, and supporting proper decision-making, that will prevent a lot of harm to people [54]. For example, testing of the motorcycle antitheft system (MATS) showed that this system had 100% accuracy at speeds of up to 70 km/h and for speeds up to 80, it had 94.4% accurate [55].

4.1.7. Smart Parking

Parking occupies a large amount of the area in a city—81% of the city area in Los Angeles, 23% in Munich, 23% in Paris, 19% in Copenhagen, and 18% in Zurich and Hamburg. Therefore, cities must use of intelligent parking systems in order to reduce congestion and help drivers [56]. In this regard, IoT technology has built a special mobile application in order to solve vehicle parking problems and this has had a remarkable effect for drivers. Based on research, from 2013 to 2018, downloading of the mobile application, increased from 17 million downloads to 80 million, which shows the benefit of this application in solving problems related to parking [57]. Therefore, the importance of smart parking in SCs should be investigated by policymakers [58], considering that finding parking spots can improve the welfare of citizens [59]. This action can be achieved by utilizing GPS data from drivers' smartphones and road-surface sensors embedded in the ground of parking spots. As a result, drivers can be notified of occupied and vacant parking spots via a real-time parking map [60].

4.2. *Strategic Policies for Boosting Economic Recovery of Smart Cities through the IoT*

The IoT technology inherent in smart cities, promises effective options that will allow cities to be more safe, inclusive, and resilient [61]. In this regard, the IoT helps cities to improve good governance and privacy which are important for the socio-economic dimensions of urban areas [62]. In addition, the advance of 5G technologies [63] and artificial neural networks (ANNs) will prompt further innovations in smart city technologies of the IoT [64]. In fact, cloud-based IoT applications that contain information gathered from citizens could help smart cities to monitor and manage crime detection, proficiency, water supply systems, healthcare facilities, electric utilities, digital libraries, transportation networks, resource management, waste management, and security mechanisms [65]. Therefore, smart technologies such as the IoT are significant when developing SCs, while maintaining emphasis on the implemented strategies and policies [66]. It is clear that the implementation of targets related to SCs requires strong and calculated strategies and

policies [67]. In fact, achieving “smartness” is not a one-time action; it is a continuous process. Therefore, policymakers should aim to devise a plan [68] that considers the individual goals of each sector whilst also evaluating the dynamic and indirect impacts on other areas within an SC. Undoubtedly, to advance SCs and continue their expansion, officials and policymakers must vigorously strive to create a unique quality of life, work, and environment for the citizens of their cities [69]. On the other hand, since the concept of SCs falls in line with the smart grid, economic issues related to the programs that are used for demand-response management (DRM) and real-time pricing should be taken into consideration [70]. In addition, it can be added that as SCs aim to improve the quality of life of urban citizens, the success of SCs depends on participation by private companies [71]. Therefore, through the use of new communication channels between the government and its citizens, policymakers should focus on the essential needs of stakeholders, such as affordable energy, urban security, and energy security [69], because, public participation will help improve quality of life and establish trust between local governments and people [72]. This means that the investment in developing SCs has advantages for both people and the community, including a reduction in the cost of living, improvement of living standards and environmental sustainability, improvement of operational efficiency, improvement of eco-friendly infrastructure, and development of smart technology through the IoT [73]. Moreover, private investment (companies) can help governments easily overcome old issues pertaining to big cities or developing cities that have not been well planned [74]. In general, investment on IoT technologies, is opening new possibilities for cities and helping them to be smart cities [75]. According to these cases, effective strategies and policies can accelerate the conversion of a standard city into an SC and, thus, should aim to attract investment, improve IT infrastructure, integrate connected local energy storage systems in order to support better renewable energy sources on the power grids, and adapt an IoT implementation strategy based on the city’s size to reduce costs, support the utilization of smart LED streetlights in major metropolises, increase the collaboration between local governments and stakeholders, increase the utilization of new technologies such as sensors, change the mentality of the citizens, and redefine the governance model with proper politics. Based on the comprehensive explanations presented above, the most important barriers and the most appropriate solutions related to IoT-based SC development are presented in Table 1.

Table 1. The most important barriers related to SC development based on the IoT and the appropriate solutions.

Sectors	Barriers	Solution	References
Public transport	More CO ₂ emissions due to increased private cars, increase in noise from private cars in cities, lack of monitoring patterns of transport use by citizens, absence or low use of monitoring systems, lack of safety and efficiency in roads, congestion and traffic, sudden accidents, and defective roads	Utilization of proper patterns and use of monitoring sensors, utilization of different types of sensors to accurately monitor roads and improve GPS systems using data from drivers’ smartphones, and improvement in the quality of roads	[29,31,34,36,76–80]
Street lighting	Lack of sensor-equipped street lights and defective lights	Streetlights with sensors and establishment of a connection between the sensors and cloud management, utilization of a monitoring-system switch to scan conditions and send signals to increase or dim the lights, and use of new lights (low consumption)	[43–46,81–84]
Utilities	Excessive consumption, extra expenses for fuel and electricity for which there is no need, lack of or improper use of smart meters and smart billing, shortage of revealing consumption patterns, and limited remote monitoring for citizens	T-equipped smart-connected meters, proper consumption patterns, and management services to improve the quality of the services	[46,85–87]

Table 1. Cont.

Sectors	Barriers	Solution	References
Smart parking	Lack of or limited smart parking options for drivers, improper parking of cars on the street, and reduction in street width due to traffic	Utilization of GPS data from drivers' smartphones or road-surface sensors embedded in the ground on parking spots	[58–60,88,89]
Waste management	Release of garbage and the resultant unpleasant odor in cities due to absence of accurate systems to monitor the proper time to collect waste in order to prevent fuel losses and empty containers	Installation of a sensor on waste containers to optimize waste-collection schedules by tracking waste levels	[90,91]
Environment	Increased CO ₂ emissions and threats public health and lack of water and air quality monitoring	Utilization of different types of sensors such as water sensors and air sensors to improve and provide more accurate monitoring	[24–28,92–94]
Public safety	Weak security of public safety in cities, increase in crimes such as robbery, lack of ethics regarding law and regulatory rights, and weak infrastructure	Utilization of IoT technologies such as CCTV cameras and acoustic sensors in different areas of cities, blockchain-based security management of IoT infrastructure for maintaining security and privacy, improvement of interoperability, leading to vendor lock-in, and control of corruption	[95–103]

To complete this Table, other investigations can be added. For example one of the greatest challenges at present is the low or inadequate quality of the life in many areas of the world. This means that in many areas, use of energy is not based on world standards and there is less use of modern technologies to manage it. This affects the quality of life of citizens, and, in particular, the economy of households [104,105]. Without a doubt, collection of the wastes in crowded areas especially cities, is important for citizens, because it prevents illness. Utilization of IoT is very important in providing more efficient waste management and overcoming other problems in this area [106]. Healthcare is one of the most important challenges for governments because low health of the citizens can have negative effects on people. In this regard, IoT systems can improve net health and increase people's health knowledge [107]. In addition, using e-health services, for instance in a global pandemic such as COVID-19, for data collection by citizens, for giving health advice through the Internet, and for increasing the health of medical staff is fruitful [108]. On the other hand, as mentioned previously, transportation systems are one of most important sectors in many countries because transport has a large effect on the environment and the movement of people. Therefore, today, the emergence of IoT systems inside cars and the conversion to smart cars (vehicles), helps the environment and can also move people easily without loss of time [109,110]. In addition, in order to reduce traffic and help the environment, greater utilization of bicycles and an increase in bike-sharing services has been implemented through IoT systems [111]. Moreover, to improve the electrical energy saving of the cars, increasing the lifetime of battery-operated devices (by up to a couple of years) by using of IoT systems is possible because IoT systems are able, during inactive periods, to keep the device in a low-power state [112]. The issue of the elderly and their care is also important in many countries. Fortunately, however, IoT systems have provided assistance applications through the use of a single wearable device in both outdoor and indoor locations. These systems are able to recognize changes in the behavior of elderly people, are low-cost, unobtrusive, have a low power consumption, and can easily prevent problems [113].

Table 2 shows a comparison between results of this work and a number of works in the literature. As can be seen, most previous work has investigated limited subjects related to the IoT and smart cities, while this work comprehensively investigated these subjects.

Table 2. A comparison of the results of this work and those of previous work.

Present Work	Types of Consideration in Other Work	Reference
Environment	Reviews of IoT-based environment monitoring systems	[114]
Public transport	Managing the public transport systems through applying digital technologies	[115]
Utilities	Investigation of the role of digitalization for smart water networks and improvement by the IoT, artificial intelligence, blockchain, and other novel technologies	[116]
Street lighting	Investigation of the street lighting framework using IoT systems to reduce cost and energy consumption	[117]
Waste management	Proposal of a proof-of-concept municipal waste management system to reduce the cost of waste classification, monitoring, and collection using the IoT	[118]
Public safety	Increase in public safety against theft using IoT systems	[119]
Smart parking	Investigation the role the Internet of Things (IoT) in overcoming the challenges of parking cars. Presentation of smart parking solutions	[120]

5. Conclusions

Future with less CO₂ [121], and relying upon renewable energy [122] as main fuels, are some of the most important goals of scientists and researchers. In this regard, the role of IoT in controlling CO₂ emissions and managing energy consumption is important. This work investigated the problems related to seven important sectors of the IoT, namely environment, public transport, utilities, smart parking, public safety, waste management, and smart lighting. Each sector was analyzed carefully to identify the challenges to be mitigated or removed such that building SCs would be possible. For instance, in the environment sector, the utilization of air and water sensors allows us to monitor air and water quality and detect the amount of CO₂, sulfur oxides, and nitrogen, water leakage, and changes in the pH level and chemical composition of water over time, as well as other factors that have potentially detrimental effects. In terms of road traffic, the determination of the speed, number, and locations of vehicles and monitoring of green-light timings can be achieved through the use of various types of sensors and GPS data collected from drivers' smartphones. Across the public-transport sector, IoT sensors can help enhance our travel experiences and achieve a higher level of safety and punctuality. In utility monitoring, the IoT allows users to control their home utilities for billing, consumption patterns, and remote monitoring. In particular, via cost-effective connectivity to utility companies' IT systems, customers can adjust their energy and water consumption more economically. For street lighting, utilizing both IoT systems and sensors connected to a cloud-management solution can ensure the confident monitoring of illuminance for the safe movement of people and vehicles. In terms of environmental effects, we can improve the lighting schedule and determine which areas require different intensities of light (some streets may only need a dim light, so less electricity would be used). In the waste-management sector, the use of IoT technologies can lead to the optimization of waste-collection schedules by tracking the waste levels, providing route optimization, and ensuring useful operational analytics. In this regard, each waste container would be implanted with a sensor that gathers data on the level of waste in a container. Then, a notification of filled containers would be sent to truck drivers via a mobile app. This is a strategic plan to avoid emptying half-full containers, resulting in less travel by trucks and reducing GHG emissions. In the public-safety sector, IoT-based SC technologies have a crucial role in offering real-time monitoring, enhancing public safety, and developing decision-making tools and analytics through CCTV cameras and acoustic sensors. At the same time, data from social media feeds can be carefully analyzed to improve public safety in a city and predict potential crime scenes. For the smart-parking sectors, IoT technologies can help drivers identify available parking spots

on a real-time map based on GPS data extracted from drivers' smartphones or road-surface sensors embedded in parking spots.

Author Contributions: Several authors contributed to this research: methodology, validation, review, and editing, A.R. (Armin Razmjoo), A.G., and M.M.; formal analysis and investigation, S.A. and A.R. (Alireza Rezvani); resources, D.A.G.; writing, and final analysis, S.M. (Seyedali Mirjalili) and S.M. (Saim Memon). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kandt, J.; Batty, M. Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities* **2020**, *109*, 102992. [\[CrossRef\]](#)
- Razmjoo, A.; Nezhad, M.M.; Kaigutha, L.G.; Marzband, M.; Mirjalili, S.; Pazhoohesh, M.; Memon, S.; Ehyaei, M.A.; Piras, G. Investigating smart city development based on green buildings, electrical vehicles and feasible indicators. *Sustainability* **2021**, *13*, 7808. [\[CrossRef\]](#)
- Razmjoo, A.; Østergaard, P.A.; Denai, M.; Nezhad, M.M.; Mirjalili, S. Effective policies to overcome barriers in the development of smart cities. *Energy Res. Soc. Sci.* **2021**, *79*, 102175. [\[CrossRef\]](#)
- Abella, A.; Ortiz-De-Urbina-Criado, M.; De-Pablos-Heredero, C. A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. *Cities* **2017**, *64*, 47–53. [\[CrossRef\]](#)
- Kundu, D. *Developing National Urban Policies: Ways Forward to Green and Smart Cities*; Springer: Singapore, 2020.
- Gil-Garcia, J.R.; Pardo, T.A. E-government success factors: Mapping practical tools to theoretical foundations. *Gov. Inf. Q.* **2005**, *22*, 187–216. [\[CrossRef\]](#)
- Park, E.; Del Pobil, A.P.; Kwon, S.J. The Role of Internet of Things (IoT) in Smart Cities: Technology Roadmap-oriented Approaches. *Sustainability* **2018**, *10*, 1388. [\[CrossRef\]](#)
- Sarin, G. Developing smart cities using Internet of Things: An empirical study. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIA Com), New Delhi, India, 16–18 March 2016.
- Li, X. Big data analysis of the Internet of Things in the digital twins of smart city based on deep learning. *Future Generation Comput. Syst.* **2022**, *128*, 167–177. [\[CrossRef\]](#)
- Cepeda-Pacheco, J.C.; Domingo, M.C. Deep learning and Internet of Things for tourist attraction recommendations in smart cities. *Neural Comput. Applic.* **2022**. [\[CrossRef\]](#)
- Hammi, B.; Khatoun, R.; Zeadally, S.; Fayad, A.; Khoukhi, L. IoT technologies for smart cities. *IET Netw.* **2018**, *7*, 1–13. [\[CrossRef\]](#)
- Ejaz, W.; Naeem, M.; Shahid, A.; Anpalagan, A.; Jo, M. Efficient Energy Management for the Internet of Things in Smart Cities. *IEEE Commun. Mag.* **2017**, *55*, 84–91. [\[CrossRef\]](#)
- Ahmad, T.; Zhang, D. Using the internet of things in smart energy systems and networks. *Sustain. Cities Soc.* **2021**, *68*, 102783. [\[CrossRef\]](#)
- Bresciani, S.; Ferraris, A.; Del Giudice, M. The management of organizational ambidexterity through alliances in a new context of analysis: Internet of Things (IoT) smart city projects. *Technol. Forecast. Soc. Chang.* **2018**, *136*, 331–338. [\[CrossRef\]](#)
- Evertzen, W.H.N.; Effing, R.; Constantinides, E. The Internet of Things as Smart City Enabler: The Cases of Palo Alto, Nice and Stockholm. In *Digital Transformation for a Sustainable Society in the 21st Century*; Pappas, I., Mikalef, P., Dwivedi, Y., Jaccheri, L., Krogstie, J., Mäntymäki, M., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2019; Volume 11701. [\[CrossRef\]](#)
- AlZubi, A.A.; Alarifi, A.; Al-Maitah, M.; Alheyasat, O. Multi-sensor information fusion for Internet of Things assisted automated guided vehicles in smart city. *Sustain. Cities Soc.* **2020**, *64*, 102539. [\[CrossRef\]](#)
- Kumar, S.; Tiwari, P.; Zymbler, M. Internet of Things is a revolutionary approach for future technology enhancement: A review. *J. Big Data* **2019**, *6*, 1–21. [\[CrossRef\]](#)
- Jha, S.; Nkenyereye, L.; Joshi, G.P.; Yang, E. Mitigating and Monitoring Smart City using Internet of Things. *Comput. Mater. Contin.* **2020**, *65*, 1059–1079. [\[CrossRef\]](#)
- Vinod Kumar, T.M. Smart Environment for Smart Cities. In *Smart Environment for Smart Cities*; Vinod Kumar, T., Ed.; Advances in 21st Century Human Settlements; Springer: Singapore, 2020. [\[CrossRef\]](#)
- Allam, Z.; Dhunny, Z.A. On big data, artificial intelligence and smart cities. *Cities* **2019**, *89*, 80–91. [\[CrossRef\]](#)
- Latre, S. City of things: An integrated and multi-technology testbed for IoT smart city experiments. In Proceedings of the 2016 IEEE International Smart Cities Conference (ISC2), Trento, Italy, 12–15 September 2016. [\[CrossRef\]](#)

22. Armin, R. Implementation of energy sustainability using hybrid power systems, a case study. *Energy Sources Part A Recovery Util. Environ. Eff.* **2019**, 1–14. [CrossRef]
23. How Smart Cities Can Help Tackle Climate Change. Available online: http://www.frontier-economics.com/uk/en/news-and-articles/articles/article-i4604-how-smart-cities-can-help-tackle-climate-change/#_ftn1 (accessed on 12 January 2022).
24. Doni, A. Survey on multi sensor based air and water quality monitoring using IoT. *Indian J. Sci. Res.* **2018**, *17*, 147–153.
25. Duangsuwan, S. A Development on Air Pollution Detection Sensors based on NB-IoT Network for Smart Cities. In Proceedings of the 2018 18th International Symposium on Communications and Information Technologies (ISCIT), Bangkok, Thailand, 26–29 September 2018. [CrossRef]
26. Kafli, N. Internet of Things (IoT) for measuring and monitoring sensors data of water surface platform. In Proceedings of the 2017 IEEE 7th International Conference on Underwater System Technology: Theory and Applications (USYS), Kuala Lumpur, Malaysia, 18–20 December 2017. [CrossRef]
27. Ullo, S.L.; Sinha, G.R. Advances in Smart Environment Monitoring Systems Using IoT and Sensors. *Sensors* **2020**, *20*, 3113. [CrossRef]
28. Randazzo, G.; Italiano, F.; Micallef, A.; Tomasello, A.; Cassetti, F.P.; Zammit, A.; D’Amico, S.; Saliba, O.; Cascio, M.; Cavallaro, F.; et al. WebGIS Implementation for Dynamic Mapping and Visualization of Coastal Geospatial Data: A Case Study of BESS Project. *Appl. Sci.* **2021**, *11*, 8233. [CrossRef]
29. Smart Cities: Digital Solutions for a More Livable Future. Available online: <https://www.mckinsey.com/~{}~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/Smart%20cities%20Digital%20solutions%20for%20a%20more%20livable%20future/MGI-Smart-Cities-Full-Report.pdf> (accessed on 12 January 2022).
30. Sanders, D. Environmental sensors and networks of sensors. *Sens. Rev.* **2008**, *28*. [CrossRef]
31. Vishal, D. IoT-driven road safety system. In Proceedings of the 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, India, 15–16 December 2017. [CrossRef]
32. Majumdar, S.; Subhani, M.M.; Roullier, B.; Anjum, A.; Zhu, R. Congestion prediction for smart sustainable cities using IoT and machine learning approaches. *Sustain. Cities Soc.* **2020**, *64*, 102500. [CrossRef]
33. Al-Sakran, H.O. Intelligent traffic information system based on integration of Internet of Things and Agent technology. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 37–43.
34. Chen, C.; Zhang, Y.; Khosravi, M.R.; Pei, Q.; Wan, S. An Intelligent Platooning Algorithm for Sustainable Transportation Systems in Smart Cities. *IEEE Sens. J.* **2020**, *21*, 15437–15447. [CrossRef]
35. Azmat, M.; Kummer, S.; Moura, L.T.; Di Gennaro, F.; Moser, R. Future Outlook of Highway Operations with Implementation of Innovative Technologies Like AV, CV, IoT and Big Data. *Logistics* **2019**, *3*, 15. [CrossRef]
36. Simmhan, Y.; Ravindra, P.; Chaturvedi, S.; Hegde, M.; Ballamajalu, R. Towards a data-driven IoT software architecture for smart city utilities. *Softw. Pr. Exp.* **2018**, *48*, 1390–1416. [CrossRef]
37. Rehman, A.; Mumtaz, T. On Board Intelligence for Public Transportation in Developing Countries using IoT. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2020; Volume 899, p. 012004.
38. Saravanan, K.; Julie, E.G.; Robinson, Y.H. Smart Cities & IoT: Evolution of Applications, Architectures & Technologies, Present Scenarios & Future Dream. In *Internet of Things and Big Data Analytics for Smart Generation*; Balas, V., Solanki, V., Kumar, R., Khari, M., Eds.; Intelligent Systems Reference Library; Springer: Cham, Switzerland, 2019; Volume 154. [CrossRef]
39. Suresh, M. A novel smart water-meter based on IoT and smartphone app for city distribution management. In Proceedings of the 2017 IEEE Region 10 Symposium (TENSYP), Cochin, India, 14–16 July 2017. [CrossRef]
40. Vishwanath, A.; Chandan, V.; Saurav, K. An IoT-Based Data Driven Precooling Solution for Electricity Cost Savings in Commercial Buildings. *IEEE Internet Things J.* **2019**, *6*, 7337–7347. [CrossRef]
41. Al-Ali, A. A smart home energy management system using IoT and big data analytics approach. In *IEEE Transactions on Consumer Electronics*; IEEE: Piscataway, NJ, USA, 2017; Volume 63. [CrossRef]
42. Shapsough, S.; Takroui, M.; Dhaouadi, R.; Zualkernan, I.A. Using IoT and smart monitoring devices to optimize the efficiency of large-scale distributed solar farms. *Wirel. Netw.* **2018**, *27*, 4313–4329. [CrossRef]
43. Arshad, R.; Zahoor, S.; Shah, M.A.; Wahid, A.; Yu, H. Green IoT: An Investigation on Energy Saving Practices for 2020 and Beyond. *IEEE Access* **2017**, *5*, 15667–15681. [CrossRef]
44. Carli, R.; Dotoli, M.; Cianci, E. An optimization tool for energy efficiency of street lighting systems in smart cities. *IFAC-Pap.* **2017**, *50*, 14460–14464. [CrossRef]
45. Yusoff, Y.M.; Rosli, R.; Karnaluddin, M.U.; Samad, M. Towards smart street lighting system in Malaysia. In Proceedings of the 2013 IEEE Symposium on Wireless Technology & Applications (ISWTA), Kuching, Malaysia, 22–25 September 2013. [CrossRef]
46. Barve, V. Smart lighting for smart cities. In Proceedings of the 2017 IEEE Region 10 Symposium (TENSYP), Cochin, India, 14–16 July 2017. [CrossRef]
47. Saad, R.; Portnov, B.A.; Trop, T. Saving energy while maintaining the feeling of safety associated with urban street lighting. *Clean Technol. Environ. Policy* **2020**, *23*, 251–269. [CrossRef]
48. Singh, A.; Aggarwal, P.; Arora, R. IoT based waste collection system using infrared sensors. In Proceedings of the 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 7–9 September 2016. [CrossRef]

49. Nasreen Banu, M.I.; Metilda Florence, S. Convergence of Artificial Intelligence in IoT Network for the Smart City—Waste Management System. In *Expert Clouds and Applications. Lecture Notes in Networks and Systems*; Jeena Jacob, I., Gonzalez-Longatt, F.M., Kolandapalayam Shanmugam, S., Izonin, I., Eds.; Springer: Singapore, 2022; Volume 209. [CrossRef]
50. Medvedev, A.; Fedchenkov, P.; Zaslavsky, A.; Anagnostopoulos, T.; Khoruzhnikov, S. Waste Management as an IoT-Enabled Service in Smart Cities. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*; Balandin, S., Andreev, S., Koucheryavy, Y., Eds.; ruSMART 2015, NEW2AN 2015. Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9247. [CrossRef]
51. Jiang, P.; Van Fan, Y.; Klemeš, J.J. Data analytics of social media publicity to enhance household waste management. *Resour. Conserv. Recycl.* **2020**, *164*, 105146. [CrossRef] [PubMed]
52. Insurance Information Institute. Available online: <https://www.iii.org/fact-statistic/facts-statistics-auto-theft#Motor%20Vehicle%20Theft,%202010-2019> (accessed on 30 March 2021).
53. Hom, E.J. Mobile Device Security: Startling Statistics on Data Loss and Data Breaches. 2017. Available online: <http://www.channelpronetwork.com/article/mobile-device-security-startling-statistics-data-loss-anddata-breaches> (accessed on 7 March 2021).
54. Fraga-Lamas, P.; Fernández-Caramés, T.M.; Suárez-Albela, M.; Castedo, L.; González-López, M. A Review on Internet of Things for Defense and Public Safety. *Sensors* **2016**, *16*, 1644. [CrossRef]
55. Papadakis, N. An IoT-Based Participatory Antitheft System for Public Safety Enhancement in Smart Cities. *Smart Cities* **2021**, *4*, 919–937. [CrossRef]
56. Kalašová, A.; Čulík, K.; Poliak, M.; Otašálová, Z. Smart Parking Applications and Its Efficiency. *Sustainability* **2021**, *13*, 6031. [CrossRef]
57. Pariama, R.E.; Manaha, R. Suyoto Parking-RR: Mobile Application Malioboro Smart Parking Based on IoT Technology. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *704*, 012037. [CrossRef]
58. Al-Turjman, F.; Malekloo, A. Smart parking in IoT-enabled cities: A survey. *Sustain. Cities Soc.* **2019**, *49*, 101608. [CrossRef]
59. Sadhukhan, P. An IoT-based E-parking system for smart cities. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017. [CrossRef]
60. Lanza, J.; Sanchez, L.; Gutiérrez, V.; Galache, J.A.; Santana, J.R.; Sotres, P.; Muñoz, L. Smart City Services over a Future Internet Platform Based on Internet of Things and Cloud: The Smart Parking Case. *Energies* **2016**, *9*, 719. [CrossRef]
61. Allam, Z. Redefining the Smart City: Culture, Metabolism and Governance. *Smart Cities* **2018**, *1*, 4–25. [CrossRef]
62. Allam, Z. The Emergence of Anti-Privacy and Control at the Nexus between the Concepts of Safe City and Smart City. *Smart Cities* **2019**, *2*, 96–105. [CrossRef]
63. Allam, Z.; Jones, D.S. Future (post-COVID) digital, smart and sustainable cities in the wake of 6G: Digital twins, immersive realities and new urban economies. *Land Use Policy* **2021**, *101*, 105201. [CrossRef]
64. Allam, Z. Achieving Neuroplasticity in Artificial Neural Networks through Smart Cities. *Smart Cities* **2019**, *2*, 118–134. [CrossRef]
65. Alam, T. Cloud-Based IoT Applications and Their Roles in Smart Cities. *Smart Cities* **2021**, *4*, 1196–1219. [CrossRef]
66. Chatterjee, S. Success of IoT in Smart Cities of India: An empirical analysis. *Gov. Inf. Q. J.* **2018**, *35*, 349–361. [CrossRef]
67. Sokolov, A.; Veselitskaya, N.; Carabias, V.; Yildirim, O. Scenario-based identification of key factors for smart cities development policies. *Technol. Forecast. Soc. Chang.* **2019**, *148*, 119729. [CrossRef]
68. Makeen, P.; Ghali, H.A.; Memon, S. Experimental and Theoretical Analysis of the Fast Charging Polymer Lithium-Ion Battery Based on Cuckoo Optimization Algorithm (COA). *IEEE Access* **2020**, *8*, 140486–140496. [CrossRef]
69. Kumar, H.; Singh, M.K.; Gupta, M.; Madaan, J. Moving towards smart cities: Solutions that lead to the Smart City Transformation Framework. *Technol. Forecast. Soc. Chang.* **2018**, *153*, 119281. [CrossRef]
70. Kumari, A.; Gupta, R.; Tanwar, S.; Tyagi, S.; Kumar, N. When Blockchain Meets Smart Grid: Secure Energy Trading in Demand Response Management. *IEEE Netw.* **2020**, *34*, 299–305. [CrossRef]
71. Dameri, R.P. Triple Helix in Smart Cities: A Literature Review about the Vision of Public Bodies, Universities, and Private Companies. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016. [CrossRef]
72. Castelnovo, W. Smart Cities Governance: The Need for a Holistic Approach to Assessing Urban Participatory Policy Making. *Soc. Sci. Comput. Rev.* **2015**, *34*, 724–739. [CrossRef]
73. Milenković, M. Using Public Private Partnership models in smart cities—Proposal for Croatia. In Proceedings of the 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 22–26 May 2017. [CrossRef]
74. Jonek-Kowalska, I.; Wolniak, R. Sharing Economies’ Initiatives in Municipal Authorities’ Perspective: Research Evidence from Poland in the Context of Smart Cities’ Development. *Sustainability* **2022**, *14*, 2064. [CrossRef]
75. Belli, L. IoT-Enabled Smart Sustainable Cities: Challenges and Approaches. *Smart Cities* **2020**, *3*, 1039–1071. [CrossRef]
76. Khan, S. Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. *Comput. Electr. Eng.* **2021**, *89*, 106906. [CrossRef]
77. Prasanth, R.S. Preventing Road Accidents using IoT. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 419–423. [CrossRef]
78. Porru, S.; Missoa, F.E.; Pani, F.E.; Repetto, C. Smart mobility and public transport: Opportunities and challenges in rural and urban areas. *J. Traffic Transp. Eng.* **2020**, *7*, 88–97. [CrossRef]

79. Harini, B.K. Increasing Efficient Usage of Real-Time Public Transportation Using IOT, Cloud and Customized Mobile App. *SN Comput. Sci.* **2020**, *1*, 159. [[CrossRef](#)]
80. Ladha, A.; Bhattacharya, P.; Chaubey, N.; Bodkhe, U. IIGPTS: IoT-Based Framework for Intelligent Green Public Transportation System. In *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*; Singh, P., Pawłowski, W., Tanwar, S., Kumar, N., Rodrigues, J., Obaidat, M., Eds.; Lecture Notes in Networks and Systems; Springer: Singapore, 2020; Volume 121. [[CrossRef](#)]
81. Umamaheswari, S. Smart Street Lighting in Smart Cities: A Transition from Traditional Street Lighting. In *Security and Privacy Applications for Smart City Development*; Tamane, S.C., Dey, N., Hassanien, A.E., Eds.; Studies in Systems, Decision and Control; Springer: Cham, Switzerland, 2021; Volume 308. [[CrossRef](#)]
82. Zhang, J.J.; Zeng, W.H.; Hou, S.L.; Chen, Y.Q.; Guo, L.Y.; Li, Y.X. A low-power and low cost smart streetlight system based on Internet of Things technology. *Telecommun. Syst.* **2022**, *79*, 83–93. [[CrossRef](#)]
83. Ali, M.; Youssef, T.; Mohamed, A. A cost-effective viable strategy for gradually transitioning Egypt's cities into truly IOT-enabled smart cities. *Int. J. Ind. Sustain. Dev.* **2020**, *1*, 1–5. [[CrossRef](#)]
84. Bhukya, K.A.; Ramasubbareddy, S.; Govinda, K.; Srinivas, T.A.S. Adaptive Mechanism for Smart Street Lighting System. In *Smart Intelligent Computing and Applications*; Satapathy, S., Bhateja, V., Mohanty, J., Udgate, S., Eds.; Smart Innovation, Systems and Technologies; Springer: Singapore, 2020; Volume 160. [[CrossRef](#)]
85. Patel, S. Role of smart meters in smart city development in India. In Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 4–6 July 2016. [[CrossRef](#)]
86. Fettermann, D.C.; Borriello, A.; Pellegrini, A.; Cavalcante, C.G.; Rose, J.M.; Burke, P.F. Getting smarter about household energy: The who and what of demand for smart meters. *Build. Res. Inf.* **2020**, *49*, 100–112. [[CrossRef](#)]
87. Avancini, D.B.; Rodrigues, J.J.P.C.; Rabêlo, R.A.L.; Das, A.K.; Kozlov, S.; Solic, P. A new IoT-based smart energy meter for smart grids. *Int. J. Energy Res.* **2020**, *45*, 189–202. [[CrossRef](#)]
88. Gupta, R.; Budhiraja, N.; Mago, S.; Mathur, S. An IoT-Based Smart Parking Framework for Smart Cities. In *Data Management, Analytics and Innovation*; Sharma, N., Chakrabarti, A., Balas, V., Martinovic, J., Eds.; Advances in Intelligent Systems and Computing; Springer: Singapore, 2021; Volume 1174. [[CrossRef](#)]
89. Said, A.M.; Kamal, A.E.; Afifi, H. An intelligent parking sharing system for green and smart cities based IoT. *Comput. Commun.* **2021**, *172*, 10–18. [[CrossRef](#)]
90. Ali, T.; Irfan, M.; Alwadie, A.S.; Glowacz, A. IoT-Based Smart Waste Bin Monitoring and Municipal Solid Waste Management System for Smart Cities. *Arab. J. Sci. Eng.* **2020**, *45*, 10185–10198. [[CrossRef](#)]
91. Shah, A.A.I. A review of IoT-based smart waste level monitoring system for smart cities. *Indones. J. Electr. Eng. Comput. Sci.* **2021**, *21*, 450–456. [[CrossRef](#)]
92. Carminati, M.; Sinha, G.R.; Mohdiwale, S.; Ullo, S.L. Miniaturized Pervasive Sensors for Indoor Health Monitoring in Smart Cities. *Smart Cities* **2021**, *4*, 146–155. [[CrossRef](#)]
93. Carminati, M.; Turolla, A.; Mezzerà, L.; Di Mauro, M.; Tizzoni, M.; Pani, G.; Zanetto, F.; Foschi, J.; Antonelli, M. A Self-Powered Wireless Water Quality Sensing Network Enabling Smart Monitoring of Biological and Chemical Stability in Supply Systems. *Sensors* **2020**, *20*, 1125. [[CrossRef](#)]
94. Geetha, S.; Gouthami, S. Internet of things enabled real time water quality monitoring system. *Smart Water* **2016**, *2*, 1. [[CrossRef](#)]
95. Da Xu, L.; He, W.; Li, S. Internet of Things in Industries: A Survey. *IEEE Trans. Ind. Inform.* **2014**, *10*, 2233–2243.
96. Salles, R.S.; Ribeiro, P.F. Smart Cities, Connected World, and Internet of Things. In *Software Defined Internet of Everything. Internet of Things (Technology, Communications and Computing)*; Aujla, G.S., Garg, S., Kaur, K., Sikdar, B., Eds.; Springer: Cham, Switzerland, 2022. [[CrossRef](#)]
97. Yan, Z.; Zhang, P.; Vasilakos, A.V. A survey on trust management for Internet of Things. *J. Netw. Comput. Appl.* **2014**, *42*, 120–134. [[CrossRef](#)]
98. Cho, J.-R.; Kim, H.S.; Chae, D.K.; Lim, S.J. Smart CCTV Security Service in IoT (Internet of Things) Environment. *J. Digit. Contents Soc.* **2017**, *18*, 1135–1142. [[CrossRef](#)]
99. Losavio, M.M.; Chow, K.P.; Koltay, A.; James, J. The Internet of Things and the Smart City: Legal challenges with digital forensics, privacy, and security. *Secur. Priv.* **2018**, *1*, e23. [[CrossRef](#)]
100. Chatzimichail, A. Smart Interconnected Infrastructure for Security and Safety in Public Places. In Proceedings of the 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, 29–31 May 2019. [[CrossRef](#)]
101. Albán, P.; Roberto, L.; Mejía, A.; Vallejo, I.; Alban, I.; Cofre, S.; Molina, E. Public Order Disruption Event Detection Based on IoT Technology. An Approach for the Improvement of Public Security Conditions. In *Developments and Advances in Defense and Security*; Rocha, Á., Paredes-Calderón, M., Guarda, T., Eds.; MICRADS 2020; Smart Innovation, Systems and Technologies; Springer: Singapore, 2020; Volume 181. [[CrossRef](#)]
102. Chatzigeorgiou, C. Increasing safety and security in public places using IoT devices. In Proceedings of the 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2–16 June 2020. [[CrossRef](#)]
103. Mehedi, S.K.T.; Shamim, A.A.M.; Miah, M.B.A. Blockchain-based security management of IoT infrastructure with Ethereum transactions. *Iran J. Comput. Sci.* **2019**, *2*, 189–195. [[CrossRef](#)]
104. Moniruzzaman, M.; Khezzr, S.; Yassine, A.; Benlamri, R. Blockchain for smart homes: Review of current trends and research challenges. *Comput. Electr. Eng.* **2020**, *83*, 106585. [[CrossRef](#)]

105. Stavarakas, V.; Flamos, A. A modular high-resolution demand-side management model to quantify benefits of demand-flexibility in the residential sector. *Energy Convers. Manag.* **2020**, *205*, 112339. [[CrossRef](#)]
106. Fan, Y.V.; Lee, C.T.; Lim, J.S.; Klemeš, J.J.; Le, P.T.K. Cross-disciplinary approaches towards smart, resilient and sustainable circular economy. *J. Clean. Prod.* **2019**, *232*, 1482–1491. [[CrossRef](#)]
107. Farahani, B.; Barzegari, M.; Shams Aliee, F.; Shaik, K.A. Towards collaborative intelligent IoT eHealth: From device to fog, and cloud. *Microprocess. Microsyst.* **2020**, *72*, 102938. [[CrossRef](#)]
108. Web Source: World Health Organization. Available online: [Who.int/health-topics/coronavirus#tab=tab_1](https://www.who.int/health-topics/coronavirus#tab=tab_1) (accessed on 2 April 2020).
109. Ajanovic, A.; Haas, R. Economic and environmental prospects for battery electric- and fuel cell vehicles: A review. *Fuel. Cell.* **2019**, *19*, 515–529. [[CrossRef](#)]
110. Chugh, A.; Jain, C.; Mishra, V.P. IoT-based multifunctional smart toy car. *Lect. Notes Netw. Syst.* **2020**, *103*, 455–461.
111. Gao, P.; Li, J. Understanding sustainable business model: A framework and a case study of the bike-sharing industry. *J. Clean. Prod.* **2020**, *267*, 122229. [[CrossRef](#)]
112. Nižetić, S. Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. *J. Clean. Prod.* **2020**, *274*, 122877. [[CrossRef](#)]
113. Villarrubia, G.; Bajo, J.; de Paz, J.F.; Corchado, J.M. Monitoring and detection platform to prevent anomalous situations in home care. *Sensors* **2014**, *14*, 9900–9921. [[CrossRef](#)]
114. Asha, P. IoT enabled environmental toxicology for air pollution monitoring using AI techniques. *Environ. Res.* **2022**, *205*. [[CrossRef](#)]
115. Bolobonov, D.; Frolov, A.; Borremans, A.; Schuur, P. Managing Public Transport Safety Using Digital Technologies. *Transp. Res. Procedia* **2021**, *54*, 862–870. [[CrossRef](#)]
116. Mounce, S.R. Data Science Trends and Opportunities for Smart Water Utilities. In *ICT for Smart Water Systems: Measurements and Data Science*; Scozzari, A., Mounce, S., Han, D., Soldovieri, F., Solomatine, D., Eds.; The Handbook of Environmental Chemistry; Springer: Cham, Switzerland, 2020; Volume 102. [[CrossRef](#)]
117. Mamoon, H. Energy Optimization for Smart Cities Using IoT. *Appl. Artif. Intell.* **2022**. [[CrossRef](#)]
118. Wang, C.; Qin, J.; Qu, C.; Ran, X.; Liu, C.; Chen, B. A smart municipal waste management system based on deep-learning and Internet of Things. *Waste Manag.* **2021**, *135*, 20–29. [[CrossRef](#)]
119. Ghasan, F.H. A review on 5G technology for smart energy management and smart buildings in Singapore. *Energy AI* **2022**, *7*. [[CrossRef](#)]
120. Alessandro, F. A Social IoT-based platform for the deployment of a smart parking solution. *Comput. Netw.* **2022**, *205*. [[CrossRef](#)]
121. Armin, R. Development of Sustainable Energy Use with Attention to Fruitful Policy. *Sustainability* **2021**, *13*, 13840. [[CrossRef](#)]
122. Armin, R. Development of smart energy systems for communities: Technologies, policies and applications. *Energy* **2022**. [[CrossRef](#)]

Review

Review of the Methods to Optimize Power Flow in Electric Vehicle Powertrains for Efficiency and Driving Performance

Izhari Izmi Mazali ^{1,*}, Zul Hilmi Che Daud ¹, Mohd Kameil Abdul Hamid ¹, Victor Tan ¹,
Pakharuddin Mohd Samin ¹, Abdullah Jubair ¹, Khairul Amilin Ibrahim ¹, Mohd Salman Che Kob ¹,
Wang Xinrui ² and Mat Hussin Ab Talib ¹

- ¹ School of Mechanical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia; hilmi@mail.fkm.utm.my (Z.H.C.D.); kameil@utm.my (M.K.A.H.); tanvictor@graduate.utm.my (V.T.); pakhar@utm.my (P.M.S.); abdullahjubair18@yahoo.com (A.J.); amilinz69@gmail.com (K.A.I.); salman_smv@yahoo.com (M.S.C.K.); mathussin@utm.my (M.H.A.T.)
- ² Department of Automotive Engineering, Zunyi Vocational and Technical College, Zunyi 563006, China; wangxinrui@graduate.utm.my
- * Correspondence: izhari@mail.fkm.utm.my

Abstract: Electric vehicles (EV) are quickly gaining a foothold in global markets due to their zero tailpipe emissions and increasing practicality in terms of battery technologies. However, even though EV powertrains emit zero emissions during driving, their efficiency has not been fully optimized, particularly due to the commonly used single-speed transmission. Hence, this paper provides an extensive review on the latest works carried out to optimize the power flow in EV powertrains using multispeed discrete transmission, continuously variable transmission and multi-motor configurations. The relevant literatures were shortlisted using a keyword search related to EV powertrain in the ScienceDirect and Scopus databases. The review focused on the related literatures published from 2018 onwards. The publications were reviewed in terms of the methodologies applied to optimize the powertrain for efficiency and driving performance. Next, the significant findings from these literatures were discussed and compared. Finally, based on the review, several future key research areas in EV powertrain efficiency and performance are highlighted.

Keywords: electric vehicle powertrain; multispeed discrete transmission; continuously variable transmission; two-motors configuration; four-motors configuration

Citation: Mazali, I.I.; Daud, Z.H.C.; Hamid, M.K.A.; Tan, V.; Samin, P.M.; Jubair, A.; Ibrahim, K.A.; Kob, M.S.C.; Xinrui, W.; Talib, M.H.A. Review of the Methods to Optimize Power Flow in Electric Vehicle Powertrains for Efficiency and Driving Performance. *Appl. Sci.* **2022**, *12*, 1735. <https://doi.org/10.3390/app12031735>

Academic Editors: Anselme Muzirafuti, Giovanni Randazzo and Dimitrios S. Paraforos

Received: 11 January 2022

Accepted: 28 January 2022

Published: 8 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Electric vehicles (EVs), which offer zero emissions during driving, are quickly gaining market share recently due to their increasing practicality; contributed by the latest technological advancements made particularly in the areas of energy storage and charging systems. Together with the recent developments in terms of the emission regulations worldwide, the market share of EV is expected to increase further contrary to that of conventional vehicles with internal combustion engines (ICEs). The latest forecast conducted by [1] from Deloitte showed that the percentage of EVs in the global market share is expected to reach 32% by the year 2030. This forecast was made based on four factors, namely customers' changing sentiments regarding EVs due to their improved practicality and ownership cost, favorable government policies, mostly in terms of financial incentives and accessibility to charging facilities, car manufacturers' business strategy of putting more emphasis on EV-related technologies, and support from companies outside of the car industry in adapting EV en masse. This trend, when viewed from a tailpipe emission perspective alone, presents a positive outlook to the global environment since the amount of harmful CO, CO₂ and NO_x emissions are expected to be reduced gradually in transportation. At the same time, it also opens up possibilities to explore numerous frontiers like vehicle connectivity (vehicle-to-grid, vehicle-to-vehicle, and vehicle-to-infrastructure), autonomous technology as well as

advanced materials for energy storage. However, new challenges will also emerge from the EVs' increasing popularity and they must be studied and addressed properly.

2. New Challenges Emerged from EVs' Popularity

EVs' increasing popularity leads to numerous new challenges that must not be conveniently ignored. These challenges can be categorized into three classes, namely; challenges in ensuring the sustainability of the EV production, challenges in meeting the increasing demand of electricity due to EV penetration, and, challenges in managing the migration of ICE-to-EV in terms of number of vehicles and the industry eco-system. In the context of EV production sustainability, it was argued in [2] that, although EVs emit zero emission, the same cannot be said for their production. This is because the production process involves a significant amount of depletable materials, like heavy rare earth materials, for the production of motors and batteries. Moreover, the process also leads to higher amounts of emissions of heavy metals like lead, nickel and molybdenum, as compared to the production of ICE vehicles, and this was claimed to be detrimental to human health. According to the study by [3], the carbon footprint from these activities is currently very high due to their localization. At the moment, these activities are mainly located in China, South Korea and Japan, where a significant portion of the power is generated by fossil fuels, resulting in a high carbon footprint. To address this, refs. [3,4] proposed either diversifying the production locations to places with high concentration of renewable power generation, or intensifying the amount of renewable power generation at the existing locations. At the same time, ref. [3] also suggested stopping the trend of increasing the battery size because it has direct relationship with the aforementioned carbon footprint issue. This suggestion can be achieved by improving the efficiency of EV powertrains.

The increasing demand for EVs also causes electricity demand to shoot up and this leads to the second challenge emerged from the increasing EV popularity. According to [5], the amount of electricity used for EVs, on a daily basis, is about the same as the average daily electricity usage of a typical household in the United States. As such, when EVs reach 20% of a total vehicle market share globally, the electricity peak demand is expected to increase by 36%. In some countries, like China, research by [6] indicated that the popularity of EVs will strain not only its national grid, but also to its national water supply. This is because in China, two major contributors to power generation are hydroelectric and coal power plants that rely heavily on the national water supply. Thus, building and operating additional hydroelectric dams and coal power plants to meet the demand for EVs will divert vast amounts of water away from household usage, causing water scarcity if not properly planned. To address this challenge, two fundamental strategies must be seriously evaluated; efficient power grid management, which can be achieved via either implementation of vehicle to grid technology or implementation of extensive battery swapping activity, and efficient, sustainable and economical EV powertrains, which include the application of optimum motors, transmissions and batteries, with, possibly, a significant amount of carry-over technologies from ICE vehicles.

Finally, it is also critical to properly manage the ICE-to-EV migration so that a smooth transition phase can be realized. Simply increasing the market share of EVs alone is not enough if the total number of existing ICE vehicles, especially those that have low emission standards, is not drastically reduced. Besides, such migration must also be managed from the perspective of the existing industrial supply chain. For instance, an appropriate strategy has to be planned for the existing ICE-related manufacturing plants which are expected to face redundancy once EVs take over ICE vehicles' market share. In this aspect, one of the strategies is to repurpose the existing manufacturing plants to focus on EV-related products. This, however, is less popular due to the high costs involved in training the existing workers and upgrading the plants [7]. Market readiness is also another major challenge in ICE-to-EV migration, especially for emerging countries. To address this, one option is to implement bridging technologies, like hybrid vehicles, that implements technologies from both ICE and EV, or the use of biofuels. The advantage of the former is that it is more practical since

it also uses gasoline for operation, which is widely available especially in the emerging markets. The advantage of the latter, on the other hand, is its renewability. Nevertheless, implementing these technologies might not lead to the desirable reduction target for the carbon emissions [8,9].

One strategy that can be applied to accelerate the ICE-to-EV migration is EV powertrain retrofitting of existing ICE vehicles. The idea here is not only to accelerate the market penetration of EV, but also to utilize the existing resources; in this case, the existing ICE vehicles on the road, which leads to, ideally, no increase in the net number of vehicles on the road. A study by [10] investigated the potential as well as the challenges of widespread EV retrofitting with an emphasis on public and business perceptions. The investigation, conducted based on the current situation in Germany, highlighted some challenges in terms of public acceptance and vehicle homologations. In general, public acceptance of EV retrofitting can be improved gradually through effective communication between the government, technology providers and the public, by highlighting the benefits in terms of sustainability, long term financial savings and reduced emissions. Simultaneously, the compatibility and flexibility of EV powertrains should also be improved so that initial retrofitting cost can be reduced. Such powertrains can also contribute in the aspect of homologations, which is a major hurdle in implementing EV retrofitting.

Therefore, it can be summarized here that, an increasing EV market share, although from one angle it reduces the carbon emissions globally, still leads to several major economic and overall sustainability challenges. If these challenges are not properly addressed, they will negate the aforementioned benefits of EVs. Figure 1 shows a summary of these challenges, and based on the figure, optimizing the performance, efficiency and sustainability of EV powertrains is the key to guarantee positive economic effects and carbon neutrality in transportation.

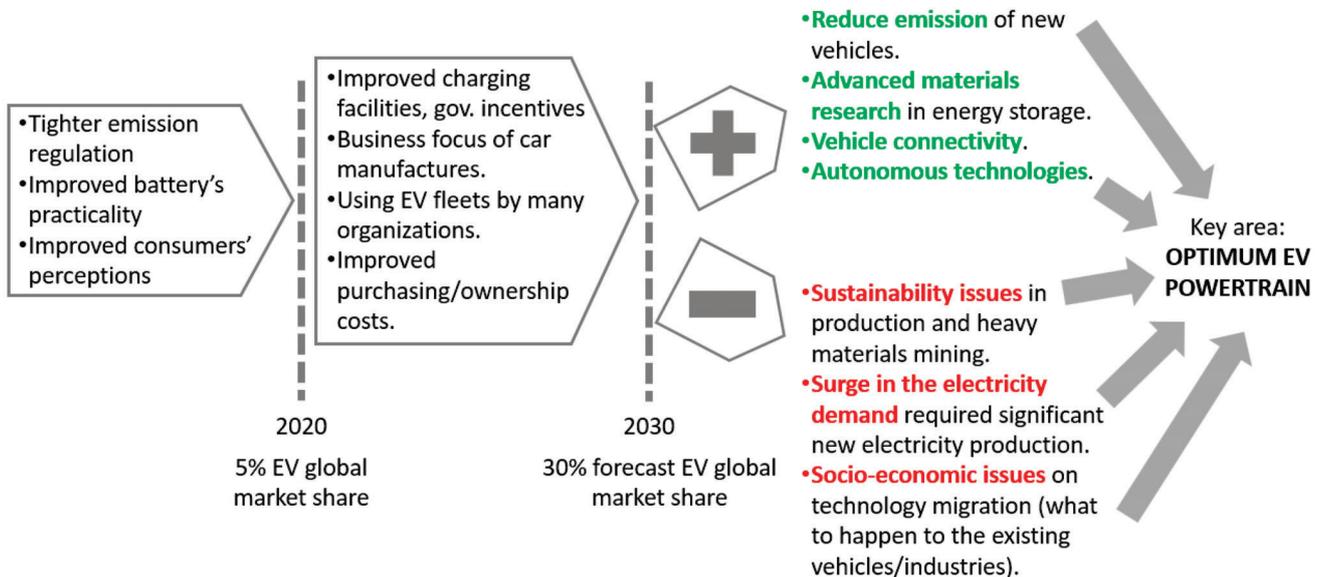


Figure 1. Summary of potentials and challenges of EV regarding environmental and other issues based on literatures in [1–10].

3. Main Components of EV Powertrains

EV powertrains mainly consist of batteries, an electric motor and transmission and their performance can be defined in terms of efficiency and practicality. A highly efficient EV powertrain means that its power consumption (kWh) per distance (km) can be kept as low as possible, thus allowing the vehicle to increase its driving mileage. For practicality, the target is to ensure that the powertrain components are cost effective; the cost for production

and operation (i.e., maintenance) can be kept as low as possible, and sustainable, i.e., with a low carbon footprint from production until application.

The purpose of the battery, the first component of an EV powertrain, is to store electricity for the electric motor's operation. To ensure that the powertrain is highly efficient and practical, the battery needs to have high energy density so that it can store high amounts of electric power without affecting its weight. Achieving this involves implementation of new cathode, anode and electrolyte materials. One of the options, suggested by [11], is to increase the nickel content in the cathode. However, this method inevitably leads to the reduction of the cathode's thermal stability, hence risking thermal runaway or damage to the battery [12]. However, according to [13], a high battery temperature, if properly managed, also presents an opportunity to enhance its performance in delivering the electricity to the motor. Because of this, many researchers have proposed either active cooling methods so that the battery's temperature can be optimized to suit various driving conditions, or emerging materials for the anode surface [14–17]. Nonetheless, such cooling methods require extra management complexity and additional power consumption for operation, while the usage of emerging materials, though promising, usually involves a significant investment for new mining and manufacturing process [18]. This is consistent with the findings by [19], which estimated that new investment of 100 Euros is required to increase the battery capacity by 1 kWh. In short, increasing the battery energy capacity, even though can avoid the increase of weight, has its own challenges in terms of safety, complexity and cost.

The next major component of an EV powertrain is the electric motor which is responsible for converting the electricity from the battery into mechanical power to move the vehicle using the electromagnetic induction principle. The motor is controlled by an inverter that regulates the required current flow from the battery to suit the driving conditions. There are two typical types of motor used in EVs: permanent magnet synchronous motors (PMSMs) and induction motors (IMs). In PMSMs, the magnetic field required to rotate the rotor is generated using permanent magnetic materials in either the stator or the rotor. On the contrary, in IMs, the electromagnetic field is produced using a current flow in the rotor conductor. Compared to ICEs, the volume of both types is relatively more compact, and, they also have a higher power to weight ratio. Even so, there are still continuous studies carried out to explore the implementation of advanced materials, like ultraconductive copper for motor windings, and grain boundary diffusion processed magnets, with the intention to increase the motors' power density even further [20]. The compactness and high-power density contribute positively to the power consumption of an EV. Not only that, but these motors also offer high torque capability at low motor speed (RPM) which eliminates the requirement of high gear ratios for vehicle start-stop. This explains the typical omission of multispeed transmissions in the existing EVs. Between these two types of motor, some researchers argued that IM ones are more robust, sustainable and low cost, partially due to the absence of a permanent magnet, while others prefer PMSMs due to their high-power density and no issue of current losses in the IMs' rotor to induce the magnetic field [21,22]. In terms of efficiency, both PMSMs and IMs have a very high peak efficiency, ranging from 85% up to 97% [23]. However, such efficiency is available only within a limited motor speed range, hence, for diverse driving conditions, the powertrain's efficiency usually falls significantly below that value. Not only that, but the construction of motor also involves the usage of heavy rare earth materials, which causes issues of high cost and less sustainable production. Therefore, sustainable and cost-effective approaches to realize the actual EV powertrains' potential in terms of driving range and performance is desired.

The final major component of EV powertrains is the transmission, responsible for ensuring that the power can be transmitted from the motor to the wheels efficiently. Because of the characteristics of the typical electric motors used in the existing EVs, the transmission used usually only provides a single speed ratio. The main benefit of using single speed transmissions is their simple construction that leads to relatively low cost for production

and maintenance. However, this limits the flexibility of the electric motor to operate optimally to suit diverse driving conditions. Therefore, it is difficult to realize the actual potential of EVs in terms of driving mileage and power consumption. A summary of the areas that can be improved to enhance the performance of an EV powertrain is illustrated in Figure 2. This figure indicates that transmission, or any method to manage power flow between the motor to the wheels optimally, is crucial in optimizing the EV powertrain performance. Once the power flow is optimized, the electric motor will have the flexibility to operate more efficiently and effectively, resulting in less power consumption from the battery. This presents a promising and cost-effective prospect of increasing EVs' driving mileage without expanding the battery size or capacity. Thus, this paper reviews and discusses the latest and most significant research works carried out to optimize the power flow in EV powertrains.

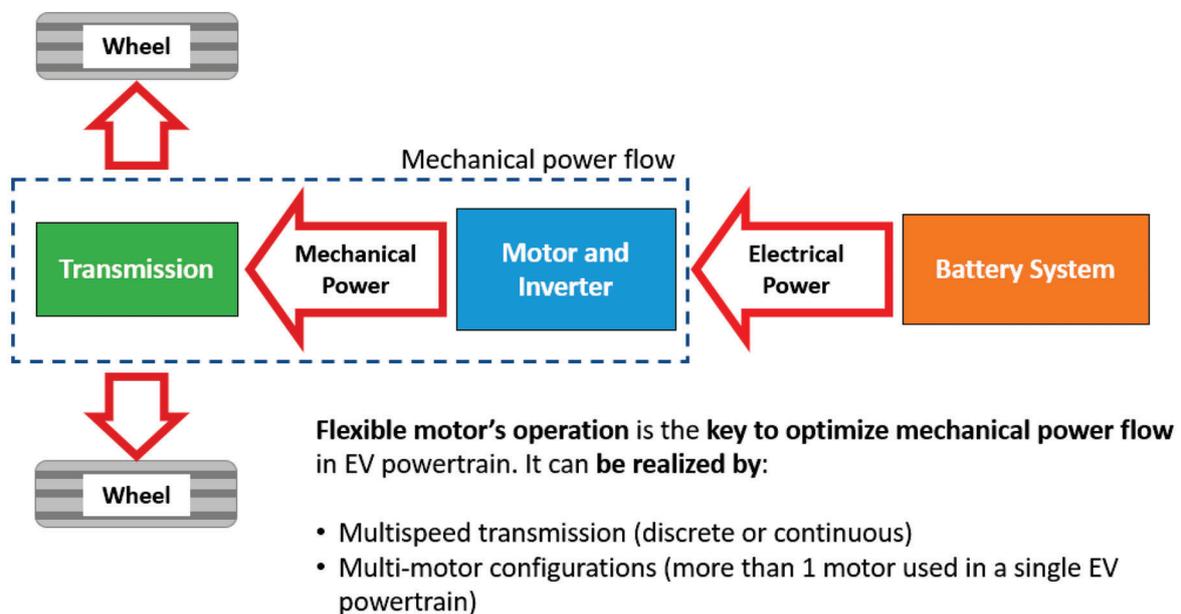


Figure 2. Various methods recently proposed to optimize EV powertrains' performance.

4. Optimizing Power Flow in EV Powertrains

Ensuring that a single motor EV powertrain can operate optimally for various driving conditions, especially when the vehicle is travelling at high speed and low load, is very challenging and because of that, their efficiency normally falls to only around 60% from about 90% for the best-case scenario [23]. One of the possibilities to avoid this is by allowing flexible power flow configurations in the powertrain. Studies [24–26] support this argument, where it is found that the powertrain efficiency and driving performance (in terms of acceleration time and comfort) can be optimized for the full EV driving experience if the driving loads can be properly distributed to two electric motors in the powertrain configuration with different transmission ratios. In the study, the possibility of implementing different hybrid EV (HEV) powertrain configurations was evaluated, and the configuration was defined in terms of coupling between the motors to the ICE, and also in terms of different transmission ratios. When the loads are properly distributed, the motors' speed can be reduced drastically during high vehicle speed, and this contributes to increasing the powertrain efficiency while ensuring the acceleration can be performed smoothly. Hence, it can be summarized here, that, flexible motor's power flow, optimized powertrain components and control are the key to optimize EV powertrains, and this can be achieved by optimizing multi-motor configurations, or by implementing multispeed transmission in the EV powertrain.

In terms of design complexity, the multispeed transmission in an EV should be less complicated than the one used in the existing ICE-powered vehicles. This is because of several factors; most notably the requirement of moving-off elements in the conventional ICE vehicles. Generally, because of the ICE idling speed condition, a moving-off element; like a dry friction clutch, or, torque converter, is required to facilitate the vehicle’s start-stop condition. For an EV, however, because of the availability of the motor’s torque from as low as 0 RPM, the implementation of moving-off elements is no longer required. On top of that, the elimination of moving-off elements also opens up the chance to implement a much simpler transmission control algorithm, since now it is no longer necessary to control the moving-off element to achieve desirable driving comfort during start-stop conditions (Figure 3). As a result, only ratio shifting control is required in an EV, although, if a discrete multispeed transmission is used, then a clutch or brake system is still required for the shifting. This is contrary to the conventional ICE vehicles, where it is absolutely critical to optimize both moving-off control and ratio shifting control. In this paper, the works related to the implementation of multi-speed transmission in EV are divided into two categories: multispeed discrete transmissions and continuously variable transmissions (CVTs). In addition, the possibilities of implementing multi-motor configurations are also reviewed here, since this approach can also lead to optimization of the motor operation for various driving conditions, which according to some scholars [25], is more effective than the implementation of multi-speed transmissions.

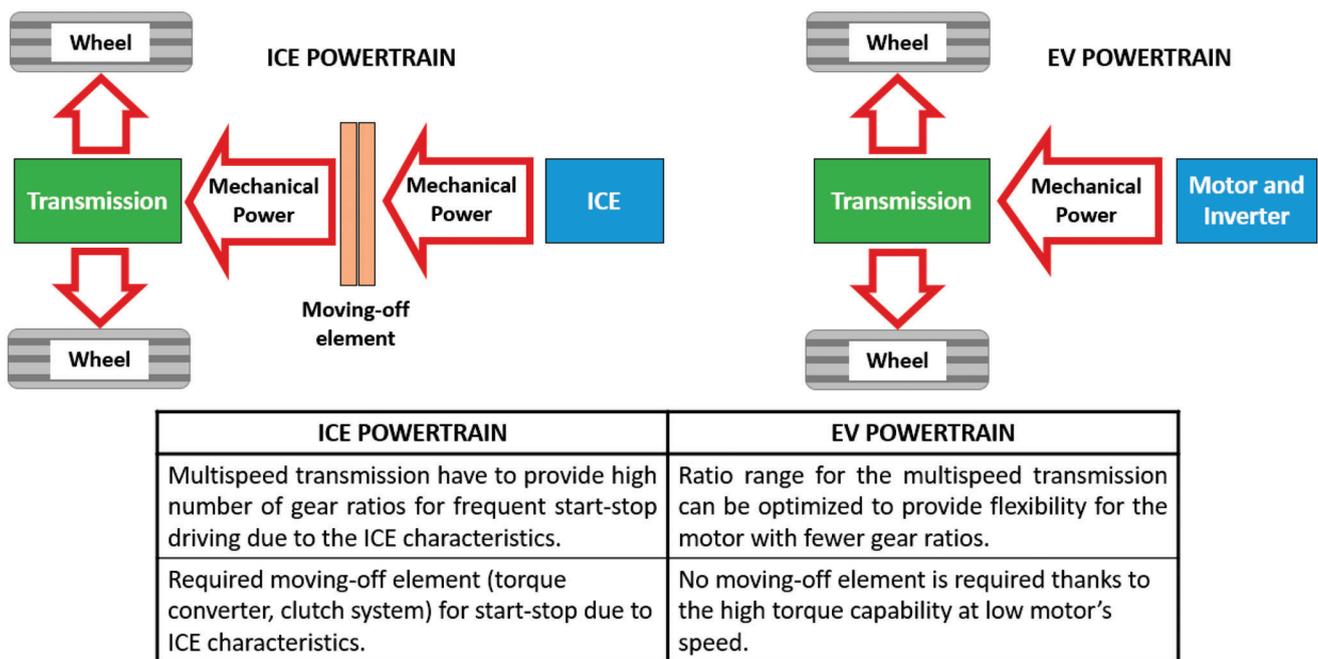


Figure 3. Differences in the powertrain requirements for ICE vehicle and EV.

This paper focuses on reviewing research works published from 2018 until early 2022. Therefore, by using keywords “multispeed transmission electric vehicle”, “continuously variable transmission electric vehicle”, “multi motors electric vehicle” and “electric vehicle powertrain” in the ScienceDirect and Scopus databases, 60 references have been identified and shortlisted as related to the topic of optimizing the power flow in EV powertrains. Among them, 24 papers describe work related to multispeed discrete transmissions, while 13 and 15 papers cover work on CVTs and multi-motor configurations, respectively. Lastly, eight papers from the 60 were review papers related to the topic of EV powertrains. The review conducted in this paper focuses on the methodologies applied and the significant findings, followed by a comparison between them. Finally, the expected key research areas

in optimizing the power flow of EV powertrain are highlighted. Figure 4 illustrates the breakdown of the selected literature reviewed in this paper.

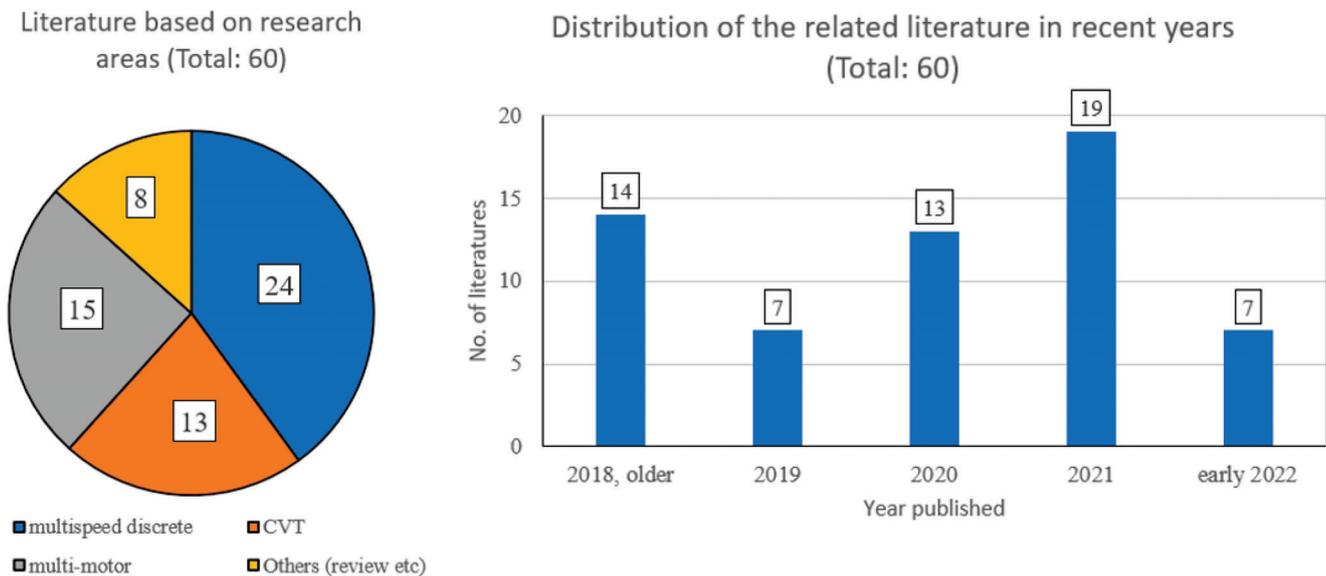


Figure 4. The number of shortlisted references related to the methods used for optimizing the power flow in an EV powertrain.

4.1. Multispeed Discrete Transmission

The main motivation of applying multispeed discrete transmission in an EV, similar to ICE vehicles, is to provide the most suitable gear ratio in the powertrain so that the motor can operate efficiently for diverse driving conditions. Since EV motors typically are capable of producing high torque output from very low RPM, the number of gears for an EV is expected to be very minimal, as few as two gears, as opposed to the ones used in the ICE vehicle. Latest works done to evaluate the efficiency difference between the EV powertrain with single-speed transmission and two-speed transmission were described in [27,28]. In [27], the comparison was conducted using simulation model-based estimation by taking into account vehicle parameters, reference motor's data and three driving cycles, namely the New European Driving Cycle (NEDC), Worldwide Light Duty Test Cycle (WLTC) and US Environmental Protection Agency (EPA) Federal Test Procedure for city driving (FTP-75). Firstly, the simulation model was run using an EV powertrain with a single speed transmission for the three driving cycles. Thereafter, the simulation results of the WLTC were used to determine the appropriate size of two gear ratios for improving the power consumption, and then the model was rerun using the newly determined ratios. Using the WLTC results, instead of NEDC and FTP-75, was logical, considering that it is the most power demanding cycle that covers diverse driving phases of urban, suburban, rural and highway scenarios. Besides, it also saves a significant amount of work and computing time as opposed to using the results from all three driving cycles. The comparison results showed that efficiency improvements were measured at the range of 1.7 to 2.4% with the two-speed transmission for the three driving cycles. It must be highlighted, however, that [27] emphasized on obtaining the gear ratios for powertrain efficiency only, without consideration of the driving performance. In terms of driving performance, the typical target is to achieve fast acceleration with minimum jerking, which leads to contradictory requirements between this and achieving maximum powertrain efficiency. Besides, the gearshifting model must also be incorporated into the powertrain simulation model to allow a realistic evaluation of the jerking. Finally, an advanced optimization method must be implemented in the model for optimizing the gear ratios for powertrain efficiency and fast acceleration with minimum jerking.

In the subsequent work by [28], the comparison was conducted using an electric bus model that runs in a specific city driving cycle with a two-speed dual clutch transmission (DCT). A DCT allows fast gearshifting thanks to its capability to pre-select the next gear before the shifting is done by the engagement of the second clutch. Such capability is not available for a single clutch automated manual transmission (AMT) and a conventional automatic transmission. The driving cycle, on the other hand, was obtained based on an existing bus route in Espoo, Finland. The powertrain model took into account not only the vehicle parameters and the motor's data, but also the efficiency mapping of the inverter. Based on the model, an exhaustive search algorithm was implemented to determine the size of the two gear ratios for optimum efficiency. The results proved that, first, the efficiency gain was in the range of 2 to 3.2%, which is consistent with the findings in [27], secondly, the application of two-speed transmission opened up the option to use a more cost-effective motor with a narrow high efficiency range, and, lastly, further studies are still required to evaluate the application particularly in terms of maintenance cost, to assess how much higher the cost will be as compared to single-speed transmission. Nevertheless, unlike [27,28] focused on maximizing the efficiency during city driving with the speed below 60 km/h and frequent starts-stops. Hence, the results are applicable only for a very specific city driving cycle. Moreover, no details on the gearshifting mechanism are provided, which means that further study to evaluate the jerking during gearshifting is required. This is particularly very important since the driving cycle studied here involves frequent start-stops. Finally, ref. [28] also considered the application of CVT with metal belt, which they found out that was less desirable due to the significant power losses in the belt. This is expected because of the high hydraulic pressure requirement to clamp the belt, especially since higher torque is required to move the bus as opposed to the passenger cars. Therefore, using such CVT in heavy vehicles, like a bus, is less practical as compared to using it in passenger cars.

Another study covering the implementation of multispeed discrete transmission in an electric bus was described in [29] where a four-speed automated manual transmission (AMT) was used. In the transmission, two DC electric motors were used for gearshifting, where one motor was used to select gear 1 and gear 2, while another motor was used to select gear 3 and gear 4. Because of this configuration, the shifting from gear 2 to gear 3 required sequential operation of both motors, thus it is expected to take longer time and higher actuation power than the shifting of gear 1 to 2 as well as gear 3 to 4. The shifting performance was evaluated experimentally on a test bench in terms of efficiency and shifting time. Based on that, a complete powertrain model for the electric bus was developed so that the optimized gearshifting strategy (defined as the optimal gearshifting points with respect to vehicle speed and throttle's opening) can be determined for minimum power consumption and minimum shifting frequency. The optimized gearshifting strategy was formulated based on the actual Beijing city driving data representing two driving conditions; high urgency driving with frequent acceleration and higher average speed, and, gentle driving with less frequent acceleration and lower average speed. Then, the four-speed AMT with the optimized gearshifting strategy was tested on a dynamometer to gauge its workability. Unlike the works in [27–29] provided the details of gearshifting mechanism in the AMT, hence analysis on the shifting time can be done realistically. However, further improvement in the shifting time here to match the DCT's performance is challenging due to the operation of two DC motors in the mechanism. This means that the shifting time can be minimized only if the number of motors can be reduced, which is possible only with the reduction of the number of gears. Therefore, the next area that can be focused on in [29] is the optimization of the gear ratios so that the possibility to reduce the number of gears can be explored. Besides, study on the jerking during gearshifting can also be carried out here thanks to the availability of the AMT's prototype.

Subsequently, research works in [30–32] described the working principle of two-speed transmissions using planetary gearset for application in an EV powertrain. In terms of the planetary gearset design, the transmission was similar to a conventional automatic

transmission for ICE vehicles. However, in terms of actuation system for its clutches and brakes, the proposed one used electro-mechanical actuation system that featured DC motor and a screw nut system. The significant benefit of using the screw nut system is that it provides self-locking mechanism, hence the desired gear can be maintained without exerting continuous hydraulic pressure on the clutches and brakes. This will improve the transmission’s efficiency since no power is required to generate the needed pressure. The challenge, however, is the complexity to integrate the design of the screw nut system with the clutches and brakes. Besides, the system also has to handle very high thrust force between the pulley (rotating based on motor’s power) and the screw (rotating only during ratio shifting to axially move the pulley). If not properly optimized, this will lead to excessive tear and wear in the screw nut system, and also power loss in the thrust bearing. In the research works conducted by [30–32], the focus was to minimize the jerking by properly implementing various gearshifting strategies with different objectives; first, to maintain a constant transmission input torque, second, to maintain a constant transmission output torque, and third, to maintain a semi-constant transmission output torque. In terms of efficiency analysis, however, no results and comparison were presented between the proposed transmission and the typical single speed transmission in an EV.

A summary of the works described in [27–31] is presented in Table 1, highlighting the potential of multispeed discrete transmissions in improving the efficiency of EV powertrains. However, these works still insufficiently discussed the topic of gear ratio optimization which is crucial to achieve not only powertrain efficiency, but also desirable driving performance. In addition, details on the gearshifting mechanism were also rarely provided, which means that analyses of the jerking and actuation power consumption during gearshifting are still lacking.

Table 1. Summary of the literature review on efficiency evaluation of using two-speed discrete transmission in an EV powertrain.

Literature, Year	Summary of the Works	Significant Findings
Hinov et al., 2021 [27]	<ul style="list-style-type: none"> - Powertrain model of a passenger car was developed and then run using single-speed transmission under three driving cycles (NEDC, WLTC and FTP-75). - Simulation results for WLTC were used to determine the size of two gear ratios for efficiency. Then, the powertrain model was rerun using the determined two gear ratios for the same driving cycles. 	<ul style="list-style-type: none"> - Improved efficiency by 1.7 to 2.4% for two-speed transmission against single-speed transmission. - Details on the shifting mechanism were not presented. - Driving performance (acceleration rate, top speed) were not considered.
Ritari et al., 2020 [28]	<ul style="list-style-type: none"> - Powertrain model of an electric bus was developed for running in a specific city driving cycle. Data for the driving cycle were obtained experimentally. - The two gear ratios were determined using exhaustive search algorithm in the powertrain model. Objective of the gear ratios was to maximize the efficiency. - Powertrain model was run using single speed transmission, two speed DCT and CVT. 	<ul style="list-style-type: none"> - Improved efficiency by 2–3.2% for two-speed DCT against single-speed transmission. - CVT suffered from belt losses, ultimately no gain in the power consumption. - Driving performance (acceleration rate, top speed) were not considered.

Table 1. Cont.

Literature, Year	Summary of the Works	Significant Findings
Lin et al., 2019 [29]	<ul style="list-style-type: none"> - Optimizing the gearshifting strategy for an electric bus using four-speed AMT. - The gear ratios were not optimized, instead they were predetermined based on literature. - Data for the driving cycle were obtained experimentally. - The actual four-speed AMT was tested on test bench for shifting time and efficiency. 	<ul style="list-style-type: none"> - Highlighting the significance contribution of optimum gearshifting strategy that can allow the motor to operate efficiently while avoiding too frequent gearshifting. - Also highlighting the gap in optimizing the gear ratios for further improvement in the powertrain efficiency.
Tian et al., 2020 and 2018 [30,31]	<ul style="list-style-type: none"> - Proposing novel two-speed transmission for EV using planetary gearset. - Gearshifting mechanism using clutches and brakes actuated by DC motors with screw and nut. - Transmission model was developed and simulated for gearshifting strategies differentiated in terms of the objectives; to maintain constant transmission input torque, to maintain constant transmission output torque, and, to maintain semi-constant transmission output torque. - Performance of the strategy was evaluated in terms of acceleration time and jerking. 	<ul style="list-style-type: none"> - Highlighting the importance of proper shifting strategy to minimize jerking during gearshifting. - Biggest jerking occurred in the shifting strategy to maintain the constant input torque, lowest jerking occurred in the other two strategies. - In the two other strategies, however, significant losses suffered in the transmission due to the slipping in the clutches and brakes.

Optimizing the best two-speed gear ratios, however, is not straightforward due to its multi-objective nature. For instance, the best ratios should be able to achieve the desirable driving performance (in terms of acceleration rate and top speed), and minimum power consumption. These objectives consistently contradict each other, and they are influenced by diverse parameters like the road gradient and instantaneous vehicle speed. Thus, advanced optimization techniques are required, for instance, a work by [33] focused on the optimization of two gear ratios based on specific gearshift scheduling strategy that took into account three parameters; vehicle speed, vehicle acceleration and road gradient. As a comparison, the usual parameter used for gearshift scheduling is the throttle position. In the work, an AMT was used and its baseline gear ratios were set at 10.00 and 5.20 for the overall gear ratio 1 (G1) and 2 (G2), respectively. For the shifting strategy, the motor speed of 3000 RPM is set as the reference for the driving due to its high efficiency in that speed, and G1 is reserved for low vehicle speed (0 to 25 km/h), and G2 is reserved for high vehicle speed (65 to 120 km/h). Between 25 to 65 km/h, the suitable ratio was selected based on the motor efficiency and power output at a particular vehicle speed, while the baseline buffer zone of 40% was set between the upshifting and the downshifting lines to avoid too frequent gear shifting. Subsequently, the baseline ratios and shifting's buffer zone were optimized using two methods: gradient descent and pattern search. Simulated under NEDC (to reflect flat road condition) and Economic Commission for Europe (ECE) Extra Urban driving cycle (to reflect gradient road), the optimized model produced a 4% and 7.5% reduction in the power consumption as compared to the baseline model, respectively. Next, the performance of the optimized model was compared against a conventional gearshift model. The conventional gearshift model consists of the same ratios as the optimized model, but it uses a conventional gearshifting strategy based on throttle. The comparison showed that the optimized model led to almost 18% energy saving over the conventional model for the gradient road driving cycle (ECE Extra Urban). However, for the flat road driving cycle (NEDC), the conventional model was slightly more efficient at about 3 to 4%. These results highlighted the contribution of two different gearshifting strategies

in optimizing powertrain efficiency for driving cycles involving diverse road gradients. Nevertheless, for the actual application of these strategies, an additional system is required to activate the suitable strategy. In this case, a gradient detection system is required so that the road gradient can be measured to activate the proposed gearshifting strategy. Therefore, further comparisons between the proposed strategy and the conventional strategy should be carried out on more driving cycles (instead of just NEDC and ECE Extra Urban) to provide clearer picture on the importance of implementing two different gearshifting strategies. Figure 5 presents a graphical summary of the work performed by [33]. In short, it can be concluded from the work, that highly flexible gear shifting strategy is crucial in optimizing EV driving mileage, and such flexibility is possible with the optimized multispeed transmission in the EV powertrain.

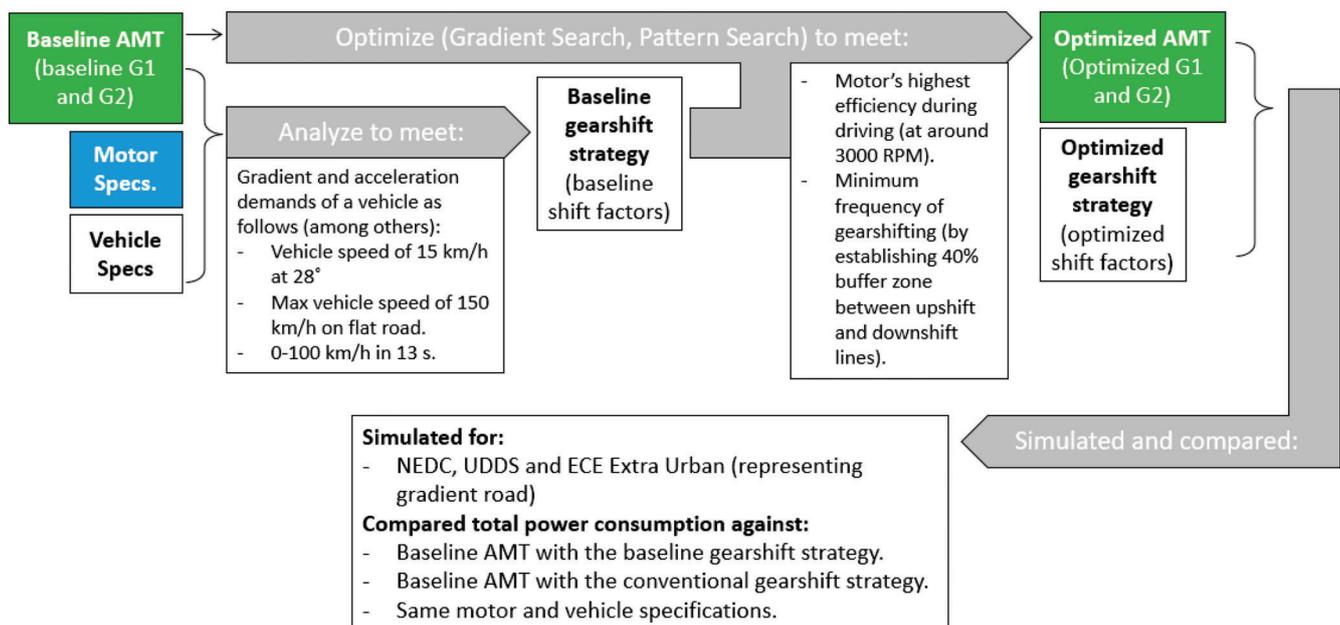


Figure 5. Summary of the work done in [33].

Other works involving the optimization of multispeed discrete transmission in EV can be found in [34–39]. However, unlike [33,34] they presented the optimization of two-speed transmission in an electric truck subjected to specific gradient route without heavy traffics, and the motor's efficiency mapping also included regenerative braking efficiency. The two-speed transmission mainly consisted of two planetary gearsets with two brakes to select the desired gear ratios. The brakes were actuated using a DC motor through a worm gear and worm wheel (Figure 6). The worm wheel was designed with an inner spiral guide, allowing it to convert its rotation about the axis of the motor's shaft into an axial movement. Depending on its rotational direction, the worm wheel, at one time, could press and lock either the first brake (engaging gear 1) or the secondary brake (engaging gear 2), accordingly. The application of worm gear here provides an advantage in terms of big torque multiplication, which leads to the possibility of using compact motor to engage the brakes. However, the worm gear is more vulnerable to tear and wear than the usual gear wheels, which means frequent gear shifting here will very likely lead to high maintenance cost. The shifting strategy used here, in contrast to [33] that took into account the road gradient, involved only throttle position and motor speed as the parameters and the driving cycle was designed to reflect an operation in an iron mine. Apart from vehicle speed, the studied driving cycle also took into account change in the weight, considering the delivery of iron ore, and also drastic change of gradient, considering the geography of the mine. Based on the aforementioned shifting strategy and driving cycle, the two ratios of the transmission were optimized for efficiency and acceleration using particle

swarm optimization (PSO). The results showed that, when compared against one-speed transmission, the optimized two-speed transmission managed to reduce the overall power consumption by 6.1%, contributed by efficiency motor's operation during driving and regenerative braking, but the gain in acceleration was very minimal. In terms of shifting strategy, for the future study, it is interesting to evaluate if there will be any efficiency improvement if the same strategy as described in [33] (consider road gradient as parameter for gearshifting) is to be implemented here in [34] (do not consider road gradient as parameter for gearshifting). The efficiency difference between them is crucial for evaluating the viability of considering the road gradient in the gearshifting strategy.

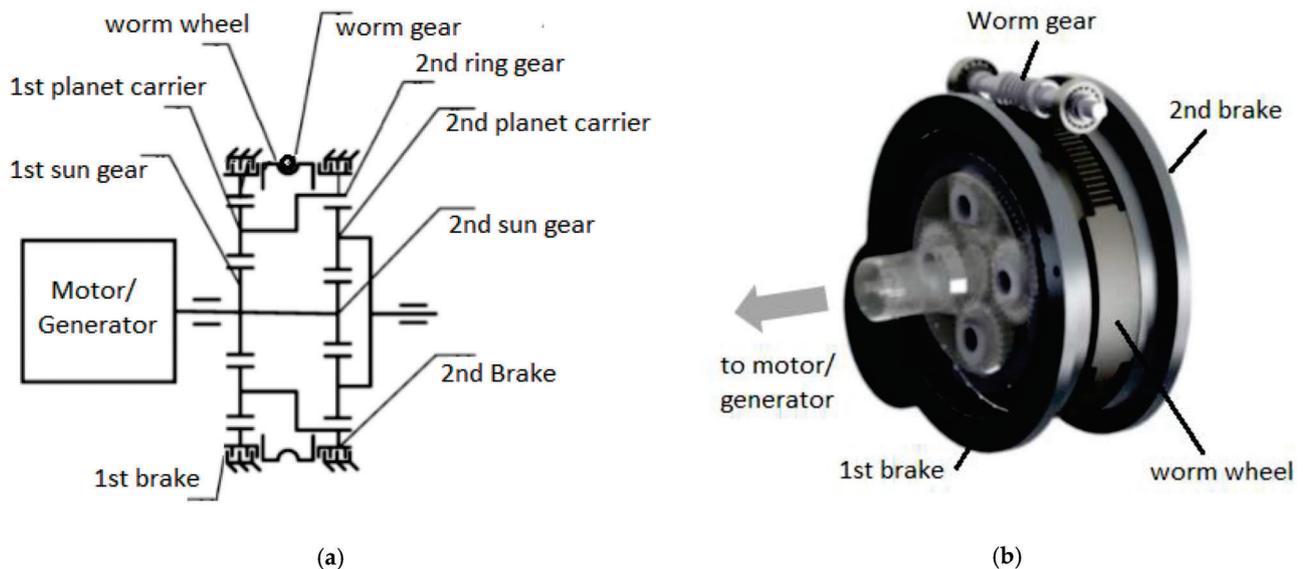


Figure 6. (a) Schematic diagram and (b) CAD model highlighting the important components of the novel two-speed transmission proposed by [34].

Another study involving regenerative braking for a two-speed transmission was described in [35], where a two-speed planetary gearset was used. Here, the main objective, instead of maximum efficiency, was to minimize jerking during the braking. The regenerative braking procedure was divided into three phases: driving phase, brake engaging phase, and braking phase. These phases were proposed to minimize torsional oscillations, that causes the jerking, by optimally synchronizing the application of hydraulic service braking and the motor's braking torque during the brake engaging phase. As a result, the jerking was reduced by around 55% as opposed to the conventional regenerative braking that does not consider such oscillations.

Next, ref. [36] presented a work carried out to optimize the gear ratios of two-speed DCT based on not only the motor's efficiency, but also the transmission efficiency. In the work, the transmission efficiency model was developed by taking into account the losses at the gear meshing, bearings, clutch and concentric shaft. Thus, different gear ratios produced different efficiency in transmitting the torque between the input and the output shafts. Based on the model, the optimum gear ratios were selected for the EV powertrain, and its performance was compared against a single-speed EV powertrain for Worldwide Harmonized Light Vehicles Test Procedure (WLTP) cycle, which showed an efficiency improvement around 10.7 to 12.1%. Regarding the transmission efficiency, research in [37] explained a possible method to improve it by modifying the tooth profile of the gears which can potentially reduce not only the loss in the gear meshing, but also the required effort for the gearshifting. In [38], on the other hand, a two-speed EV powertrain model were optimized using genetic algorithm (GA) with the objectives to achieve quick 0–100 km/h acceleration (driving performance) and minimum power consumption (efficiency) under NEDC. The type of transmission used in the model were not specifically mentioned, hence

the shifting mechanism involved was unknown. However, the model included regenerative braking efficiency model. Unlike many papers that emphasized on optimizing the size of the gear ratios, the work in [38] optimized not only the gear ratios, but also the maximum motor's output torque in Nm and its maximum rotational speed in RPM. The motor's torque and RPM were optimized within the range of 150–200 Nm and 8000–12,000 RPM, respectively. In order to obtain the balance optimization results, specific weightage was given to both of the objectives: driving performance and efficiency. The results showed that it was possible to achieve a balance (compromised) solution between the driving performance (quick acceleration) and efficiency (power consumption) by optimizing the two gear ratios and the motor's maximum torque and speed. Nevertheless, different set of gear ratios, or a continuous ratio range between 1.341 and 3.050, were required to achieve the fastest possible acceleration and highest possible powertrain efficiency. This meant that, to achieve maximum performance and efficiency in a single powertrain system, the gear numbers must be higher than 2. In a discrete transmission, however, increasing the gear number must be done together with redesigning the gearshifting mechanism which leads to increased design complexity and cost. Figure 7 shows summary of the optimization work done in [38] using GA to determine the optimum gear ratios and motor's outputs for driving objectives. Based on the results, if a continuous ratio range between 1.341 to 3.050 can be provided by one transmission (like a continuously variable transmission), then all the objectives can be met, instead of opting for a compromised two gear ratios in a two-speed transmission.

Another interesting work regarding EV powertrain with discrete gear transmission was presented in [40], which experimentally evaluated the performance of three different transmission ratios—6.00, 8.00 and 10.00—for one driving cycle. Among the ratios, 8.00 served as the benchmark for the results' analysis. In the experiment, the test vehicle was tested on the same track using three different one-speed transmissions, corresponding to the three ratios. The results showed that, with the gear ratio of 10.00, the power consumption was higher by 4.2% as compared against the benchmark. The authors argued that the increase was caused by the possibility of fast acceleration provided by the ratio, hence the driver has the tendency to often accelerate suddenly. On the other hand, the power consumption can be reduced by 2.4% when the ratio of 6.00 was used, since it was claimed that with that ratio, the driver was forced to drive with using gentler acceleration. From the work, three important conclusions can be drawn. Firstly, driving style is critical for the power consumption of an EV with one-speed transmission, thus, encouraging drivers to drive economically plays an important role in increasing EVs' efficiency. Secondly, the size of the gear ratio has some influence on a person's driving style which ultimately affects the driving power consumption, and thirdly, multispeed transmissions can offer flexibility to suit drivers' driving preferences, which means that the EV can then be driven either to achieve maximum efficiency, or with an aggressive driving style.

A summary of the works presented in [33–40] is provided in Table 2 and they highlighted the significance of optimizing the gear ratios and the gearshifting strategy to achieve powertrain efficiency and driving performance. Some of these works have started to discuss on the gearshifting mechanism, but analysis on jerking and actuation power usage during shifting was still limited. Moreover, since some of the gearshifting mechanisms are novel, new study areas concerning their durability and practicality must also be covered in the future.

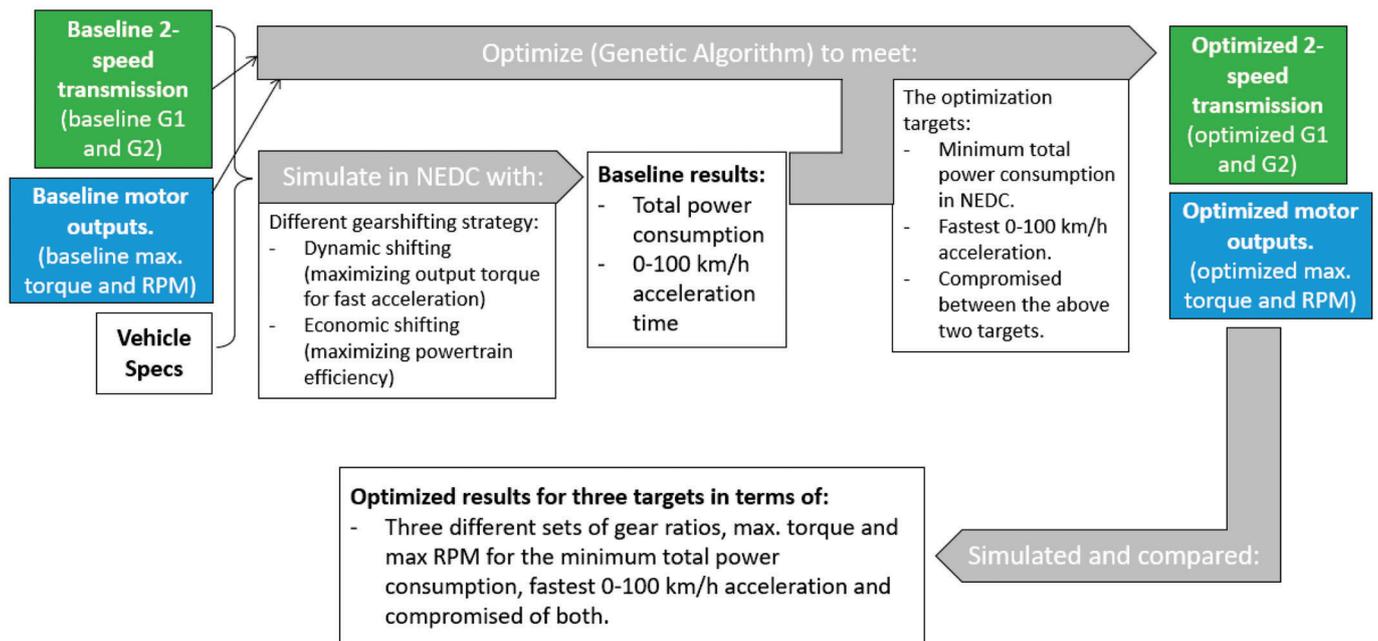


Figure 7. Summary of the work done in [38].

Table 2. Summary of the literature review on optimizing two-speed transmission for efficiency and driving performance in an EV powertrain.

Literature, Year	Summary of the Works	Significant Findings
Ahssan et al., 2020 [33]	<ul style="list-style-type: none"> - Optimizing gear ratios of two-speed AMT for efficiency and driving based on a proposed shifting strategy in a passenger EV. - The proposed shifting strategy considered vehicle speed, acceleration rate and road gradient. Traditional shifting strategy throttle opening only. - Baseline for the two gear ratios was predetermined at 10.00 (Gear 1) and 5.20 (Gear 2) based on literature. Then, the two ratios were optimized using gradient descent and pattern search. - Using the optimized gear ratios and the proposed shifting strategy, a powertrain model was simulated under ECE Extra Urban cycle (for gradient road) and NEDC (for flat road). Then, the same model was simulated using the optimized gear ratios and the traditional shifting strategy for comparison. 	<ul style="list-style-type: none"> - Against the baseline gear ratios, the optimized gear ratios yielded 4% and 7.5% reduction in power consumption for the gradient road and the flat road cycles, respectively. - Comparison between the two shifting strategies (both using the optimized gear ratios) showed 18% reduction and 4% increase in power consumption for the gradient road and the flat road cycles, respectively. - Showed that besides optimum gear ratios, shifting strategy must also be formulated based on the road profiles for optimum powertrain efficiency. - Shifting mechanism was not explained in details.

Table 2. Cont.

Literature, Year	Summary of the Works	Significant Findings
Tan et al., 2018 [34]	<ul style="list-style-type: none"> - Optimizing gear ratios of a novel two-speed transmission for efficiency and driving performance. The transmission featured planetary gearsets and brakes as its shifting mechanism. The brakes were actuated by worm gears. - The transmission was used in an EV truck powertrain model for a specific hilly road cycle with minimum traffics. The model also considered significant change in vehicle weight (load/unload raw materials). - Traditional shifting strategy was employed (based on throttle opening). - Gear ratios were optimized using PSO, and then its performance was compared against a single-speed transmission. 	<ul style="list-style-type: none"> - The optimized two-speed transmission reduced the total power consumption by 6.1% against the single-speed transmission. - Acceleration for both transmissions was about the same. - Details of the shifting mechanism was included. However, controller for the mechanism was not studied.
Kwon et al., 2021 [36]	<ul style="list-style-type: none"> - Optimizing gear ratios of a two-speed DCT by taking into account the motor's efficiency and transmission efficiency model. The model provided precise transmission efficiency for different gear ratios (usually assumed to be a constant). - The optimized two-speed DCT was simulated in an EV powertrain model for WLTP cycle. Then, it was compared against a single-speed transmission powertrain model for the same driving cycle. 	<ul style="list-style-type: none"> - Efficiency improvement between 10.7% to 12.1%. In terms of transmission efficiency, work in [37] explained an option by optimizing the gear tooth profile. - Acceleration was about the same for both transmissions. - Details of the shifting mechanism was not provided.
Li et al., 2020 [38]	<ul style="list-style-type: none"> - Optimizing two-speed transmission for fast 0–100 km/h acceleration and high efficiency using GA. - Apart from gear ratios, motor's parameters (max. torque and RPM) were also optimized within a specified range of 150–200 Nm and 8000–12,000 RPM. - Compromised solutions between the acceleration and efficiency were also determined based on weightage. - Powertrain model include regenerative braking efficiency. Losses during regenerative braking can be minimized by vibration suppression method discussed in [35]. 	<ul style="list-style-type: none"> - Different gear ratios required to meet different objectives (fastest acceleration, lowest power consumption, compromised based on weightage) as shown in Figure 7. - Highlighting the limitation of two-speed discrete transmission, potential of using CVT which allows continuous ratio range, hence possible to meet the different objectives.

Table 2. Cont.

Literature, Year	Summary of the Works	Significant Findings
Han et al., 2019 [39]	<ul style="list-style-type: none"> - Optimizing the number of gears (between 2, 3 and 4 gear ratios and continuous) and gear ratios using dynamic programming based on a specific vehicle powertrain system with battery and motor's efficiency models - Simulated under four different driving cycles namely Urban Dynamometer Driving Schedule, (UDDS), NEDC, Highway Fuel Economy Test (HWFET) and LA92. - Based on the optimized data, the optimum gearshifting strategy was determined. 	<ul style="list-style-type: none"> - Increasing the number of gears led to improvement in power consumption. - Continuous ratio provided the best results in power consumption, hence indicating the potentials of CVT. - However, increasing the number of discrete gears usually led to more complex shifting mechanism, which may negate the gain in power consumption.
Spanoudakis et al., 2020 [40]	<ul style="list-style-type: none"> - Experimentally evaluating three different gear ratios; 0.167, 0.125 (benchmark) and 0.100, on a specific driving cycle. - The tested was conducted on a small EV on a test track. Every gear ratio tested represented three different transmissions used in the EV. 	<ul style="list-style-type: none"> - Highlighting the driving tendency of the driver with different gear ratios. For instance, it was argued that the gear ratio 0.100 allowed for faster acceleration, hence the driver tend to accelerate more aggressively, resulting in 4.2% more power consumed. - Highlight the contributions of multispeed gear ratios in providing flexibility for driving performance and efficiency.

The latest works related to gearshifting mechanism and its control are described in [41–50]. Researchers in [41–43] argued that criteria for the EV motor to operate efficiently is not just the application of the multispeed discrete transmission but also smooth gearshifting process with minimum jerking and actuation power usage. Reference [44], on the other hand, explained jerking effects in relation to friction clutch, one-way clutch and types of the driveline. The jerking effects were evaluated under three common shifting scenarios: upshifting during driving, downshifting during driving, and downshifting during braking. In general, smooth shifting is not only beneficial for driving comfort, where it avoids excessive jerking and torque interruption, but it also helps in terms of the overall powertrain efficiency. Thus, a novel clutchless AMT was proposed in [41–43] featuring a unique synchronizer called bilateral Harpoon-shift synchronizer. Such a synchronizer uses a torque spring, constructed based on multiple bended coil springs, inside the dog body's internal groove to keep the dog gear damped to the guide ring (Figure 8). This results in quick synchronization of the guide ring and the dog gear without using frictional cones, and also smooth shifting due to the spring's damping effect. Additionally, the spring also helps in reducing the required axial force for shifting; minimizing the required DC motor's work to actuate the fork for shifting. Hence, faster and more efficient shifting process can be done with a compact DC motor. However, the spring also causes additional normal force between the guide ring and the sleeve, and this causes friction force between them during the shifting process. This eventually leads to an extra load that must be overcome by the motor.

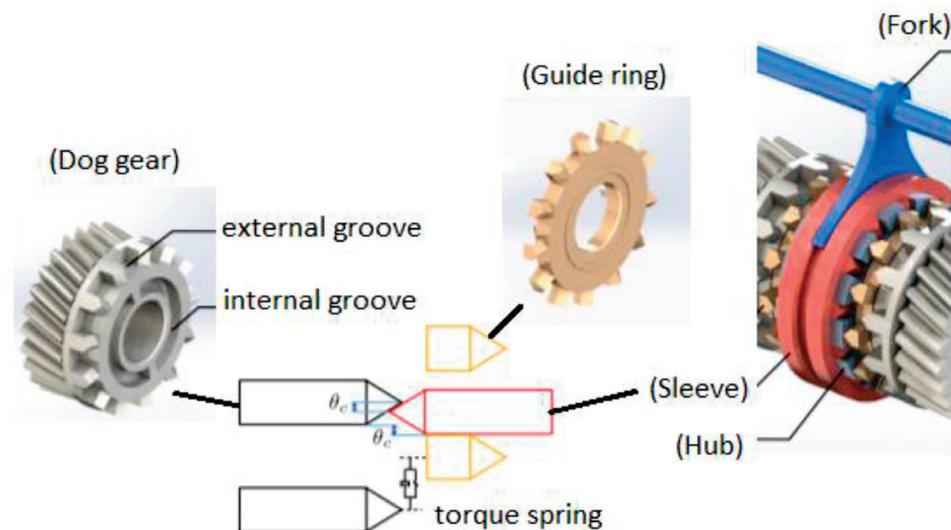


Figure 8. The proposed bilateral Harpoon-shift synchronizer by [41–43].

Another novel synchronizer design that featured springs inside it was also presented in [45]. Meanwhile, the work in [46] reported the optimization of the gearshifting with the objectives to minimize shifting time, friction work, due to the engagement and disengagement of the clutch during shifting, and jerking. The optimization was conducted using the Legendre pseudospectral method and the gearshifting model was simplified as two degree of freedom (2-DOF) and 4-DOF dynamic models based on a friction clutch and a sleeve shifting mechanism in a two-speed transmission. The results were divided into four different patterns: the least shifting time, the least friction work, the least jerking, and finally, the compromised solution. In the compromise solution, obtained in the 2-DOF model simulation, the shifting time was recorded at 0.92 s, with the square of continuous jerking measured at $0.48 \text{ (m/s}^3\text{)}^2$, and the friction work at 1856 J. In the work, however, no detailed descriptions of the actual actuators used for the clutch and the gearshift sleeve was provided, which can be the focus for future works.

Subsequent research work performed to analyze the performance of shifting mechanism was described in [47,48], where a two-speed dry clutch inverse AMT (I-AMT) was used to vary the gear ratios with very minimal torque interruptions with help from two one-way clutches. The one-way clutches, on the other hand, were integrated into the first gear and the second gear separately, hence the shifting can be achieved by actuating only a single dry clutch. Prior to that works, another study has been carried out, as described in [50], to evaluate the clutch control of a wet dual clutch two-speed transmission for EV application. The objective of the study was to experimentally quantify the clutch control's performance in terms of jerking and engagement time. However, because of the usage of the wet clutch, some power was lost due to the clutch actuation. Not only that, but the gearshift quality was also less desirable due to the high jerking at around 10 m/s^3 , signifying noticeable torque interruption during the shifting. Besides, the sticking characteristics of the wet clutch, due to its hydraulic system, caused difficulty to optimize the clutch control for minimum jerking and engagement time. For improvement, other type of clutch, like a dry clutch system with electro-mechanical actuator, can be implemented, so that the clutching and gearshifting can be precisely controlled based on the motor's torque to minimize jerking and shifting time. Returning to the work explained in [47,48], a dry clutch was used, and its slip control was optimized using high-order disturbance observer to minimize jerking and shifting time, and the clutch control was then tested experimentally in a small EV during upshifting and downshifting. The dry clutch was actuated by a DC motor. The results were encouraging, with the jerking measured at most around 3 m/s^3 , which is significantly lower than the widely accepted threshold of 10 m/s^3 . Nevertheless, the operation of I-AMT involved frequent slipping in the dry clutch,

hence its durability is expected to be compromised. This leads to the possible increase in the maintenance cost against a simpler single-speed transmission EV powertrain. Thus, detailed study in this aspect is crucial to quantify its long-term operation. Further studies on optimizing the gearshifting mechanism were explained in [49], in which the application of torque sensor was proposed in a two-speed DCT so that precise clutch engagement force can be regulated to fit the desired clutch torque for optimum shift quality. The torque sensor allowed precise real time torque measurement which is crucial to regulate the clutch engagement for optimum engagement time with minimum jerking. However, the application of torque sensor required significant cost, which will increase the transmission's cost tremendously. This unfortunately makes the option of implementing the torque sensor in the actual transmission impractical.

More advanced studies related to EV powertrain with two-speed discrete transmission focused on the shifting strategy that adapts driver behaviors are described in [51]. Previous studies on the methods to recognize driver behaviors can be read in [52–54], covering its application in a fuel cell vehicle and HEV, but none of them was conducted specifically for EV powertrains with two-speed transmissions. However, all of the literatures have certain similarities, in the sense that, the throttle opening rate was used as the indicator for the behavior, and then fuzzy logic was applied to predict the suitable driving style corresponding to the modified standard driving cycles based on driving aggressiveness. Subsequently, the baseline driving style (usually established based on conventional practice) was optimized in the literatures by embedding the correcting factor representing the fuzzy logic's output. Therefore, in [51], they also proposed a predictive model based on a fuzzy neural network (FNN) to recognize the driver's intention via the actual throttle opening rate. Simultaneously, the learning vector quantization neural network (LVQNN) method was used to select the appropriate driving cycle by comparing the actual vehicle speed data against three predetermined different driving cycles. These predetermined cycles were obtained offline based on samples generated from the driving cycles of New York City Cycle (NYCC) UDDS and HWFET. Finally, a correcting factor, representing the outputs from FNN and LVQNN, was introduced to the baseline shifting strategy to optimize it for efficiency and driving performance. The baseline shifting strategy was formulated by taking into account the motor's efficiency, throttle opening and battery's SOC at 40% and 70%. Comparison between the baseline shifting strategy with and without the correcting factor, through simulation and dynamometer testing, showed an average efficiency improvement of up to 2%, proving the benefits of adapting driver behaviors in the shifting strategy.

A summary of the works related to the gearshifting mechanisms and the adaption of driver behaviors in [41–51] is presented in Table 3. In terms of gearshifting mechanisms, the works reviewed here mostly discussed their standalone performance in terms of jerking and shifting time, while limited discussions were carried out to evaluate their performance when integrated in a powertrain system. This means that the question on the potential improvement efficiency and driving performance in a complete powertrain system is still not properly answered. Nevertheless, with the numerous novel designs of gearshifting mechanism proposed recently by the researchers, the outlook of developing and implementing multispeed discrete transmission, especially two-speed transmission, in a commercialized EV looks promising.

Table 3. Summary of the literature review on gearshifting mechanism of multispeed discrete transmission in EV powertrain.

Literature, Year	Summary of the Works	Significant Findings
<p>Mo et al., 2020, 2019, and 2018 [41–43] and Beaudoin and Boulet, 2021 [44]</p>	<ul style="list-style-type: none"> - Discussing the importance of optimum gearshifting process for efficiency and driving comfort in EV. - In [44], analysis and comparison have been carried out between different shifting mechanisms (friction clutch and one-way clutch, among others) for efficiency and jerking performance. There were analyzed for three conditions; upshifting during driving, downshifting during driving and downshifting during braking. - In [41–43], novel shifting mechanism named bilateral Harpoon-shift synchronizer was proposed (Figure 8), eliminated the usage of clutch in the AMT. - Using torsional spring (bended compression spring) inside the dog gear internal groove to provide spring force to the guide ring. Other example of similar design was described in [45]. 	<ul style="list-style-type: none"> - In [44], results showed the importance of selecting the right shifting mechanism to minimize the jerking so that maximum efficiency can be achieved. - In [41–43], the proposed Harpoon-shift synchronizer minimized the required shifting effort thanks to the spring force on the guide ring. - It also provided damping effect between the dog gear and the sleeve, hence the jerking can be minimized during the shifting.
<p>Liu et al., 2020 [46]</p>	<ul style="list-style-type: none"> - Optimizing gearshifting to minimize shifting time, friction work and jerking using Legendre pseudospectral method. - Gearshifting model was simplified as 2-DOF and 4-DOF dynamic models based on a friction clutch and a sleeve shifting mechanism in a two-speed transmission. 	<ul style="list-style-type: none"> - In the compromised solution (balanced between all the objectives), the shifting time was recorded at 0.92 s, with the square of continuous jerking measured at $0.48 \text{ (m/s}^3\text{)}^2$, and the friction work at 1856 J. - For now, no details descriptions on the actual actuators used for the clutch and the gearshift sleeve.
<p>Walker et al., 2017 [50]</p>	<ul style="list-style-type: none"> - Evaluating wet clutch control on a dynamometer test rig in terms of shifting time and jerking. - Based on a two-speed DCT applied in an EV powertrain. 	<ul style="list-style-type: none"> - Some power was lost in the wet clutch actuation. - High jerking at around 10 m/s^3, signifying noticeable torque interruption during the shifting. - Difficulty to optimize the shifting due to the sticking characteristic of the wet clutch. - Highlighting the importance of precise clutch control in shifting, can be potentially done by an electro-mechanical actuator.

Table 3. Cont.

Literature, Year	Summary of the Works	Significant Findings
Hong et al., 2021 [47] and Yue et al., [48]	<ul style="list-style-type: none"> - Based on two-speed dry clutch I-AMT. The gearshifting mechanism consisted two one-way clutches on the first and second gears and a dry clutch with actuator. - The one-way clutches allowed shifting to be done by only using one actuator at the dry clutch. - Clutch slip control was optimized using high-order disturbance observer to minimize jerking and shifting time. - The study was conducted experimentally on a small EV on dynamometer. 	<ul style="list-style-type: none"> - Jerking measured at around 3 m/s^3, well lower than the widely accepted threshold of 10 m/s^3. - Nevertheless, it involved frequent slipping in the dry clutch, hence its durability is expected to be compromised. - Further study in this aspect is crucial to quantify its significance.
Ogawa et al., 2021 [49]	<ul style="list-style-type: none"> - Proposed the implementation of torque sensor at the input of the 1st gear of a two-speed DCT to precisely control the torque transmitted. - Simulation analysis was conducted using mathematical equations describing the torque transmission on the clutch and the gears. 	<ul style="list-style-type: none"> - Possible to precisely control the clutch to achieve the appropriate output torque of the clutch (input torque at the 1st gear). - However, costly for actual implementation.
Lin et al., 2021 [51]	<ul style="list-style-type: none"> - Optimizing shifting strategy by incorporating driver's behaviors. First, a baseline shifting strategy was formulated by taking into account the motor's efficiency, throttle opening and battery's SOC at 0.4 and 0.7. - Next, a predictive model based on FNN is applied to recognize the driver's intention via actual throttle opening rate. - LVQNN method was used to select the appropriate driving cycle by comparing the actual vehicle speed data against three predetermined different driving cycles namely NYCC, UDDS and HWFET. - Output from the FNN and LVQNN would be in the form of correcting factor, then be included in the baseline shifting strategy to optimize it for efficiency and driving performance. 	<ul style="list-style-type: none"> - Comparison between the baseline shifting strategy with and without the correcting factor, through simulation and dynamometer testing, showed an average efficiency improvement of up to 2%, proving the benefits of adapting driver behaviors in the shifting strategy.

Overall, it can be summarized that the research works on multispeed discrete transmission for EV mainly focused on the implementation of two-speed discrete transmission which can be in the form of AMT or DCT. The two-speed design is very compact, which means the additional weight relative to the usual single-speed transmission can be kept to a minimum. Besides, the two gear ratios provide the necessary flexibility in the EV driving modes' selection for optimum efficiency and driving performance. The main challenge, however, is how to optimize the gear ratios and the shifting strategy so that the gains in powertrain efficiency and driving comfort can be maximized. Several optimization methods have been implemented to optimize the gear ratios and the shifting strategy, and the results highlighted the capabilities of the two-speed transmission to reduce the power consumption by up to 16% for some driving cycles. However, more work is still required to evaluate the operation and controls of the gearshifting mechanisms in a complete powertrain in terms of efficiency and driving performance. For now, the studies on gearshifting mechanism mostly focused on assessing its jerking and shifting time, with very limited discussion to answer question on its contribution to the overall powertrain's efficiency.

4.2. Continuously Variable Transmission (CVT)

The main motivation of utilizing CVT in EV, identical to the multispeed discrete transmission, is to provide variable transmission ratios so that the motor can operate optimally for diverse driving conditions. In general, there are many types of CVT available for automotive application, but in this review paper, the focus will be on CVT that uses pulleys and metal belt which is the most widely used type currently in automotive. Unlike multispeed discrete transmissions, CVTs with metal belts offer a continuous ratio range, which mean more ratios are available to be chosen to suit the driving conditions. In this sense, a CVT is more flexible than any multispeed discrete transmission, hence, the motor has a much better chance to operate optimally for a longer duration of the driving. However, this type of CVT has certain limitations in terms of power loss in the metal pushing V-belt, or metal chain, used to transmit the torque between the primary pulley; connected to the motor, and the secondary pulley; connected to the vehicle's wheels. The loss is caused by a portion of motor power consumed to produce the required high clamping force to clamp the belt for the torque transmission between the pulleys. Research in [55] discussed the possibility of controlling the appropriate CVT ratio using fuzzy logic based on the motor's efficiency mapping as the reference. The fuzzy control algorithm was tested in a simulation model developed based on three driving routes differentiated in terms of road gradients. While the controller helped in enhancing the motor's efficiency throughout the routes, more detail studied are still required particularly for the ratio and clamping force actuation system of the CVT which was not explained in the paper. Subsequently, in 2017, ref. [56] suggested that a CVT, with a possibility to clamp the belt using an electro-mechanical actuation system with self-lock capability, has the potential to increase the powertrain efficiency. They explained that, unlike conventional CVT that uses engine power to generate hydraulic pressure to clamp its metal belt, such CVT eliminates the required power consumption for the clamping since the self-lock mechanism can held the clamping force. Thus, more power can be transmitted to the wheels, and its ratio can also be selected more efficiently. For an EV, this is particularly beneficial since the motor can operate with high flexibility, resulting in improved powertrain efficiency and increased driving mileage. However, to incorporate the self-locking mechanism required extensive design modifications on the CVT's pulleys as well as integration of the DC motor to actuate the mechanism accordingly.

Next, refs. [57,58] reported their research involving an evaluation of four different DCTs and a CVT applied in an EV. The four DCTs were differentiated in terms of the number of gear ratios (from single to four-speed), and the size of the gear ratios were determined based on the gradient climbing requirement (first gear), high speed driving (top gear) and the progression factor for the intermediate gears. Thus, the gear sizes, as well as the gearshifting strategy, were not optimized based on suiting any driving cycle. The ratio range of the CVT, on the other hand, was defined based on the continuous ratio range between the first and the top gears of the DCTs. In addition, the CVT was also considered to use electro-mechanical actuation system, instead of the conventional hydraulic system, to vary its ratio and to clamp the belt. Therefore, its efficiency was considered to be significantly higher as compared to the existing CVT used in any ICE vehicle. Also, the manufacturing cost for the CVT was considered to be lower than the two-speed transmission in the research. All the transmissions were then simulated based on a hybrid driving cycle established by combining FTP-75 and HWFET. The results showed that, CVT was the best performer in terms of efficiency for a B-segment car, reduced the power consumption by 31.9% against the single-speed transmission, followed by the three-speed DCT (19.1%), four-speed DCT (18.2%) and the two-speed DCT (16.4%). This result highlighted the magnitude of improvement that can be gained by eliminating the losses in the hydraulic actuation system conventionally used in a CVT with metal belt. Moreover, it also emphasized the saturation in the increment of the gear numbers in a multispeed discrete transmission, which in this case, it can be observed by the reduction of the efficiency improvement between the three-speed DCT and the four-speed DCT. This

means that the saturation point here is at three gears, and further increases in the gear numbers will only cause significant actuation losses in the additional shifting mechanism added for the extra gear ratios. The rather low saturation point is typical for a small car (i.e., B-segment) due to its narrower range of the power required as opposed to a bigger car (i.e., E-segment). For a E-segment car, the results showed that CVT was the best performer (23.6%), followed by the four-speed DCT (15.2%), three-speed DCT (9.0%) and two-speed DCT (9.6%). This result suggested that the saturation point for E-segment EV could be higher than four gears for a multispeed discrete transmission, which is logical considering its wider range of power required as opposed to B-segment car. Based on these results, CVT seems more promising, provided that a reliable electro-mechanical actuation system can be successfully integrated in its pulleys system. To achieve this, further research works are still required, especially in the areas of the workability and durability of the electro-mechanical actuation system in the CVT, since such actuation system is still relatively new and has not been implemented previously in any commercialized CVT with metal belt.

Other works discussing the application of CVT in EV powertrain were presented in [59,60]. In [59], the potential of CVT's continuous ratio range to improved EV power consumption was assessed against the single-speed, two-speed AMT and two-speed DCT EV powertrains. In the assessment, all types of transmissions were considered to have the same constant efficiency of 97%. The assessment was conducted based on an analytical model of the motor's efficiency, and it showed that with CVT, the powertrain efficiency can be improved by about 3% for WLTP cycle against the other discrete multispeed transmissions. However, more detail analysis, particularly on the CVT's efficiency, is required, because the application of CVT conventionally involves high hydraulic pressure for clamping and ratio shifting. Thus, without optimization on the hydraulic actuation system, it is inappropriate to assume that the CVT has the same efficiency as the discrete multispeed transmission. Subsequently, in [60], the CVT was considered to be using the optimized electro-hydraulic actuation system (more compact and requiring low power for generating the belt's clamping force) and the novel single loopset belt (as opposed to the typical metal pushing V-belt, hence more compact design and reduced power losses). On top of that, the possibility of downsizing of the motor was also studied, where it was achieved through the reduction of the rotor's diameter and inertia. Nevertheless, the work did not take into account the optimization of the transmission ratios and the shifting strategy. When the powertrain was simulated under WLTC, it showed a 12.7% efficiency improvement against the EV powertrain with single speed transmission. That, however, was less than the two-speed AMT that produced a 13.5% improvement. The lower improvement was very likely caused by the hydraulic actuation system. Even though the system was optimized, the required belt's clamping force was still very high (depending on the EV motor's torque) and must be provided continuously during operation. Hence, continuous power to generate the clamping force, albeit lower thanks to the optimization, was still needed. The AMT, on the contrary, used geartrain to transmit the power, hence no requirement for the belt's clamping force. This means that the continuous power for the clamping force was eliminated entirely. This situation also affected the power flow in the powertrain, which can compromise the driving performance and this can be observed in the 0–100 km/h acceleration time, where the AMT and the single speed transmission yielded 6.9 s, while the CVT achieved 7.4 s. Based on the results, it appeared that AMT is the better transmission for EV than the CVT, although it must be noted that with the latter, it is possible to eliminate torque interruption during ratio shifting.

Research work described in [61] explained optimization and discretization of the CVT ratios so that optimum power consumption can be realized with as minimum shifting as possible. The CVT featured an electro-hydraulic actuation system, where an electric pump was used to precisely control the required hydraulic pressure for clamping and ratio shifting (Figure 9). The rationale of discretizing the ratios was to avoid too frequent shifting would lead to uncomfortable driving due to jerking, as well as power losses in the hydraulic actuation system. The discretizing process started by first establishing the

appropriate number of ratios based on the relation between the energy cost and the ratio number. Hence, the number of ratios was set at four, and the ratio sizes were determined through an equal ratio series method. Then, these ratios were optimized using GA for optimum efficiency when undergoing a combined driving cycle that comprised of UDDS, NYCC and HWFET. For the driving cycle simulation, three ratio shifting strategy models were employed; first, continuously variable ratio shifting strategy, where the best ratio was selected continuously during driving for maximum efficiency, second, the discrete ratio shifting strategy based on the ratios established through the equal ratio series method, and third, the discrete ratio shifting strategy based on the ratios optimized through GA. Comparison between the results confirmed that the third strategy performed the best, with the minimum total power consumption and average jerking, measured at 8.10 kWh and 4.32 m/s^3 , respectively. The first and second strategies, meanwhile, recorded 8.16 kWh and 5.35 m/s^3 , as well as 8.69 kWh and 4.65 m/s^3 , respectively. To summarize, the work reported in [61] highlighted two very important findings. Firstly, CVT provides a continuous ratio range, hence the ratios can be discretized and optimized to suit diverse vehicle parameters, which means the same CVT can be implemented for several type of EVs for optimum driving performance and efficiency. Secondly, high ratio number presents better flexibility for motor's operation, but it also leads to complicated shifting logic which will cause too frequent shifting, resulting in the power losses in the actuation and compromised driving comfort. This presented an opportunity to apply the same CVT with different sets of discretized ratios to suit the requirements of diverse EV segments, which can contribute in terms of cost reduction in the transmission production.

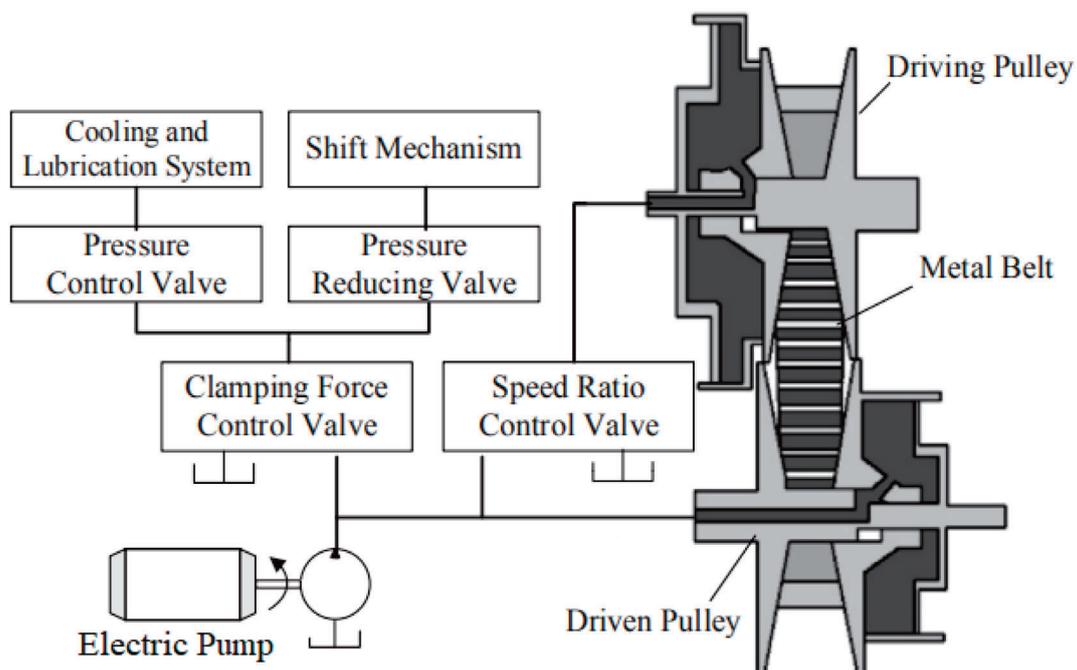


Figure 9. CVT with electro-hydraulic actuation system presented in [61].

One of the latest works describing the application of CVT in an EV powertrain can be accessed in [62]. This work presented an optimization of the EV ownership cost by taking into account the components' cost and the electricity cost for all the components involved in the powertrain; battery, motor (with inverter), and, CVT. The optimization was carried out using convex programming design optimization, with the targets to minimize the cost; by a means of optimizing the size of the motor and battery, without compromising the driving performance in terms of 0–100 km/h acceleration time (at most 11 s), top speed of 165 km/h and gradeability of 30%. In the work, three powertrain models were evaluated; first, the base powertrain model taken from the actual EV that used single speed

transmission, second, the modified powertrain model, which essentially the based model with CVT instead of the single speed transmission, and lastly, the optimized powertrain model, in which the CVT ratio as well as the size of the main powertrain components were optimized based on the actual driving data obtained from road and dynamometer tests. Similar to [61], the CVT evaluated in [62] also used electro-hydraulic actuation system for belt's clamping and ratio changing. However, the design integrated the cooling system of both the CVT and the motor, where the heat from the CVT fluid was dissipated to the motor's coolant through a heat exchanger, and then the coolant would be cooled down by the radiator (Figure 10). As a result, an extra radiator for the CVT was unnecessary, and this led to a more compact and cost-effective thermal management system. By simulating all the powertrain models under WLTC, the results showed that the optimized powertrain model performed the best in terms of efficiency (11.19 kWh/100 km, means 2.1% improvement against the base model with 11.43 kWh/100 km) and cost (2% cost reduction against the base model). In terms of cooling power consumption, the optimized model also gained an improvement of 30% as compared to the base model, and this means that the integrated thermal management system was not only cost effective, but it was also very efficient in controlling the operating temperature of the motor and the CVT. Besides, the integrated system also presents an opportunity for further integration with the battery's thermal management system which will potentially lead to further improvements in terms of cost and efficiency. Nevertheless, power losses in the hydraulic actuation system can still be expected, and it is interesting to evaluate how the optimized CVT performs against a two-speed transmission with optimized gear ratios.

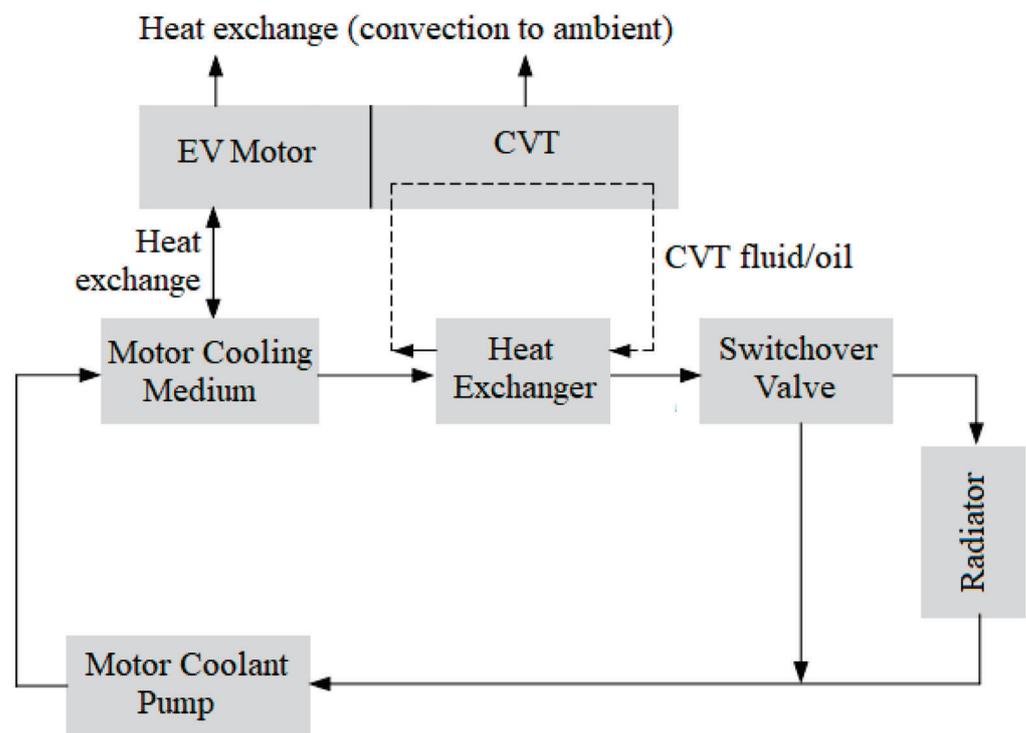


Figure 10. Integrated thermal management system for motor and CVT proposed in [62].

Another paper that discussed the potential cost benefit of using CVT in an EV can be found in [63]. Here, the application of CVT was complemented by the hybrid battery technology that incorporated a supercapacitor, and the cost benefits considered not only the component and electricity cost, but also the battery replacement cost. The CVT used an electro-mechanical actuation system for clamping and ratio changing, which means it featured a self-lock mechanism to maintain the belt's clamping force without using hydraulic pressure. The actuation system comprised of two DC motors with presumably

power screw mechanism, each for actuating the primary pulley and the secondary pulley. So, theoretically, it was more efficient than the CVT described in [61,62] thanks to the self-lock capability. However, detail description on the CVT's actual electro-mechanical actuation system was not described in the paper. The CVT losses model was developed to estimate its efficiency, and based on the model, the efficiency was estimated between the range of 78% to 89%. When simulated under a combined driving cycle of HWFET and FTP-75, the usage of the electro-mechanical CVT reduced the motor's losses by 37.9%, which was translated into 8.3% improvement in the power consumption of the vehicle when compared against single speed transmission. In terms of battery degradation, using CVT reduced the degradation by 7.2% as opposed to the single speed transmission, and, with the proposed hybrid battery technology, the improvement rate can be increased further to 17.5%. The battery degradation was defined in terms of capacity loss percentage, estimated using the LiFePO₄ cell's dynamic model that took into account the charging rate and the battery temperature during driving [64,65]. The latest review paper providing further descriptions on the estimation techniques for battery state of health in an EV can be accessed in [66]. Finally, in terms of cost benefits for 11-year of operation, when compared the application of CVT against single speed transmission (both using the hybrid battery technology), a saving of around USD 4541 can be expected for the consumers for the battery cost, resulted from the reduction of the required battery capacity due to the improvement in the powertrain efficiency. In addition, a further saving of USD 1768 can also be gained due to the reduction in the electricity cost. In overall, after reflecting the battery replacement cost as well as the penalty cost for using CVT, the total cost benefit was estimated at USD 3178 for 11-year of operation.

More advanced research on CVT application in an EV powertrain was explained in [67] which involved the optimization of an eco-driving strategy. The optimization objective was to minimize the reduction of the battery's SOC during the driving by taking into account not only the motor's efficiency, but also the instantaneous SOC as well as the efficiency of the CVT. Here, the CVT efficiency model was developed using mathematical equations introduced in [68–70]. NEDC was used as the driving cycle, in which it was divided into three driving conditions, namely, constant driving speed, acceleration condition and deceleration condition. In these conditions, the powertrain efficiency was analyzed for different SOC, CVT ratios and ambient temperature. One interesting aspect of this work is that it analyzed the SOC until the range of below 10%. Such SOC is rarely discussed in other literatures because in the actual application, the cut-off set point for the battery is usually set at around 20% to avoid any risk of damage. The analysis showed that, low SOC (at about 10%) decreased the powertrain efficiency by 33.12% and the acceleration time became longer by 68.8%. Such inefficiency was caused by the degradation of the battery that becomes significant starting from 10% SOC and lower. Because of the degradation, the battery internal resistance increases and its open circuit voltage decreases, which resulting in the increase losses percentage in the current flow from the battery to the EV motor. Moreover, at 10% SOC, the battery also became more sensitive to ambient temperature, which caused the losses in the current flow to increase even further. Above 10% SOC, however, the ambient temperature was insignificant in influencing these parameters, and the battery degradation was negligible, hence powertrain efficiency became more stable. By incorporating this knowledge in the eco-driving strategy, the constant driving speed condition in the NEDC can be increased from 61 to 70%, and the total driving time can be reduced by 12.1%, resulting in a more economical driving. This study highlighted the importance of eco-driving method that can only be implemented if the EV powertrain has the required flexibility in providing diverse driving modes. In addition, the study also served as the starting point to integrating battery's health conditions in the powertrain analysis.

To summarize, CVTs can provide better flexibility in ratio selection due to their continuous ratio range as opposed to any multispeed discrete transmission. This flexibility allows the EV's motor to operate optimally for various driving conditions. However, for the actual operation of the CVT in an EV, an appropriate shifting strategy must be formulated, either

continuous shifting or discretized shifting. The first strategy leads to better motor efficiency, but requires higher actuation power and advanced shifting logic. The second strategy, on the other hand, compromises the motor’s efficiency slightly as compared to the first strategy, but its shifting logic can be made simpler and the actuation power consumption can be reduced. Another area that has to be studied is the actuation system for ratio shifting and belt’s clamping in the CVT. Here, three possibilities can be explored; either optimizing the hydraulic actuation system typically used in the existing CVT, or implementing electro-mechanical actuation system to replace the hydraulic actuation system in the CVT, or, developing geartrain-based design of CVT. By optimizing the hydraulic actuation system, the power required to generate continuous pressure for CVT ratio and belt’s clamping force can be reduced. However, since the required belt’s clamping force is still very high (around at least 10 kN and it increases with the increment of the motor’s torque), thus the amount of required power will always be significant. As compared to any multispeed discrete transmission, such issue is eliminated thanks to the application of geartrain. Implementing electro-mechanical actuation system, on the other hand, eliminates the power requirement for the continuous pressure due to its self-lock mechanism, but, designing and integrating such system in a CVT requires extensive study to confirm not only its workability but also its reliability. A summary of the literature review on the CVT application for EV powertrains can be found in Table 4.

Table 4. Summary of the literature review on CVT application in an EV powertrain.

Literature, Year	Summary of the Works	Significant Findings
Fernandes, 2016 [55] and Hofman and Janssen, 2017 [56]	<ul style="list-style-type: none"> - In [55], the objective was to control the appropriate CVT ratio using fuzzy logic based on the motor’s efficiency mapping as the reference. - The fuzzy control algorithm was simulated based on three driving routes with different gradients. - The work in [56] described the potential of CVT in providing flexibility for motor to operate efficiently thanks to its continuous ratio. - However, it also emphasized on the importance of the self-lock mechanism for the CVT’s pulley system. 	<ul style="list-style-type: none"> - In [55], with the control algorithm, the efficiency was improved through the routes. However, no details description on the CVT used, only its ratio range was provided. - Ref. [56] highlighted the possibility of using electro-mechanical actuation system for the pulley for its self-lock mechanism. be selected more efficiently.
Ruan et al., 2018 [57] and Ruan et al., 2018 [58]	<ul style="list-style-type: none"> - To study the potential of CVT for improving efficiency of an EV powertrain. Its ratio range was determined based on the capability for climbing a gradient (maximum underdrive) and reaching top speed (maximum overdrive). - The CVT was considered to be using electro-mechanical actuation system for its pulley, hence no losses due to the hydraulic system. It was also considered more cost-effective for production against other transmission types. - It was compared against DCTs (two, three and four-speed) and a single speed transmission. The DCTs’ gear ratios were determined based on progression factor between the maximum underdrive and maximum overdrive. 	<ul style="list-style-type: none"> - Based on a combined FTP-75 and HWFET cycle, the simulation results showed CVT as the most efficient for a B-segment car, 17.4% more efficient against the single-speed transmission, followed by the two-speed DCT (16.4%) and the three-speed DCT (3.3%). - For a E-segment car, CVT was also the most efficient with a bigger margin compared to the second spot (24% to 9.6%). - Proving wide range of CVT application across EV segments. However, more details on the actuation system have to be considered. The DCTs’ gear ratios were also not optimized, hence their full potential was also not fully considered.

Table 4. Cont.

Literature, Year	Summary of the Works	Significant Findings
Hofman and Salazar, 2020 [59]	<ul style="list-style-type: none"> - To study an EV powertrain model that used CVT, with its efficiency set constant at 97%. - The model considered motor’s efficiency map, and the model was simulated for WLTP. - The results were also compared with two-speed AMT and two-speed DCT with the same efficiency. - No details on the actuation system/gearshift mechanism for all transmissions studied. 	<ul style="list-style-type: none"> - Results showed 3% improvement in efficiency of the powertrain with CVT against the other two transmissions. - However, 97% efficiency for the CVT (same as the AMT and DCT) can be considered too ambitious, the range of 80% to 89% was more realistic considering the power losses in the belt and the hydraulic actuation system. Such losses were not applicable for the AMT and DCT. - Highlighting that efficiency of the CVT is the key. It can be increased by electro-mechanical actuation system plus precise clamping control.
Sluis et al., 2019 [60]	<ul style="list-style-type: none"> - To evaluate CVT with optimized hydraulic actuation system and novel metal belt design. - Also include downsizing of the electric motor. - Simulation was run under WLTC for comparison against a single-speed transmission and a two-speed AMT EV powertrain. - No details on the gearshifting mechanism for the two-speed AMT. 	<ul style="list-style-type: none"> - Results showed 12.7% efficiency improvement against the EV powertrain with single speed transmission. - However, the two-speed AMT performed better with 13.5% efficiency improvement. - For the 0–100 km/h acceleration, the AMT and the single speed transmission produced 6.9 s, while the CVT achieved 7.4 s. - CVT was not the best performer, but it eliminated torque interruption during ratio shifting.
Hu et al., 2019 [61]	<ul style="list-style-type: none"> - To discretize the CVT ratios so that shifting frequency can be optimized for efficiency with minimum frequency. - It featured electro-hydraulic actuation system, where an electric pump was used to precisely control the required hydraulic pressure for clamping and ratio shifting (Figure 9). - Ratio number was set at 4 based on the optimum energy cost. Then, the four ratios were optimized using GA for optimum efficiency under UDDS, NYCC and HWFET. - Three ratio shifting strategy models were employed; first, continuously variable ratio shifting strategy, involving continuous ratio change for maximum motor’s efficiency, second, the discrete ratio shifting strategy based on the ratios established through the equal ratio series method, and third, the discrete ratio shifting strategy based on the ratios optimized through GA. 	<ul style="list-style-type: none"> - Results confirmed the third strategy as the best, with the minimum total power consumption and average jerking, measured at 8.10 kWh and 4.32 m/s³, respectively. - The first and second strategies recorded 8.16 kWh and 5.35 m/s³, as well as 8.69 kWh and 4.65 m/s³, respectively. - Highlighted two very important findings; firstly, potential of discretizing the CVT ratio for optimum performance for diverse EV segments, secondly, continuous ratio changes also resulted in actuation power losses and compromised driving comfort, apart from gain in the motor’s efficiency.

Table 4. Cont.

Literature, Year	Summary of the Works	Significant Findings
Wei et al., 2021 [62]	<ul style="list-style-type: none"> - To optimize the EV's ownership cost considering cost of the powertrain's components and the electricity cost. - Using convex programming design optimization, objective to minimize the costs of the motor and the battery without compromising 0–100 km/h acceleration time (max 11 s), top speed of 165 km/h and gradeability of 30%. - Three powertrain models were evaluated; first, the base powertrain model taken from the actual EV that used single speed transmission, second, the modified base powertrain model with CVT, and lastly, the optimized powertrain model comprising optimized CVT ratio, motor and battery. - CVT used electro-hydraulic actuation system for belt's clamping and ratio changing with an integrated cooling system for both the CVT and the motor. 	<ul style="list-style-type: none"> - Results for WLTC showed that the optimized powertrain managed 11.19 kWh/100 km power consumption, which was 2.1% lower than the based powertrain model. Also, the powertrain's cost was 2% lower, and the cooling power consumption was reduced by 30%. - The integrated cooling system was more compact and more cost-effective, particularly because of the usage of single radiator (Figure 10).
Ruan et al., 2019 [63]	<ul style="list-style-type: none"> - To evaluate the application of CVT with the hybrid battery technology incorporating supercapacitor, mainly in terms of electricity cost and battery's replacement cost, defined based on the degradation status. - Degradation was determined using capacity loss percentage using LiFePO4 cell's dynamic model from [64,65]. - The CVT used electro-mechanical actuation system for clamping and ratio changing, with two DC motors for ratio and clamping control. However, details on the system and controls were not included. - Efficiency of the CVT was estimated between 78% to 89%. 	<ul style="list-style-type: none"> - Simulated under HWFET and FTP-75, the motor's losses were reduced by 37.9%, equivalent to 8.3% improvement in power consumption against single speed transmission. - For battery degradation, using CVT reduced the degradation by 7.2% as opposed to the single speed transmission, and, with the proposed hybrid battery technology, the improvement rate can be increased further to 17.5%. - For 11-year cost benefits, a saving of around USD 4541 for the CVT can be achieved due to the reduction of the battery's required capacity. - USD 1768 electricity cost savings can be realized.
Liao et al., 2021 [67]	<ul style="list-style-type: none"> - To optimize CVT operation in an EV with eco-driving strategy. - Optimization objective was to maximize the battery's SOC considering not only motor's efficiency, but also the instant SOC and efficiency of the CVT. - CVT efficiency model was developed using mathematical equations introduced in [68–70]. - The powertrain efficiency was analyzed for different SOC, CVT ratios and ambient temperature for constant driving speed, acceleration and deceleration phases in NEDC. 	<ul style="list-style-type: none"> - Results showed that at low SOC (about 10%), the powertrain efficiency was reduced by 33.12%, while the acceleration time became longer by 68.8%. - Ambient temperature was insignificant for the powertrain efficiency. - Objective of the eco-driving was to maximize the instant SOC by increasing the constant driving phase, without compromising much on the acceleration time. - Results with the eco-driving showed that the constant driving phase can be increased from 61% to 70% of the total driving time, when the SOC was optimized.

4.3. Multi-Motor Configuration

The idea of optimizing EV powertrains using a multi-motor configuration involves properly distributing the driving loads, typically between two motors or four motors, so that they can operate optimally under various driving conditions. Additionally, in some situations, these motors can also provide boosting power for faster acceleration and higher top speed. As a result, the capacity of the motors can be reduced without compromising the driving performance and this leads to potentially significant cost savings. There are three common types of multi-motor configuration studied for EV; the first one is two-motors configuration where both motors are connected to a transmission before the wheels, the second is the two-motors configuration where both motors are directly coupled to the wheels, and finally, the four-motors configuration where all motors are directly coupled to the wheels (Figure 11).

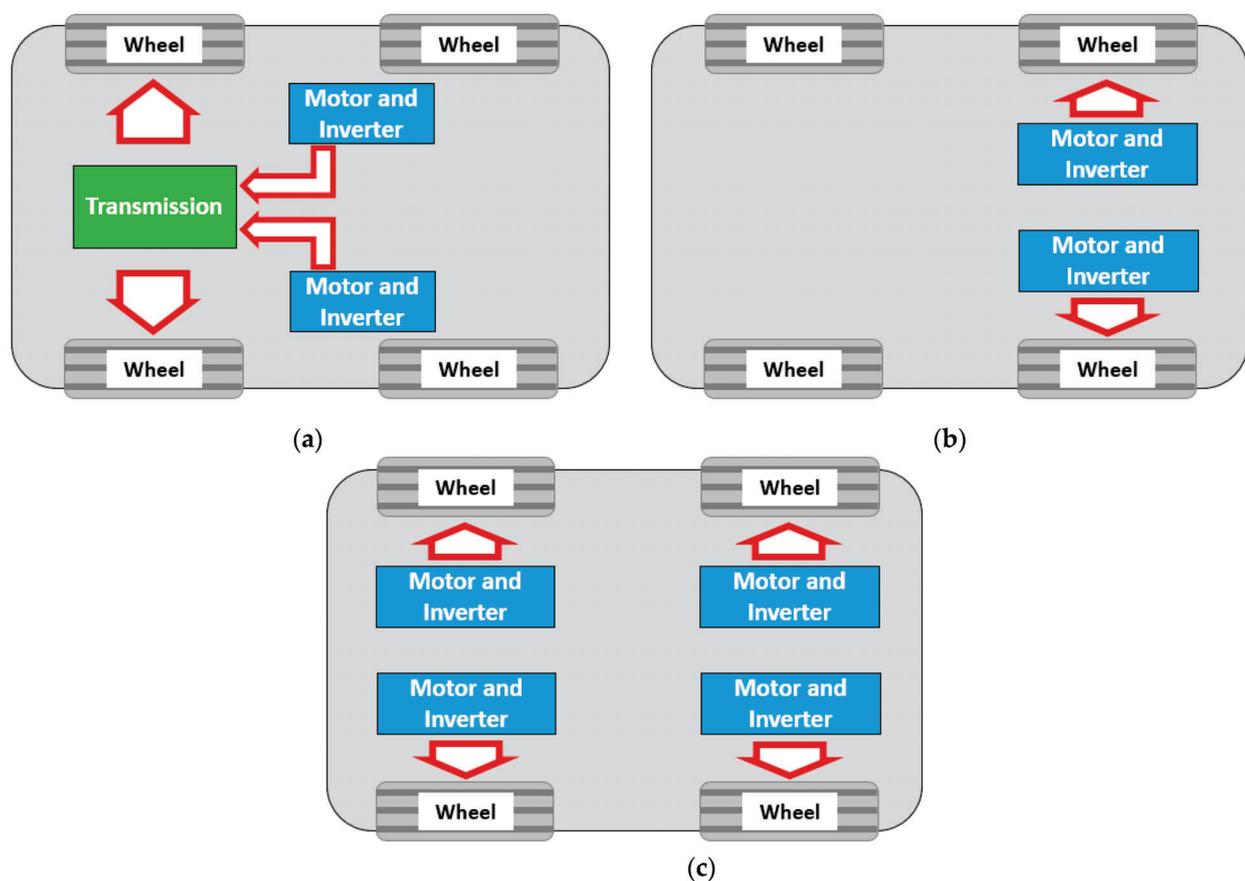


Figure 11. (a) Two-motors configuration where both motors connected to a transmission before the wheels, (b) both motors connected directly to the wheels, and (c) four-motors configuration where all motors directly connected to the wheels.

In the two-motors configuration where two motors are connected to a transmission before the wheels, allowing torque and speed couplings between the motors is crucial to suit diverse driving conditions. As shown in Figure 12, torque coupling means the torque from the ICE and the motor is combined through direct gearwheel, resulting in the shared torque requirement between them at the wheels. Equations describing the output torque and speed of the coupling are shown in Equations (1) and (2). Speed coupling, on the other hand, means the speed of the ICE is combined with the motor's speed through planetary gearset, resulting in higher speed, hence higher power, at the wheels (Figure 13). Equations (3) and (4) explain the relationship between the motors' inputs and the coupling's outputs. Torque coupling is generally useful for start-stop condition, while

speed coupling is usually applied to achieve fast acceleration and high-speed driving. Latest examples for the optimization of HEV powertrain configurations can be read in [71–73]. In Peng et al. [71], various HEV powertrain configurations based on CVT with metal belts and discrete gear automatic transmissions as the torque coupling, and a planetary gearset as the speed coupling, were generated using a fundamental matrix. In the work, feasible driving modes of these configurations were determined using an adjacency matrix, and based on these modes, the powertrain configurations were evaluated and compared against the benchmark configuration (Figure 14a) in terms of 0–100 km/h acceleration time and average power consumption under WLTC. The results demonstrated that the best configuration, as depicted in Figure 14b, managed to reduce the acceleration time and the average power consumption by 8.7% and 12.2%, respectively. Such improvements were possible because of the flexible driving modes provided by the proposed configuration that resulted in the reduced ICE power required for some driving conditions (due to the planetary gearset at the motor’s output), and more efficient regenerative braking (due to the several torque coupling possibilities at clutch C3 and C2). Results in [71] are also consistent with those discussed in [73], where it was found that the combination of CVT and planetary gearset is crucial to optimize the HEV’s efficiency for various vehicle speeds.

$$T_{out} = \frac{R_{out}}{R_{motor1}} T_{motor1} + \frac{R_{out}}{R_{motor2}} T_{motor2} \tag{1}$$

$$\omega_{out} = \frac{R_{motor1}}{R_{out}} \omega_{motor1} = \frac{R_{motor2}}{R_{out}} \omega_{motor2} \tag{2}$$

$$T_{out} = \frac{2R_{carrier}}{R_{sun}} T_{motor1} = \frac{2R_{carrier}}{R_{ring}} T_{motor2} \tag{3}$$

$$\omega_{out} = \frac{R_{sun}}{2R_{carrier}} \omega_{motor1} + \frac{R_{ring}}{2R_{carrier}} \omega_{motor2} \tag{4}$$

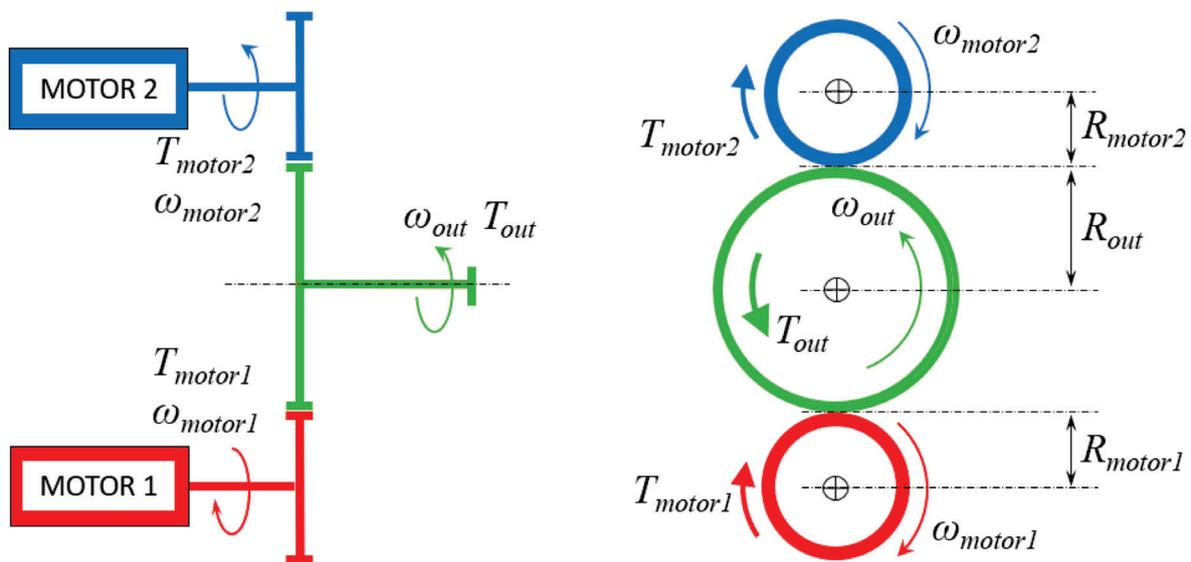


Figure 12. Torque coupling realized through direct gearwheel meshing.

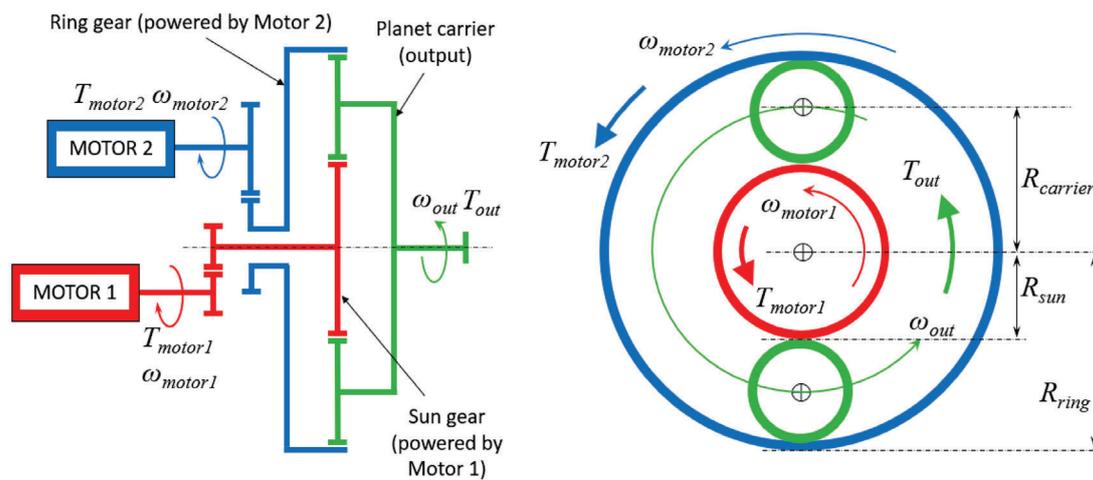


Figure 13. Speed coupling realized through planetary gearset.

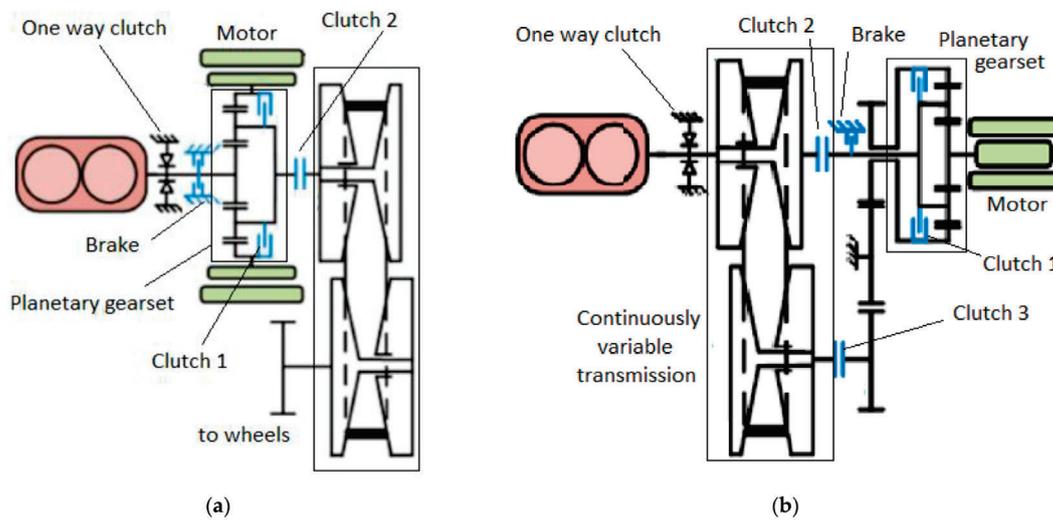


Figure 14. (a) The benchmark and (b) the proposed HEV powertrain configuration described in [71].

Outcomes in [71–73] highlight the importance of a proper strategy for torque and speed coupling between two or more motors that can lead to flexible driving modes for an EV. Thus, to evaluate its implementation in EV [74–77], studied the effects of using multi-motor configurations in EV powertrain on the power consumption based on several driving cycles. In [74], the same two motors were used in a two-motor EV powertrain configuration, and the powertrain was evaluated based on three different torque distribution strategies between the motors, where the final strategy was optimized using adaptive non-linear PSO. In [75], on the contrary, two motors with different maximum torque were used where they were connected through planetary gearset. Here, only speed coupling is possible, and these motors were controlled using a combination of speed feedback control strategy and torque feedforward control strategy to minimize jerking during the shifting of the driving modes. Other papers describing the implementation of dual motor configuration with planetary gearset can be read in [76,77], and because they allowed only speed coupling, the flexibility in terms of driving modes was limited.

Meanwhile, in [78], three configurations of two motors EV powertrain without multispeed transmission were considered; configuration for torque coupling (Figure 15a), configuration for speed coupling (Figure 15b), and a configuration for both torque and speed couplings (Figure 15c). Therefore, this configuration offered significant improvement in driving modes flexibility. In the first configuration, a single planetary gearset was used, in which the first motor was directly connected to the gearset’s sun gear, while the second

motor was connected to the same sun gear through a clutch. For the second configuration, two planetary gearsets were used with a brake at their ring gears. The first motor was meshed to the sun gear of the first planetary gearset, and the planet carrier here was rotatable. The second motor, on the contrary, was connected to the second planetary gearset's sun gear, where its planet carrier was fixed to the casing. The engagement of the brake on the ring gears of both gearsets was used to control the speed coupling in the configuration. Finally, the third configuration was essentially a heavily redesigned second configuration, that now has a clutch between the sun gears, and the rotation of the second planet carrier was controllable through another brake. As a result, the third configuration was more flexible, thanks to its capability to provide both torque (through the clutch between the sun gears) and speed couplings (through the brake at the ring gears). The benchmark in the study was a single motor EV powertrain with single-speed transmission. The gear ratios of all these powertrain configurations were optimized using non dominated sorting GA (NSGA-II) for optimum efficiency under UDDS, HWFET and NEDC. Compared to the single motor EV powertrain with single-speed transmission, in average, the single motor with two-speed transmission was about 2% more efficient, while the first, second and third two motors EV powertrain configurations were 5.77, 5.57 and 6.40% more efficient, respectively. It was interesting to note here, that the efficiency of the first two motors configuration (capable of torque coupling only) and the second configuration (capable of speed coupling only) were pretty much the same, although, in terms of mechanical system, the latter was significantly more complex than the former due to the application of two planetary gearsets. Next, for the second configuration with speed coupling and torque coupling options, the design was much more complex since it required three actuators for the two brakes and a clutch. With the difference of only 0.63% in terms of efficiency gain between the first and the third configurations, it was logical to choose the first configuration for an actual implementation. Therefore, in the future work for [78], more studies can be carried out to compare the driving and gearshifting performance of these configurations so that more aspects can be evaluated.

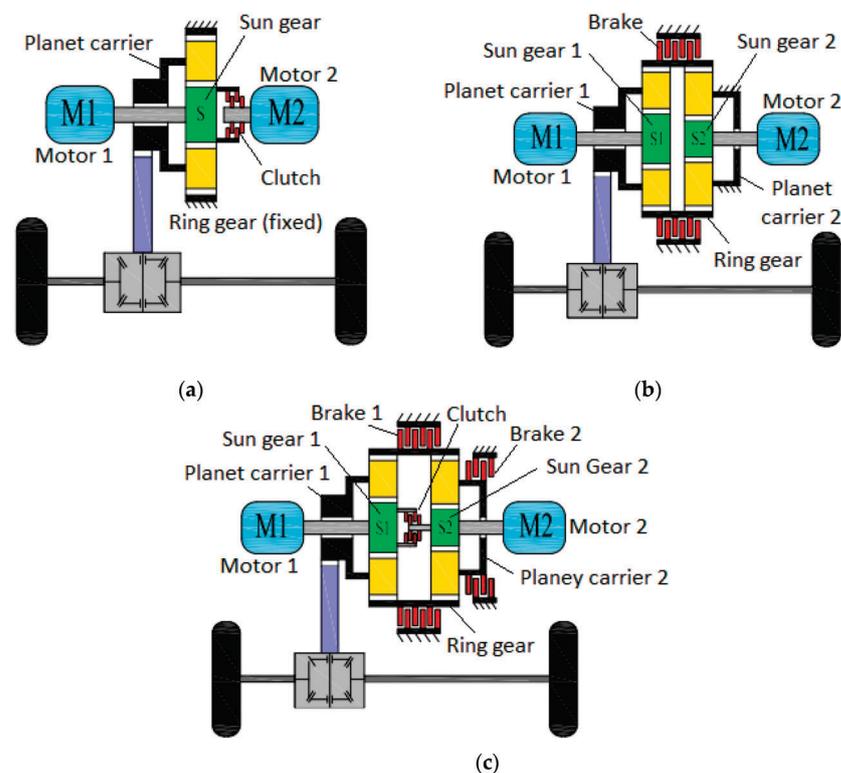


Figure 15. (a) First, (b) second and (c) third two motors EV powertrain configurations proposed in [78].

Next, in a latest work described by [79], a two-motor configuration based on a Simpson planetary gearset was proposed in a two-motor EV powertrain configuration. Unlike the configuration described in [78], this proposal provided two gear ratios in the powertrain, resulting in more flexibility for the driving modes. The Simpson planetary gearset consisted of two planetary gearsets with a brake for each set's ring gear, and another brake was used at the first motor's shaft. The ring gear of the first planetary gearset was connected to the second motor, and the ring gear of the second planetary gearset was connected to the first gearset's planet carrier. The planet carrier of the second gearset, on the other hand, was connected to the wheels through a differential, while the sun gears of both planetary gearsets was connected to the first motor. Figure 16 depicts the diagram of the Simpson planetary gearset configuration proposed in [79], which was capable of providing six driving modes (two modes with two motors, and four modes with one motor). The two modes with two motors represented the possibility of torque coupling and speed coupling of the motors, and the four single-motor modes represented the power flow from the first and second motor through two-speed gear ratios. The powertrain configuration's motor power and gear ratios were optimized using GA for minimum average efficiency in six driving cycles (LA92, JP1015, NEDC, WLTP and HWFET), high gradeability (40% at 10 km/h), high top speed (at least 190 km/h) and fast 0–100 km/h acceleration time (at around 10 s). The proposal was then compared with the typical parallel axle dual motor configuration with fixed gear ratio for evaluation. In terms of acceleration, the proposal provided faster 0–50 km/h acceleration time, but no significant difference was observed for the 0–100 km/h. For the average efficiency, the proposed Simpson planetary gearset configuration was more efficient than the typical parallel axle dual motor configuration by around 2.88 to 8.33% when employed in driving cycles with frequent acceleration and deceleration. However, in a driving cycle with relatively constant vehicle speed (like HWFET), the typical configuration was slightly more efficient by 0.45%, very likely due to the losses in the planetary gearsets. Moreover, the proposed configuration also required three actuators for the three brakes to properly control coupling between the two motors. In this aspect, advanced control algorithm for the actuators is crucial to ensure that they can be operated systematically not only for powertrain efficiency but also for driving comfort by minimizing the jerking during the driving mode shifting. Thus, in the future study, it is imperative to evaluate the workability and control of these actuators so that their performance in terms of powertrain efficiency, jerking and shifting time can be quantified and compared.

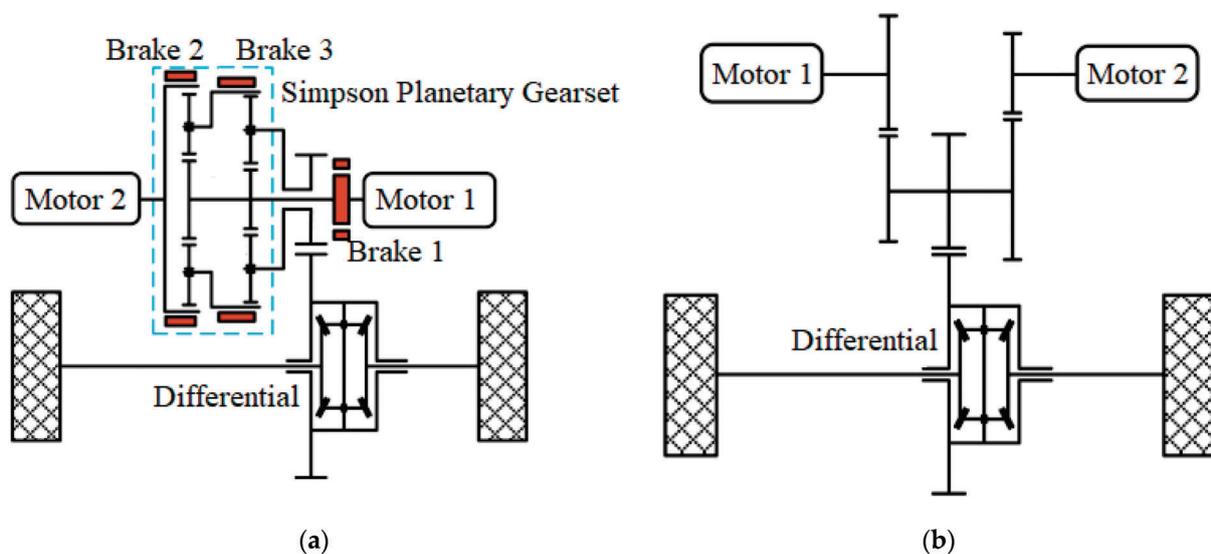


Figure 16. (a) Proposed two motors powertrain configuration with Simpson planetary gearset and (b) the typical parallel axle two motors configuration [79].

The subsequent work on multi-motor configuration is described in [80], which like [79], studied the possibility of applying an EV powertrain with a two motor configuration and a multispeed discrete transmission, in this case four speed AMT. The powertrain was applied in a city bus for a specific Nuremberg City Cycle and also for NYCC, where the focus was to optimize the driving strategy by properly coordinating the gearshifting. This was to avoid gear hunting, which not only affected the powertrain efficiency, but was also detrimental to the driving comfort. Next, in [81], the objective was to determine and compare the performance of the configuration against single motor configuration with four-speed AMT. The study started by generating two optimized configurations of two-motors-four-speed-AMT EV powertrain, where the optimization objectives were to obtain minimum operating cost (defined as minimum total power consumption for the aforementioned driving cycle) and high driving performance (defined as the minimum acceleration time for 0–40 km/h). The control variables of the optimization were motor scale factor; expressed as the ratio of motor 1's power divided by the power of motor 2, and the four gear ratios of the AMT. The optimum combinations of the variables to meet the optimization objectives were determined using NSGA-II. As a result, three powertrain configurations were finalized; Configuration 1 that consisted of one motor configuration and two optimized gear ratios, Configuration 2 that consisted of two motors configuration with the motor scale factor of 0.42 and four optimized gear ratios, and Configuration 3 that consisted of two motors configuration with the motor scale factor of 1.00 and four optimized gear ratios. All of them achieved the same acceleration time of 8.5 s, but in terms of power consumption, Configuration 1 recorded the worst at 7.48 kWh, while the second and third configurations managed to improve over the first one by 4.82% and 5.08%, respectively. Next, the optimum driving strategy was formulated for each configuration in terms of shifting schedule and motors coupling modes so that the driving efficiency and performance can be improved further. The simulation results of the three powertrain configurations plus the optimal driving strategy showed that, both Configurations 2 and 3 allowed the motors to operate with at least 85% efficiency rate for 65% of the driving cycle. As a comparison, in Configuration 1, the motor was allowed to operate at the same efficiency rate only for about 32% of the same driving cycle. Because of that, the total power consumptions obtained for Configurations 2 and 3 were lower than Configuration 1, at 7.219 kWh and 7.216 kWh, against 7.627 kWh, respectively. Nevertheless, it must be mentioned that in Configuration 1, the gear shifting occurred only 46 times during the cycle, which was significantly lower than 84 and 80 each for Configurations 2 and 3. These findings indicated that the one motor EV powertrain configuration with two-speed transmission is potentially advantageous in terms of overall cost (production, operation and maintenance costs) than those two motors configurations, even though it performed the worst in terms of efficiency. Not only that, the acceleration time was also the same for all configurations, and this reinforce Configuration 1 as the overall best choice as opposed to the other two. Therefore, further study on the operation cost can be carried out in the future by relating the data of the gearshifting frequency and the wear and tear of the shifting mechanism. Next, the shifting performance especially in terms of jerking must also be evaluated for all the configurations so that more aspects can be compared to determine the overall best configuration between the three. Other works describing the application of multi-motor configurations with multispeed transmission can also be read in [82], which reviewed the methodologies of the multi-motor configurations, and [83], which presented an optimization of gear ratios and torque distribution of two-motor EV powertrain configuration with two-speed transmission using a surrogated model developed based on an effective adaptive sampling method.

Apart from using two motors powertrain configurations, there were also other studies performed to evaluate the application of four motors in the EV powertrain. For instance, in [84,85], four motors were used with each of them assembled in the EV's four wheels (Figure 17). The main idea here was to split the weight distribution of the powertrain evenly to all wheels, thus enabling the increase of driving flexibility without the application of bigger motor and transmission. Each motor was also coupled to a two-speed AMT

designed based on planetary gearset, where the appropriate gear ratio was actuated using DC motor and worm gear mechanism. In addition, the gear ratio actuator also featured ball-ramp self-energizing that consisted of translation plate with steel balls and spiral ramps. The purpose of this mechanism was to amplify the clutch engagement force during the gearshifting, resulting in the reduction of DC motor's required power. To simulate the powertrain performance, the AMT's gear ratios as well as its complete parameters were determined based on a transmission-equipped wheel hub motor described in [86], while the gearshifting schedule was developed based on the vehicle speed and the throttle position. To simplify the simulation model, a dog clutch model was used as the surrogate model for the proposed gear ratio actuator. Two shifting approaches were applied—synchronous and asynchronous—where synchronous means that the gearshifting occurred simultaneously for the front and the rear wheels, while asynchronous means the gearshifting was done independently between the front and the rear wheels with a delay of 0.2 s. The key benefit lay in the asynchronous approach, which minimized the jerking thanks to the delay that reduced the torque interruption in the front wheels' gearshifting by compensating it with the torque at the unshifted rear wheels (and vice versa). As a result, the jerking can be kept within the range of 4 m/s^3 to 6 m/s^3 . However, the proposed powertrain configuration involved four independent two-speed AMT actuators, which means, although it was highly flexible in terms of gearshifting, it required sophisticated control logic to avoid too frequent shifting and gear hunting. Too frequent shifting and gear hunting, if not properly optimized, will cause driving discomfort and increased losses in the actuators.

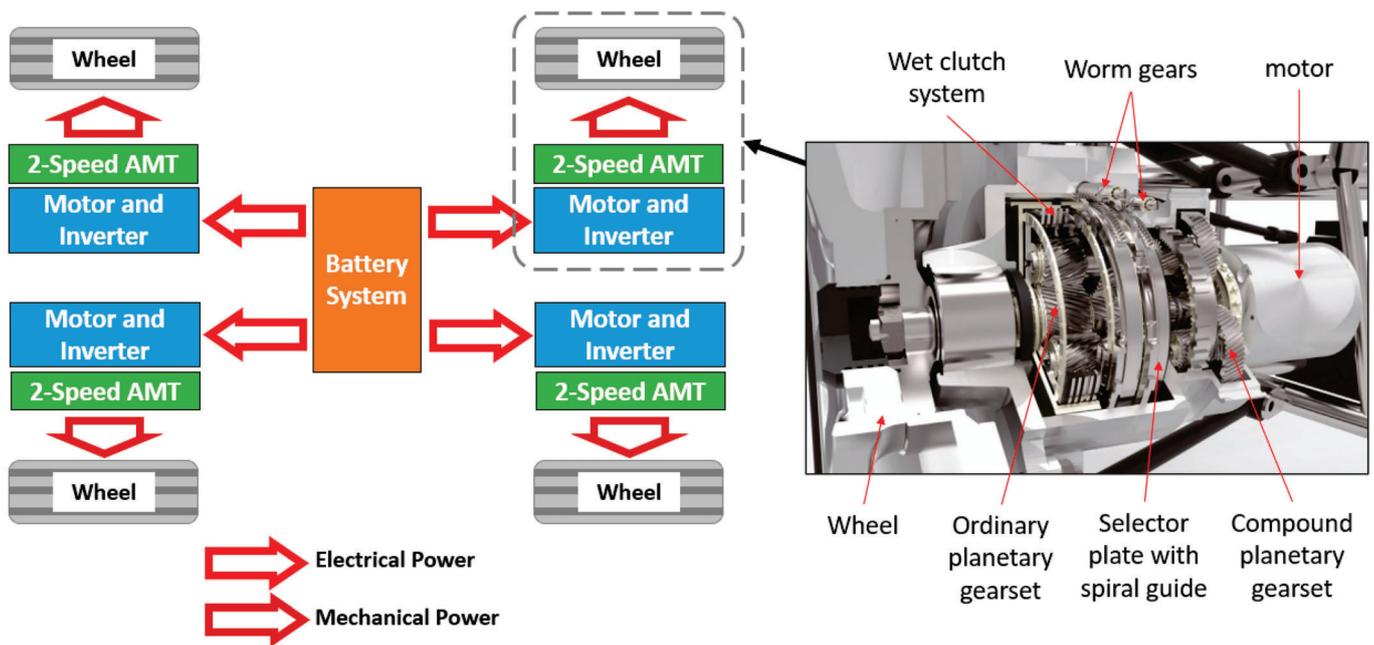


Figure 17. Four motors EV powertrain configuration with two-speed AMT [84,85].

The latest studies on four motors EV powertrain configuration focused on not only achieving optimum efficiency, but also enhancing driving safety and steering assist [87,88]. Unlike the powertrain configuration proposed in [83–85], the four motors EV powertrain introduced in [87,88] lacked the gearshifting options since it used single speed transmission. However, the omission of gearshifting options increase its simplicity in terms of operation and controls. In [87], the element of driving safety, together with the optimum efficiency in terms of power consumption, can be realized by implementing integrated motors' torque vectoring control strategy in the wheels. The integrated strategy was intended to achieve multiple objectives such as reasonable traction torque distribution on the wheels for yaw stability control and steering assists, proper motor's output torque to enable it to operate at

its most efficient range for optimum power consumption, and reducing the wheels' dynamic slip for driving stability as well as optimum power consumption. The proposed integrated strategy was then compared against the conventional axis distribution and maximization of stability margin strategies through simulation under WLTC, where the results showed that the integrated strategy reduced the wheels slip by 14.38% compared to the axis distribution strategy that led to the improved power consumption by 5.37%. The simulation results were then validated experimentally based on a single seat EV prototype that was driven at 60 km/h on a slippery road and then executing standard lane change maneuver. In the experiment, the wheels slip was reduced by 12.75% due to the implementation of the integrated torque vectoring control strategy as opposed to the axis distribution strategy. Meanwhile, in [88], a fuzzy logic algorithm was applied in the traction distributions on the wheels. The algorithm was responsible to make sure that sufficient traction can be given to each of the wheels to achieve the desired vehicle speed during the driving, while also ensuring that the vehicle trajectory follows the desired driving path. Components of the powertrain used here were optimized using PSO for minimizing the weight of the battery and motors, minimizing drop in the battery's SOC, and reducing driver steering efforts. With the optimized EV powertrain and the fuzzy logic algorithm for the wheels' traction distribution, driver steering efforts can be reduced by 78.5% and the driving mileage can be increased even with the reduced size of the battery. The results obtained in [87,88] demonstrated the potentials of implementing independent motors for all wheels not only for efficiency but also for the vehicle handling that can possibly contribute in the vehicle safety and autonomous vehicle technologies. For the future scope, the work in [87,88] can be expanded to evaluate the proposed configuration's performance in controlling the vehicle maneuvering against conventional traction control methods.

Summary of the reviewed literature on the multi-motor configurations for EV powertrain is presented in Table 5. There are two common multi-motor configurations studied in the literature; the two motor configuration, and the four motor configuration. The objective of the two motor configuration is to allow the operation of one motor to be supported by the other motor through either torque coupling or speed coupling. As a result, the operation of these motors can be optimized for diverse driving conditions. The challenge, however, is to come up with the proper mechanism for the couplings, which typically involves multi clutches and brakes. The next challenge will be to effectively and systematically control these clutches and brakes through actuators so that the driving mode shifting can be executed smoothly and efficiently. Regarding the four motors configuration, the objective is to minimize the transmission power loss, since the motor is coupled directly to each wheel. Besides, the traction on the wheels can also be distributed independently, which can improve not only the power consumption, but also the driving stability and safety. Nevertheless, to achieve these, advanced control algorithm is required to integrate effectively the operation of the four motors at the wheels.

Table 5. Summary of the literature review on multi-motor configurations for EV powertrain.

Literature, Year	Summary of the Works	Significant Findings
Peng et al., 2020 [71]	<ul style="list-style-type: none"> - To optimize speed and torque couplings of HEV powertrain configurations using different arrangements of CVT and planetary gearset based on fundamental matrix. - Next, feasible driving modes for different powertrain configurations were determined using adjacency matrix. The configurations were simulated and compared against the benchmark configuration shown in Figure 14a in terms of 0–100 km/h acceleration time and average power consumption under WLTC. 	<ul style="list-style-type: none"> - The simulation results demonstrated that the best configuration (Figure 12b on ref. [71]) reduced the acceleration time and the average power consumption by 8.7% and 12.2%, respectively. - Such improvements were gained from the reduced ICE required power for some driving conditions thanks to the planetary gearset at the motor’s output, and, more efficient regenerative braking thanks to the diverse torque coupling possibilities at clutch C3 and C2.
Zheng et al., 2020 [74] and Wu et al., 2021, 2018, 2018 [75–77]	<ul style="list-style-type: none"> - In [74], two-motor powertrain configuration was proposed consisting two motors with the same power output. - It was evaluated based on three different torque distribution strategies between the motors. - The final strategy was optimized using adaptive non-linear PSO. - In [75], two-motor powertrain configuration was proposed consisting two motors with different maximum torque, connected through planetary gearset, allowing speed coupling. Other similar works was described in [76,77]. - Jerking was minimized during the shifting of the driving modes using speed feedback control strategy and torque feedforward control strategy. 	<ul style="list-style-type: none"> - These works involved either torque coupling or speed coupling, hence the driving modes flexibility was not maximized. - Nevertheless, some improvements in terms of efficiency and jerking against one-motor configurations were still reported in the works.
Du et al., 2021 [78]	<ul style="list-style-type: none"> - Three different two-motors configurations of two motors EV powertrain. the first one torque coupling, the second one allowed speed coupling, and, the last one allowed both torque and speed couplings. - These configurations were differentiated in terms the connection between the planetary gearsets’ components and the engagement components (brakes and clutches) (Figure 14). - These configurations were compared against single motor EV powertrains with single-speed transmission and two-speed transmission. - Gear ratios of all the powertrains were optimized using non dominated sorting GA (NSGA-II) for optimum efficiency under UDDS, HWFET cycle and NEDC. 	<ul style="list-style-type: none"> - Against the single motor EV powertrain with single-speed transmission, the two-speed transmission one was about 2% more efficient, while the first, second and third two motors EV powertrain configurations were 5.77%, 5.57% and 6.40% more efficient, respectively. - Efficiency gain for exclusively torque coupling (first configuration) or speed coupling (second configuration) was almost similar. Significant gains can be achieved by combining both couplings capability - However, the second one was significantly more complex mechanically, compared to the first one due to the application of two planetary gearsets.

Table 5. Cont.

Literature, Year	Summary of the Works	Significant Findings
Hong et al., 2022 [79]	<ul style="list-style-type: none"> - Two-motor configuration based on Simpson planetary gearset was proposed with two gear ratios. - The planetary gearset consisted of two planetary gearsets with a brake for each set's ring gear. Detail on its mechanical design was provided in Figure 15. - The configuration was capable of providing six driving modes (two modes with two-motor, and four modes with one-motor). The two-motor modes represented the torque coupling and speed coupling of the motors. The four one-motor modes represented the power flow from the first and second motor through two-speed gear ratios. - The motor power and gear ratios were optimized using GA for minimum average efficiency in six driving cycles (LA92, JP1015, NEDC, WLTP and HWFET), high gradeability (40% at 10 km/h), high top speed (at least 190 km/h) and fast 0–100 km/h acceleration time targeted at around 10 s. - Further reads on two-motor configuration with two-speed transmission can be accessed in [82,83]. Ref. [82] is a review paper on the methodologies for EV powertrain, while [83] presented an optimization of two gear ratios and motors' torque distribution using a surrogated model developed based on an effective adaptive sampling method 	<ul style="list-style-type: none"> - Against the typical parallel axle two-motor configuration with single-speed transmission, the proposed configuration yielded faster 0–50 km/h acceleration, but similar 0–100 km/h acceleration rate. - The proposal's average efficiency was improved by 2.88% to 8.33% for driving cycles with frequent start-stop. - However, the average efficiency was slightly worse (about 0.45%) for the driving cycle with relatively constant vehicle speed. - Main challenge is to coordinate the actuation of the three brakes so that driving mode shifting (motors coupling) can be executed smoothly and efficiently.
Nguyen et al., 2021, 2022 [80,81]	<ul style="list-style-type: none"> - To study the application of EV powertrain with two-motor configuration and a four speed AMT. The powertrain was used in a city bus for Nuremberg City Cycle and NYCC. - To optimize efficiency and driving comfort by coordinating the gearshifting effectively so that losses and discomfort caused by gear hunting can be avoided. - The proposed powertrain was optimized for minimum total power consumption and minimum 0–40 km/h acceleration time. The control variables were motor scale factor (power ratio of motor 1 to motor 2) and the AMT's four gear ratios. The optimum combinations of the variables were determined using NSGA-II. - Three configurations produced from the optimization; Configuration 1 with one motor configuration and two optimized gear ratios, Configuration 2 with two motors configuration with the motor scale factor of 0.42 and four optimized gear ratios, and, Configuration 3 with two motors configuration with the motor scale factor of 1.00 and four optimized gear ratios. - Further reads on multi-motor configurations with multi 	<ul style="list-style-type: none"> - Against one-motor configuration with four-speed AMT, all three configurations achieved 8.5 s acceleration time, but for power consumption, Configuration 1 recorded the worst at 7.48 kWh, while the second and third configurations improved by 4.82% and 5.08%, respectively. - With the optimized gearshifting schedule and motors coupling modes, Configurations 2 and 3 allowed the motors to operate with at least 85% efficiency for 65% of the driving cycle's total duration. Meanwhile, Configuration 1 managed the same condition for only about 32% of the total duration. - The total power consumptions obtained for Configurations 2 and 3 were lower than Configuration 1, at 7.219 kWh and 7.216 kWh, against 7.627 kWh, respectively. - But, in terms of gearshifting frequency, Configuration 1 was the best with only 46 times (Configurations 2 and 3 recorded 84 and 80 each). Thus, Configuration 1 was potentially more advantageous in the overall cost (production and operation).

Table 5. Cont.

Literature, Year	Summary of the Works	Significant Findings
Meng et al., 2021, 2021 [84,85]	<ul style="list-style-type: none"> - To evaluate the operation of four-motor configuration; a motor for each of the EV's four wheels (Figure 16). The main objective was to distribute the powertrain's weight evenly to all wheels, thus improving driving flexibility without bigger motor and transmission. - Each motor was coupled to a two-speed AMT featuring planetary gearsets, in which the gear ratio was actuated using worm gear mechanism with DC motor. - The actuator also featured ball-ramp self-energizing capability for amplifying the clutch engagement force during the gearshifting, resulting in the reduction of the required DC motor's power. Complete AMT's parameters based on [86] and dog clutch model was used to represent the actuator in the simulation. - Two gearshifting approaches were applied; synchronous (simultaneous shifting for front and rear wheels) and asynchronous (independent shifting for front and rear wheels with 0.2 s delay). 	<ul style="list-style-type: none"> - In the asynchronous approach, the jerking was minimized. This was due to the delay that reduced the torque interruption in the front wheels' gearshifting by compensating it with the torque at the unshifted rear wheels (and vice versa). The jerking recorded within the range of 4 m/s³ to 6 m/s³. - However, the proposed powertrain configuration involved four independent two-speed AMT actuators, which means, it required sophisticated control logic to avoid too frequent shifting and gear hunting.
Wei et al., 2022 [87]	<ul style="list-style-type: none"> - To optimize power consumption and also vehicle dynamics by implementing integrated motors' torque vectoring control strategy in the wheels. - The four-motor configuration used single-speed transmission in each wheel. - Objectives of the optimization were to provide reasonable traction torque distribution on the wheels for yaw stability control and steering assists, to provide proper motor's output torque to enable it to operate at its most efficient range for optimum power consumption, and, to reduce the wheels dynamic slip for driving stability and safety. - The optimized proposed powertrain was then compared against the conventional axis distribution and maximization of stability margin strategies through simulation under WLTC. 	<ul style="list-style-type: none"> - Simulation results showed that the proposal reduced the wheels slip by 14.38% compared to the conventional one. This led to the improved power consumption by 5.37%. - Experimental results validated the simulation results. The experimental results were obtained based on a single seat EV prototype that was driven at 60 km/h on a slippery road and then executing standard lane change maneuver. - The experiment recorded that the wheels slip was reduced by 12.75% in the proposal as opposed to the conventional one.
Miranda et al., 2022 [88]	<ul style="list-style-type: none"> - To apply fuzzy logic algorithm in the wheels' traction distributions based on a four-motor configuration with single-speed transmission. - The algorithm was tasked to provide sufficient traction to all wheels to achieve the desired vehicle speed during the driving. During the process, the vehicle trajectory must follow the desired driving path for driving stability. - Components of the proposed powertrain were optimized using PSO for minimizing the weight of the battery and motors, maximizing the battery's SOC, and minimum steering efforts. 	<ul style="list-style-type: none"> - Results showed that steering efforts can be reduced by 78.5%. - Driving mileage was increased even with the reduced size of the battery. - The results demonstrated the potentials of implementing independent motors for all wheels not only for efficiency, but also for vehicle handling that can lead to improved vehicle safety and autonomous vehicle technologies.

5. Comparison and Future Works Related to the Methods for Optimizing Power Flow in EV Powertrains

In the previous section, the methods to optimize power flow in an EV powertrain are reviewed and divided into three: applying multispeed discrete transmissions, applying CVTs with metal belts, and implementing multi-motor configurations. Although these methods can lead to improvements in driving efficiency and performance, their advantages and disadvantages relative to each other must also be properly assessed. Therefore, in this section, the three methods are compared and evaluated extensively to discuss their potential advantages and disadvantages. Afterwards, key areas for future research works in the context of optimizing EV's power flow are presented and discussed.

5.1. Comparison of the Methods

Among the three methods, multispeed discrete transmission is the most common one studied by scholars. Within this class, the two-speed transmission is the most popular, which can be described in the form of either AMT or DCT. Most of the works related to the application of two-speed transmission reviewed here involved optimization of the gear ratios and shifting strategy to achieve optimum power consumption and driving performance, in terms of acceleration rate and top speed [33–40]. As a result, the EV powertrain becomes more efficient and more capable, and this opens up the possibility of optimizing the size and capacity of the motor and the battery. Such a possibility is beneficial for production sustainability, because the usage of heavy materials for motors and batteries can be reduced. At the same time, since the two-speed transmission is very compact and shares significant degree of similarity with the traditional one in ICE vehicles, the existing transmission manufacturing process can also be utilized which will be cost-effective for total EV production cost. The significant challenges, however, are the jerking in the gearshifting mechanism, the limited flexibility and the additional maintenance cost for the two-speed transmission.

Similar to multispeed discrete transmission, CVTs with metal belts also involve providing multiple ratios for optimum efficiency and driving performance in an EV powertrain. However, unlike a multispeed discrete transmission, CVTs are capable of providing a continuous ratio range, which addresses the limited flexibility problem faced by the multispeed discrete transmission approach. This presents an opportunity to implement it for diverse driving conditions and various vehicle segments. At the same time, it is also beneficial in terms of production sustainability and technology migration, because of the possibility to reuse the existing manufacturing processes and facilities. This is because the metal belt-based CVT for EV shares a significant number of common components with the existing ones used in ICE vehicles. The main challenges for this type of CVT, however, are its metal belt's operation and maintenance cost. The metal belt's operation requires high hydraulic pressure, significantly higher than the requirement for AMT and DCT, for maintaining its clamping force and ratio. In addition, the belt's operation also inevitably involves micro slippage between its components. These two factors cause transmission loss which is higher than that suffered in AMT and DCT. Also, these factors require slightly costlier maintenance than the other two types of discrete transmission commonly studied for EV application.

The last method reviewed here is the implementation of multi-motor configurations, which can be divided into two typical approaches: two-motor configurations and four-motor configurations. For the two-motor configuration, the possibility for both torque and speed couplings certainly leads to flexibility for various driving modes. However, in some cases, providing either speed coupling or torque coupling can already be sufficient to achieve higher powertrain efficiency than the application of two-speed discrete transmission. The two-motor configuration can also be applied with a multispeed transmission, which can be particularly useful for optimizing the power consumption of heavy vehicles such as trucks and busses. For the four-motor configuration, it allows even more diversity in the driving modes than the two-motor configuration, thanks to the possibility to distribute

traction to each wheel independently. This presents an opportunity to optimize not only the powertrain efficiency, but also the driving dynamic of the EV (which is not possible for both the multispeed discrete transmission and CVT) and the motors' capacity and size. The main challenges of the multi-motor configurations, however, are the complexity in terms of the mechanical design and control due to the application of multi clutches, brakes and EV motors, as well as the production sustainability due to the high number of motors involved. A summary of the comparison between these three proposed methods of optimizing the EV powertrain is presented in Table 6.

Table 6. Comparison between the application of multispeed discrete transmission, CVT and multi-motor configuration in EV powertrain.

Methods	Advantages	Disadvantages
Multispeed discrete transmission	<ul style="list-style-type: none"> - The gear ratios can be optimized for different driving conditions, more flexible than the single-speed transmission commonly used in EV. - Possibility to optimize the size and capacity of the motor and the battery for improved sustainability, production and ownership cost. - Share significant number of components similarity with the ICE vehicles, hence possible to reuse the existing production facilities and process. 	<ul style="list-style-type: none"> - Limited gear ratios, which limits the flexibility for more diverse and demanding driving conditions. Also prevent the possible application in wider vehicle segments. - Torque interruption and jerking during gearshifting. - Extra maintenance cost as compared to the single-speed transmission.
Continuously variable transmission (CVT)	<ul style="list-style-type: none"> - It is capable of providing continuous ratio range, hence greater flexibility than any multispeed discrete transmission. - Possibility to optimize the size and capacity of the motor and the battery for improved sustainability, production and ownership cost. - Share significant number of components similarity with the ICE vehicles, hence possible to reuse the existing production facilities and process. - Presents the possibility for application across wider vehicle segments, thanks to the greater flexibility. 	<ul style="list-style-type: none"> - The high requirement for the metal belt's operation and maintenance cost involving higher hydraulic pressure than AMT and DCT for maintaining its clamping force and ratio. - Inevitable micro slippage between the belt's components, thus higher transmission loss than AMT and DCT. - Slightly costlier maintenance AMT and DCT based on the usage in ICE vehicles.
Multi-motor configuration	<ul style="list-style-type: none"> - Possible to be implemented with multispeed transmission for much greater flexibility. - Providing diverse driving modes, more flexible than the single-speed transmission and the multispeed discrete transmission. - Splitting power requirement between the motors, thus their capacity and size can be reduced and optimized. - For four-motor configuration, it presents potential to also optimize driving dynamics and safety, apart from efficiency. 	<ul style="list-style-type: none"> - Complexity in mechanical design to allow the speed and torque couplings between the motors (for two-motor configuration). Even higher complexity when involved multispeed transmission. - Complexity in controls, especially for four-motor configuration that needs precise coordination between all motors at the wheels. Even higher complexity when involved multispeed transmission. - The production sustainability due to the high number of motors involved.

5.2. Key Areas for Future Research Works

The latest research works on two-speed transmission for application in EV powertrains reviewed in this paper mostly focused on optimizing the gear ratios for powertrain

efficiency and performance. Besides, these works also emphasized the proper gearshifting strategy, which should be formulated accurately by taking into account the road conditions and driver's input so that a balance between the powertrain efficiency and performance can be realized. This was consistent with the research trends previously discussed and reviewed in [89,90]. However, these works still did not sufficiently discuss the control algorithm of the gearshifting mechanism in detail. Examples of works on this topic can be read in [41–43,47,48], though these works evaluated only on the jerking of the mechanism alone without being integrated with the multispeed transmission inside the powertrain. Besides, they also did not assess the effect in terms of transmission efficiency and actuation power consumption. Therefore, in the future, works in the control of the gearshifting mechanism are expected to be intensified.

On the CVT with metal belt for EV powertrains, most of the previous related literatures focused on analyzing and comparing its efficiency and performance against multispeed discrete transmissions and multi-motor configurations. However, due to the application of hydraulic actuation system, the CVT suffered significant power losses, which prompted some scholars to study the practicality of replacing it with an electro-mechanical actuation system. At the moment, very limited works have been carried out to analyze thoroughly the application of electro-mechanical CVT for EV. Based on the latest review paper on an electro-mechanical CVT with metal belt in [91], some of the designs are capable not only of self-locking the ratio and belt's clamping force, but also precisely controlling them. Controlling the belt's clamping force, particularly, is the key to optimize the CVT efficiency as well as to optimize the durability of the electro-mechanical actuation system, as extensively explained in [92]. This means that transmission losses can be minimized as much as possible, though thorough studies still need to be performed since electro-mechanical CVT is still not a mature technology and so far, not being implemented for commercialization. Therefore, key research area here is the optimization of the ratio and clamping control algorithms in the electro-mechanical actuation system for CVT with metal belt. Another area that can be focused on is the possibility of implementing geartrain-based CVT, which eliminates the application of the metal belt entirely.

In the context of multi-motor configurations, key research areas that can be pursued are the control algorithm for traction distribution on the wheels and the durability of the powertrain system. Apart from optimum power consumption and driving performance, multi-motor configuration offers the chance to implement steering assist and wheels' traction control. However, since the powertrain is now attached directly to wheels, it now becomes part of the unsprung mass. As a result, the powertrain is now subjected to harsh operating condition involving direct vibrations caused by the road surface, as well as water splash and debris from the road surface. This factor very likely will affect the durability and maintenance routine of the powertrain, which requires further study to evaluate its significance to the overall ownership cost of the EV.

In terms of maintenance, refs. [93,94] discussed the gap between the operation cost for ICE and EV. In general, the purchasing cost for EV is higher due to the high battery cost, while the cumulative maintenance cost for ICE will be greater over time due to the frequent maintenance requirement for its powertrain. Depending on the vehicle segment, EV can achieve cost parity with the equivalent ICE model in around 8 years of ownership. Therefore, detail studies to determine the acceptable cost parity between ICE vehicle, multispeed discrete transmission, CVT with metal belt, multi-motors configuration and the conventional single-speed transmission EV are still required.

Another topic that is relevant for future research work here is the lubrication and cooling of the powertrain, which is particularly crucial for the CVT with metal belt. If an integrated cooling and lubrication system can be developed for all components (i.e., battery, motor and transmission) in an EV powertrain, the ownership cost of the EV can be reduced significantly. Study by [62] have started to evaluate the possibility of integrating the cooling system for the CVT fluid and the motor's fluid, and this helps in making the powertrain system more compact and cost effective. Latest review on the lubrication for

EV powertrain can be accessed here [95,96] which explored the possibility of integrating the cooling system for all components of an EV powertrain. The literatures also evaluate various possible lubricants for specific EV powertrain that have different requirements than the conventional ICE powertrain.

The subsequent research area that worth studying in the future is the possibility to implement the same EV powertrain configuration for diverse vehicle segments for cost savings in production [97,98]. As highlighted in [99], CVTs with metal belts or chains present the opportunity for application in various EV segments thanks to their continuous ratio range. However, the powertrain’s performance when applied in different segments have to be properly analyzed so that the advantages and disadvantages of its application can be determined. Besides, certain modifications on this type of CVT must also be studied, since different segments normally involve different motor power requirements, which will require different specifications for the metal belt. To address this, the possibility of using geartrain-based CVT should be explored, which can eliminate the belt’s application.

The final research area proposed here is the implementation of a holistic eco-driving method. According to [100], the fundamental aspect of eco-driving is to maximize the constant vehicle speed range so that losses in the powertrain can be reduced, which based on the study can be reduced by as much as 27% depending on the vehicle segments. A holistic eco-driving method, for the future study, should consider not only maximizing the constant vehicle speed range, but also optimizing the motor’s efficiency range, minimizing the transmission power losses, optimizing regenerative braking, as well as maximizing the battery’s health and durability without compromising much on the driving comfort [101]. Table 7 summarizes the potential key research areas that can be pursued in optimizing EV powertrains in the near future

Table 7. Expected key research areas on optimizing power flow in EV powertrain.

From 2018 to Early 2022	2022 Onwards
<ul style="list-style-type: none"> - Optimizing the gear ratios for powertrain efficiency and performance using advanced optimization techniques. Highlighted the importance of determining the appropriate gears’ size for optimum efficiency and driving performance. - Optimizing gearshifting strategy by taking into account the road conditions and driver’s input for the powertrain efficiency and performance. - Preliminary works on the gearshifting mechanism and controls. Here, only evaluating the mechanism in terms of jerking, without being integrated into the multispeed transmission. - Optimizing and discretizing the CVT ratios for optimum power consumption without too frequent shifting based on hydraulic actuation system. - Optimizing the distribution of traction on the wheels (four-motor configuration) and the shifting of the speed and torque couplings (two-motor configuration) for efficiency and driving performance. - Implement Eco-driving strategy for optimum efficiency (power consumption) and driving performance by taking into account motor’s efficiency map, transmission losses, drivers’ behaviors and battery’s SOC. 	<ul style="list-style-type: none"> - Research on the control of the gearshifting mechanism are expected to be intensified, evaluating the jerking and efficiency of the powertrain completely. - Research in evaluating the performance and practicality of electro-mechanical actuation system in CVT for EV powertrain. - Research on the mechanism for mode shifting in the multi-motor configuration. - Research on the steering assists in the four-motor configuration for driving dynamics and safety. Also, durability and cost analysis of the configuration. - Research on the lubricant and cooling system for optimizing the powertrain performance and lifespan of the battery. Focus on the integrated lubrication and cooling system for multi-components in the powertrain. - Research on the detail maintenance cost. The objective is to determine the cost threshold acceptable to wider consumer demography. - Research on the possibility of applying single type of powertrain for wider vehicle segments for cost and sustainability. - Research on the implementation of holistic eco-driving strategy for not only optimum power consumption and driving performance, but also for driving safety and battery state of health. This can be achieved by considering the motor’s efficiency, transmission losses, drivers’ behaviors, battery’s SOC and temperature characteristics and surrounding traffic.

6. Conclusions

EV market penetration globally is expected to intensify in the near future thanks to their improved practicality, reduced ownership cost and governments' policy on emissions, among other factors. In one aspect, this development is expected to reduce greenhouse gas emissions from new vehicles. In other aspects, however, the increased EV market share also leads to new challenges such as production sustainability, excessive increase in electricity demand and technology migration issues. If not properly addressed, these challenges will cause not only excessive cost to the manufacturers and customers, but also potentially reverse the environmental gains from the reduced tailpipe emissions. Therefore, optimizing the power flow of the EV powertrain is the key to addressing those challenges, which can be divided into three methods: multi-speed discrete transmission, CVT and multi-motors configuration.

In this paper, the latest literatures on the three methods have been reviewed extensively in terms of the methodology and significant findings. Next, the methods are compared to assess their advantages and disadvantages. In short, multispeed discrete transmission, especially two-speed discrete transmission, features an advantageous compact design which makes it very practical for EV powertrains. As a result, the extra weight due to the inclusion of the transmission in an EV can be minimized, and the shifting strategy can be made simpler and more effective to avoid too frequent gearshifting that will compromise driving comfort. However, the two-speed discrete transmission lacks flexibility due to its limited number of gears, hence it is not practical for diverse driving modes and vehicle segments. In this aspect, CVTs and multi-motor configurations are more flexible, due to their continuous ratio range, and options for independent traction distribution on the wheels, respectively. Nevertheless, CVT suffers from significant losses in its hydraulic actuation system and belt, while multi-motors configuration requires advanced control algorithm to precisely distribute the wheels' traction, as well as extra cost due to the high number of motors being used.

From the review, several key research areas have been identified for the future study. The latest literature mostly focused on optimizing the gear ratios considering motor's efficiency and driving conditions (for multi-speed discrete transmission and CVT), optimizing the shifting strategy for diverse driving cycles (for multi-speed discrete transmission, CVT and multi-motors configuration), and optimizing the traction distribution on the wheels for reduced power consumption and improved vehicle dynamics (for multi-motor configurations). Thus, the identified key research areas are; optimizing the gearshifting mechanism and its control (for multi-speed transmission and multi-motors configurations that feature two-speed AMT), evaluating and optimizing the electro-mechanical actuation system for CVT, optimizing the wheels' traction distribution for steering assists and driving safety (for multi-motors configuration), optimized and integrated lubrication and cooling system for all EV powertrain's components, detailed cost and environmental assessments of their application in EV, and finally, implementation of advanced eco-driving strategy considering not only motor's efficiency, but also transmission losses and battery SOC. These areas are crucial for optimizing EV powertrains' efficiency and performance for a more sustainable and cost-effective EV.

Author Contributions: Provision of resources in terms of the relevant literatures on EV powertrain, I.I.M., Z.H.C.D., M.K.A.H., V.T. and A.J.; Research and review on the literatures related to EV powertrain optimization using multispeed discrete transmission, I.I.M., M.S.C.K. and M.H.A.T.; Research and review on the literatures related to EV powertrain optimization using continuously variable transmission, I.I.M., Z.H.C.D. and M.S.C.K.; Research and review on the literatures related to EV powertrain optimization using multi-motor configuration, I.I.M., P.M.S., K.A.I. and W.X.; Research project supervision, I.I.M., Z.H.C.D. and M.K.A.H.; Writing of the original draft, I.I.M., V.T. and A.J.; Reviewing and editing of the final article, I.I.M., P.M.S. and M.H.A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Malaysian Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (FRGS/1/2021/TK0/UTM/02/44).

Informed Consent Statement: Not applicable.

Acknowledgments: We are grateful Opia Anthony Chukwunonso for Reviewing and editing of the final article.

Conflicts of Interest: The authors declare no conflict of interest in terms of financial and personal that could influence the work presented in this paper.

References

- Hamilton, J.; Walton, B.; Ringrow, J.; Albert, G.; Smith, S.F.; Day, E. *Electric Vehicles Setting a Course for 2030*; Deloitte University: Diegem, Belgium, 2020.
- Sisani, F.; Di Maria, F.; Cesari, D. Environmental and human health impact of different powertrain passenger cars in a life cycle perspective. A focus on health risk and oxidative potential of particulate matter components. *Sci. Total Environ.* **2022**, *805*, 150171. [[CrossRef](#)]
- Hung, C.R.; Völler, S.; Agez, M.; Bettez, G.M.; Strømman, A.H. Regionalized climate footprints of battery electric vehicles in Europe. *J. Clean. Prod.* **2021**, *322*, 129052. [[CrossRef](#)]
- Sathre, R.; Gustavsson, L. A lifecycle comparison of natural resource use and climate impact of biofuel and electric cars. *Energy* **2021**, *237*, 121546. [[CrossRef](#)]
- Gerossier, A.; Girard, R.; Kariniotakis, G. Modeling and Forecasting Electric Vehicle Consumption Profiles. *Energies* **2019**, *12*, 1341. [[CrossRef](#)]
- Yu, M.; Bai, B.; Xiong, S.; Liao, X. Evaluating environmental impacts and economic performance of remanufacturing electric vehicle lithium-ion batteries. *J. Clean. Prod.* **2021**, *321*, 128935. [[CrossRef](#)]
- VW Plots New German Electric Car Factory to Counter Tesla. Available online: <https://www.bloomberg.com/news/articles/2021-11-09/vw-plots-new-german-electric-car-factory-to-counter-tesla> (accessed on 17 November 2021).
- Logan, K.G.; Nelson, J.D.; Brand, C.; Hastings, A. Phasing in electric vehicles: Does policy focusing on operating emission achieve net zero emissions reduction objectives? *Transp. Res. Part A Policy Pract.* **2021**, *152*, 100–114. [[CrossRef](#)]
- Morfeldt, J.; Kurland, S.D.; Johansson, D.J. Carbon footprint impacts of banning cars with internal combustion engines. *Transp. Res. Part D Transp. Environ.* **2021**, *95*, 102807. [[CrossRef](#)]
- Hoefl, F. Internal combustion engine to electric vehicle retrofitting: Potential customer's needs, public perception and business model implications. *Transp. Res. Interdiscip. Perspect.* **2021**, *9*, 100330. [[CrossRef](#)]
- Zhang, M.; Yang, K.; Duan, S.; Liu, H.; Gao, F.; Geng, M. Thermal stability of high energy density LiNi_{0.815}Co_{0.15}Al_{0.035}O₂/Li₄Ti₅O₁₂ battery. *High Volt. Eng.* **2017**, *43*, 2221–2228.
- Barkhotlz, H.; Preger, Y.; Ivanov, S.; Langendorf, J.; Castro, L.; Lamb, J.; Chalamala, B.; Ferreira, S. Multi-scale thermal stability study of commercial lithium-ion batteries as a function of cathode chemistry and state-of-charge. *J. Power Sources* **2019**, *435*, 226777. [[CrossRef](#)]
- Daud, Z.H.C.; Asus, Z.; Bakar, S.A.A.; Husain, N.A.; Mazali, I.I.; Chrenko, D. Thermal characteristics of a lithium-ion battery used in a hybrid electric vehicle under various driving cycles. *IET Electr. Syst. Transp.* **2020**, *10*, 243–248. [[CrossRef](#)]
- Tete, P.R.; Gupta, M.M.; Joshi, S.S. Developments in battery thermal management systems for electric vehicles: A technical review. *J. Energy Storage* **2021**, *35*, 102255. [[CrossRef](#)]
- Zhao, C.; Zhang, B.; Zheng, Y.; Huang, S.; Yan TLiu, X. Hybrid Battery Thermal Management System in Electrical Vehicles: A Review. *Energies* **2020**, *13*, 6257. [[CrossRef](#)]
- Hwang, J.-Y.; Park, S.-J.; Yoon, C.S.; Sun, Y.-K. Customizing a Li-metal battery that survives practical operating conditions for electric vehicle applications. *Energy Environ. Sci.* **2019**, *12*, 2174–2184. [[CrossRef](#)]
- Ibrahim, A.; Jiang, F. The electric vehicle energy management: An overview of the energy system and related modeling and simulation. *Renew. Sustain. Energy Rev.* **2021**, *144*, 111049. [[CrossRef](#)]
- Berg, H.; Zackrisson, M. Perspectives on environmental and cost assessment of lithium metal negative electrodes in electric vehicle traction batteries. *J. Power Sources* **2019**, *415*, 83–90. [[CrossRef](#)]
- Funke, S.; Plötz, P.; Wietschel, M. Invest in fast-charging infrastructure or in longer battery ranges? A cost-efficiency comparison for Germany. *Appl. Energy* **2019**, *235*, 888–899. [[CrossRef](#)]
- Husain, I.; Ozpineci, B.; Islam, S.; Gurpinar, E.; Su, G.-J.; Yu, W.; Chowdhury, S.; Xue, L.; Rahman, D.; Sahu, R. Electric Drive Technology Trends, Challenges, and Opportunities for Future Electric Vehicles. *Proc. IEEE* **2020**, *109*, 1039–1059. [[CrossRef](#)]
- Kimiabeigi, M.; Widmer, J.D.; Long, R.; Gao, Y.; Goss, J.; Martin, R.; Lisle, T.; Vizan, J.M.S.; Michaelides, A.; Mecrow, B. High-Performance Low-Cost Electric Motor for Electric Vehicles Using Ferrite Magnets. *IEEE Trans. Ind. Electron.* **2015**, *63*, 113–122. [[CrossRef](#)]
- El-Refaie, A.; Osama, M. High specific power electrical machines: A system perspective. *CES Trans. Electr. Mach. Syst.* **2019**, *3*, 88–93. [[CrossRef](#)]

23. Wen, J.; Zhao, D.; Zhang, C. An overview of electricity powered vehicles: Lithium-ion battery energy storage density and energy conversion efficiency. *Renew. Energy* **2020**, *162*, 1629–1648. [[CrossRef](#)]
24. Chen, Y.-S.; Chen, I.-M.; Liu, T. A design approach for multi-configuration hybrid transmission mechanisms. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2020**, *234*, 2744–2758. [[CrossRef](#)]
25. Zhang, B.; Zhang, J.; Shen, T. Optimal control design for comfortable-driving of hybrid electric vehicles in acceleration mode. *Appl. Energy* **2021**, *305*, 117885. [[CrossRef](#)]
26. Krüger, B.; Keinprecht, G.; Filomeno, G.; Dennin, D.; Tenberge, P. Design and optimisation of single motor electric powertrains considering different transmission topologies. *Mech. Mach. Theory* **2022**, *168*, 104578. [[CrossRef](#)]
27. Hinov, N.; Punov, P.; Gilev, B.; Vacheva, G. Model-Based Estimation of Transmission Gear Ratio for Driving Energy Consumption of an EV. *Electronics* **2021**, *10*, 1530. [[CrossRef](#)]
28. Ritari, A.; Vepsalainen, J.; Kivekas, K.; Tammi, K.; Laitinen, H. Energy Consumption and Lifecycle Cost Analysis of Electric City Buses with Multispeed Gearboxes. *Energies* **2020**, *13*, 2117. [[CrossRef](#)]
29. Lin, C.; Zhao, M.; Pan, H.; Yi, J. Blending gear shift strategy design and comparison study for a battery electric city bus with AMT. *Energy* **2019**, *185*, 1–14. [[CrossRef](#)]
30. Tian, Y.; Ruan, J.; Zhang, N.; Wu, J.; Walker, P. Modelling and control of a novel two-speed transmission for electric vehicles. *Mech. Mach. Theory* **2018**, *127*, 13–32. [[CrossRef](#)]
31. Tian, Y.; Zhang, N.; Zhou, S.; Walker, P.D. Model and gear shifting control of a novel two-speed transmission for battery electric vehicles. *Mech. Mach. Theory* **2020**, *152*, 103902. [[CrossRef](#)]
32. Roozegar, M.; Angeles, J. Gear-shifting in a novel modular multi-speed transmission for electric vehicles using linear quadratic integral control. *Mech. Mach. Theory* **2018**, *128*, 359–367. [[CrossRef](#)]
33. Ahssan, R.; Ektesabi, M.; Gorji, S. Gear Ratio Optimization along with a Novel Gearshift Scheduling Strategy for a Two-Speed Transmission System in Electric Vehicle. *Energies* **2020**, *13*, 5073. [[CrossRef](#)]
34. Tan, S.; Yang, J.; Zhao, X.; Hai, T.; Zhang, W. Gear Ratio Optimization of a Multi-Speed Transmission for Electric Dump Truck Operating on the Structure Route. *Energies* **2018**, *11*, 1324. [[CrossRef](#)]
35. Wang, F.; Ye, P.; Xu, X.; Cai, Y.; Ni, S.; Que, H. Novel regenerative braking method for transient torsional oscillation suppression of planetary-gear electrical powertrain. *Mech. Syst. Signal Process.* **2022**, *163*, 108187. [[CrossRef](#)]
36. Kwon, K.; Jo, J.; Min, S. Multi-objective gear ratio and shifting pattern optimization of multispeed transmissions for electric vehicles considering variable transmission efficiency. *Energy* **2021**, *236*, 121419. [[CrossRef](#)]
37. Han, J.O.; Jeong, W.H.; Lee, J.S.; Oh, S.H. The Structure and Optimal Gear Tooth Profile Design of Two-Speed Transmission for Electric Vehicles. *Energies* **2021**, *14*, 3736. [[CrossRef](#)]
38. Li, Y.; Zhu, B.; Zhang, N.; Peng, H.; Chen, Y. Parameters optimization of two-speed powertrain of electric vehicle based on genetic algorithm. *Adv. Mech. Eng.* **2020**, *12*, 1–16. [[CrossRef](#)]
39. Han, K.; Wang, Y.; Filev, D.; Dai, E.; Kolmanovsky, I.; Girard, A. Optimized Design of Multi-Speed Transmissions for Battery Electric Vehicles. In Proceedings of the 2019 American Control Conference, Philadelphia, PA, USA, 10–12 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 816–821.
40. Spanoudakis, P.; Moschopoulos, G.; Stefanoulis, T.; Sarantinoudis, N.; Papadokokolakis, E.; Ioannou, I.; Piperidis, S.; Doitsidis, L.; Tsourveloudis, N.C. Efficient Gear Ratio Selection of a Single-Speed Drivetrain for Improved Electric Vehicle Energy Consumption. *Sustainability* **2020**, *12*, 9254. [[CrossRef](#)]
41. Mo, W.; Wu, J.; Walker, P.D.; Zhang, N. Shift characteristics of a bilateral Harpoon-shift synchronizer for electric vehicles equipped with clutchless AMTs. *Mech. Syst. Signal Process.* **2021**, *148*, 107166. [[CrossRef](#)]
42. Mo, W.; Walker, P.D.; Zhang, N. Dynamic analysis and control for an electric vehicle with harpoon-shift synchronizer. *Mech. Mach. Theory* **2019**, *133*, 750–766. [[CrossRef](#)]
43. Mo, W.; Walker, P.D.; Fang, Y.; Wu, J.; Ruan, J.; Zhang, N. A novel shift control concept for multi-speed electric vehicles. *Mech. Syst. Signal Process.* **2018**, *112*, 171–193. [[CrossRef](#)]
44. Beaudoin, M.-A.; Boulet, B. Fundamental limitations to no-jerk gearshifts of multi-speed transmission architectures in electric vehicles. *Mech. Mach. Theory* **2021**, *160*, 104290. [[CrossRef](#)]
45. Wang, Y.; Wu, J.; Zhang, N.; Mo, W. Dynamics modeling and shift control of a novel spring-based synchronizer for electric vehicles. *Mech. Mach. Theory* **2022**, *168*, 104586. [[CrossRef](#)]
46. Liu, Y.; Lin, Z.; Zhao, K.; Ye, J.; Huang, X. Multiobjective gearshift optimization with Legendre pseudospectral method for seamless two-speed transmission. *Mech. Mach. Theory* **2020**, *145*, 103682. [[CrossRef](#)]
47. Hong, J.; Gao, B.; Yue, H.; Chen, H. Dry Clutch Control of Two-Speed Electric Vehicles by Using an Optimal Control Scheme with Persistent Time-Varying Disturbance Rejection. *IEEE Trans. Transp. Electrification* **2021**, *7*, 2034–2046. [[CrossRef](#)]
48. Yue, H.; Zhu, C.; Gao, B. Fork-less two-speed I-AMT with overrunning clutch for light electric vehicle. *Mech. Mach. Theory* **2018**, *130*, 157–169. [[CrossRef](#)]
49. Ogawa, K.; Aihara, T. Development of two-speed dual-clutch transmission for seamless gear shifting in EVs. *Transp. Eng.* **2021**, *6*, 100097. [[CrossRef](#)]
50. Walker, P.; Zhu, B.; Zhang, N. Powertrain dynamics and control of a two speed dual clutch transmission for electric vehicles. *Mech. Syst. Signal Process.* **2017**, *85*, 1–15. [[CrossRef](#)]

51. Lin, X.; Li, Y.; Xia, B. An online driver behavior adaptive shift strategy for two-speed AMT electric vehicle based on dynamic corrected factor. *Sustain. Energy Technol. Assess.* **2021**, *48*, 101598.
52. Gao, B.; Cai, K.; Qu, T.; Hu, Y.; Chen, H. Personalized adaptive cruise control based on online driving style recognition technology and model predictive control. *IEEE Trans. Veh. Technol.* **2020**, *69*, 12482–12496. [[CrossRef](#)]
53. Guo, Q.Y.; Zhao, Z.G.; Shen, P.H.; Zhan, X.W.; Li, J.W. Adaptive optimal control based on driving style recognition for plug-in hybrid electric vehicle. *Energy* **2019**, *186*, 115824. [[CrossRef](#)]
54. Lin, X.Y.; Zeng, S.R.; Li, X.F. Online correction predictive energy management strategy using the Q-learning based swarm optimization with fuzzy neural network. *Energy* **2021**, *223*, 120071. [[CrossRef](#)]
55. Fernandes, F.M.C. Fuzzy controller applied to electric vehicles with continuously variable transmission. *Neurocomputing* **2016**, *214*, 684–691. [[CrossRef](#)]
56. Hofman, T.; Janssen, N. Integrated Design Optimization of the Transmission System and Vehicle Control for Electric Vehicles. In Proceedings of the 20th IFAC World Conference, Toulouse, France, 9–14 July 2017; International Federation of Automatic Control: Kidlington Oxford, UK, 2017; Volume 50, pp. 10072–10077.
57. Ruan, J.; Walker, P.; Wu, J.; Zhang, N.; Zhang, B. Development of continuously variable transmission and multi-speed dual-clutch transmission for pure electric vehicle. *Adv. Mech. Eng.* **2018**, *10*, 1–15. [[CrossRef](#)]
58. Ruan, J.; Walker, P.; Zhang, N. Comparison of Power Consumption Efficiency of CVT and Multi-Speed Transmissions for Electric Vehicle. *Int. J. Automot. Eng.* **2018**, *9*, 268–275. [[CrossRef](#)]
59. Hofman, T.; Salazar, M. Transmission Ratio Design for Electric Vehicles via Analytical Modeling and optimization. In Proceedings of the 2020 IEEE Vehicle Power and Propulsion Conference, Gijon, Spain, 18 November–16 December 2020; pp. 1–6.
60. Van Der Sluis, F.; Romers, L.; Van Spijk, G.-J.; Hupkes, I. CVT, Promising Solutions for Electrification. In Proceedings of the WCX SAE World Congress Experience, Detroit, MI, USA, 9–11 April 2019. [[CrossRef](#)]
61. Hu, J.; Mei, B.; Peng, H.; Guo, Z. Discretely variable speed ratio control strategy for continuously variable transmission system considering hydraulic energy loss. *Energy* **2019**, *180*, 714–727. [[CrossRef](#)]
62. Wei, C.; Hofman, T.; Caarls, E.I. Co-Design of CVT-Based Electric Vehicles. *Energies* **2021**, *14*, 1825. [[CrossRef](#)]
63. Ruan JSong, Q.; Yang, W. The application of hybrid energy storage system with electrified continuously variable transmission in battery electric vehicle. *Energy* **2019**, *183*, 315–330. [[CrossRef](#)]
64. Song, Z.; Hofmann, H.; Li, J.; Han, X.; Ouyang, M. Optimization for a hybrid energy storage system in electric vehicles using dynamic programming approach. *Appl. Energy* **2015**, *139*, 151–162. [[CrossRef](#)]
65. Rezvanizani, S.M.; Liu, Z.; Chen, Y.; Lee, J. Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility. *J. Power Sources* **2014**, *256*, 110–124. [[CrossRef](#)]
66. Noura, N.; Boulon, L.; Jemei, S. A Review of Battery State of Health Estimation Methods: Hybrid Electric Vehicle Challenges. *World Electr. Veh. J.* **2020**, *11*, 66. [[CrossRef](#)]
67. Liao, P.; Tang, T.-Q.; Liu, R.; Huang, H.-J. An eco-driving strategy for electric vehicle based on the powertrain. *Appl. Energy* **2021**, *302*, 117583. [[CrossRef](#)]
68. Akehurst, S.; Vaughan, N.D.; Parker, D.A.; Simner, D. Modeling of loss mechanisms in a pushing metal V-belt continuously variable transmission, part 1: Torque losses due to band friction. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2004**, *218*, 1269–1281. [[CrossRef](#)]
69. Akehurst, S.; Vaughan, N.D.; Parker, D.A.; Simner, D. Modeling of loss mechanisms in a pushing metal V-belt continuously variable transmission, part 2: Pulley deflection losses and total torque loss validation. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2004**, *218*, 1283–1293. [[CrossRef](#)]
70. Akehurst, S.; Vaughan, N.D.; Parker, D.A.; Simner, D. Modeling of loss mechanisms in a pushing metal V-belt continuously variable transmission, part 3: Belt slip losses. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2004**, *218*, 1295–1306. [[CrossRef](#)]
71. Peng, H.; Qin, D.; Hu, J.; Fu, C. Synthesis and analysis method for powertrain configuration of single motor hybrid electric vehicle. *Mech. Mach. Theory* **2020**, *146*, 103731. [[CrossRef](#)]
72. Zhuang, W.; Li, S.; Zhang, X.; Kum, D.; Song, Z.; Yin, G.; Ju, F. A survey of powertrain configuration studies on hybrid electric vehicles. *Appl. Energy* **2020**, *262*, 114553. [[CrossRef](#)]
73. Mantriota, G.; Reina, G.; Ugenti, A. Performance Evaluation of a Compound Power-Split CVT for Hybrid Powertrains. *Appl. Sci.* **2021**, *11*, 8749. [[CrossRef](#)]
74. Zheng, Q.; Tian, S.; Zhang, Q. Optimal Torque Split Strategy of Dual-Motor Electric Vehicle Using Adaptive Nonlinear Particle Swarm Optimization. *Math. Probl. Eng.* **2020**, *2020*, 1204260. [[CrossRef](#)]
75. Wu, J.; Zhang, N. Driving mode shift control for planetary gear based dual motor powertrain in electric vehicles. *Mech. Mach. Theory* **2020**, *158*, 104217. [[CrossRef](#)]
76. Wu, J.; Liang, J.; Ruan, J.; Zhang, N.; Walker, P. A robust energy management strategy for EVs with dual input power-split transmission. *Mech. Syst. Signal Process.* **2018**, *111*, 442–455. [[CrossRef](#)]
77. Wu, J.; Liang, J.; Ruan, J.; Zhang, N.; Walker, P.D. Efficiency comparison of electric vehicles powertrains with dual motor and single motor input. *Mech. Mach. Theory* **2018**, *128*, 569–585. [[CrossRef](#)]
78. Du, W.; Zhao, S.; Jin, L.; Gao, J.; Zheng, Z. Optimization design and performance comparison of different powertrains of electric vehicles. *Mech. Mach. Theory* **2021**, *156*, 104143. [[CrossRef](#)]

79. Hong, X.; Wu, J.; Zhang, N.; Wang, B.; Tian, Y. The dynamic and economic performance study of a new Simpson planetary gearset based dual motor powertrain for electric vehicles. *Mech. Mach. Theory* **2022**, *167*, 104579. [CrossRef]
80. Nguyen, C.T.; Walker, P.D.; Zhang, N. Optimization and coordinated control of gear shift and mode transition for a dual-motor electric vehicle. *Mech. Syst. Signal Process.* **2021**, *158*, 107731. [CrossRef]
81. Nguyen, C.T.; Walker, P.D.; Zhou, S.; Zhang, N. Optimal sizing and energy management of an electric vehicle power-train equipped with two motors and multi-gear ratios. *Mech. Mach. Theory* **2022**, *167*, 104513. [CrossRef]
82. Li, Z.; Khajepour, A.; Song, A.J. Comprehensive review of the key technologies for pure electric vehicles. *Energy* **2019**, *182*, 824–839. [CrossRef]
83. Kwon, K.; Seo, M.; Min, S. Efficient multi-objective optimization of gear ratios and motor torque distribution for electric vehicles with two-motor and two-speed powertrain system. *Appl. Energy* **2020**, *259*, 114190. [CrossRef]
84. Meng, D.; Tian, M.; Miao, L.; Wang, Y.; Hu, J.; Gao, B. Design and modeling of an in-wheel two-speed AMT for electric vehicles. *Mech. Mach. Theory* **2021**, *163*, 104383. [CrossRef]
85. Meng, D.; Wang, F.; Wang, Y.; Gao, B. In-Wheel Two-Speed AMT with Selectable One-Way Clutch for Electric Vehicles. *Actuators* **2021**, *10*, 220. [CrossRef]
86. Yamamoto, S.; Morita, R.; Oike, M. Transmission-equipped wheel hub motor for passenger cars. *ATZ Worldw.* **2018**, *120*, 28–33. [CrossRef]
87. Wei, H.; Ai, Q.; Zhao, W.; Zhang, Y. Modelling and experimental validation of an EV torque distribution strategy towards active safety and energy efficiency. *Energy* **2022**, *239*, 121953. [CrossRef]
88. Miranda, M.H.R.; Silva, F.L.; Lorencó, M.A.M.; Eckert, J.J.; Silva, L.C.A. Electric vehicle powertrain and fuzzy controller optimization using a planar dynamics simulation based on a real-world driving cycle. *Energy* **2022**, *238*, 121979. [CrossRef]
89. Yildiz, A.; Oezil, M.A. A Comparative Study of Energy Consumption and Recovery of Autonomous Fuel-Cell Hydrogen-Electric Vehicles Using Different Powertrains Based on Regenerative Braking and Electronic Stability Control System. *Appl. Sci.* **2021**, *11*, 2515. [CrossRef]
90. Ahssan, R.; Ektesabi, M.M.; Gorji, S.A. Electric Vehicle with Multi-Speed Transmission: A Review on Performances and Complexities. *SAE Int. J. Altern. Powertrains* **2018**, *7*, 169–181. [CrossRef]
91. Mazali, I.I.; Husain, N.A.; Tawi, K.B.; Daud, Z.H.C.; Kob, M.S.C.; Asus, Z. Review of latest technological advancement in electro-mechanical continuously variable transmission. *Int. J. Veh. Des.* **2019**, *81*, 166–190. [CrossRef]
92. Liu, J.; Sun, D.; Ye, M.; Liu, X.; Li, B. Study on the transmission efficiency of electro-mechanical continuously variable transmission with adjustable clamping force. *Mech. Mach. Theory* **2018**, *126*, 468–478. [CrossRef]
93. Liu, Z.; Song, J.; Kubal, J.; Susarla, N.; Knehr, K.W.; Islam, E.; Nelson, P.; Ahmed, S. Comparing total cost of ownership of battery electric vehicles and internal combustion engine vehicles. *Energy Policy* **2021**, *158*, 112564. [CrossRef]
94. Machado, F.A.; Kollmeyer, P.; Barroso, D.G.; Emadi, A. Multi-Speed Gearboxes for Battery Electric Vehicles: Current Status and Future Trends. *IEEE Open J. Veh. Technol.* **2021**, *2*, 419–435. [CrossRef]
95. Shah, R.; Gashi, B.; González-Poggini, S.; Colet-Lagrille, M.; Rosenkranz, A. Recent trends in batteries and lubricants for electric vehicles. *Adv. Mech. Eng.* **2021**, *13*, 1–10. [CrossRef]
96. Cabrera, L.I.S. Tribology of electric vehicles: A review of critical components, current state and future improvement trends. *Tribol. Int.* **2019**, *138*, 473–486. [CrossRef]
97. Verbruggen, F.J.R.; Silvas, E.; Hofman, T. Electric Powertrain Topology Analysis and Design for Heavy-Duty Trucks. *Energies* **2020**, *13*, 2434. [CrossRef]
98. Jones, B.; Elliott, R.J.; Nguyen-Tien, V. The EV revolution: The road ahead for critical raw materials demand. *Appl. Energy* **2020**, *280*, 115072. [CrossRef] [PubMed]
99. Continuously Variable Transmission with Pushbelt for Electric Vehicles. For More Efficiency and Power Output in Electric Vehicles. Available online: <https://www.bosch-mobility-solutions.com/en/solutions/transmission-technology/transmission-cvt4ev/> (accessed on 27 December 2021).
100. Gao, Z.; LaClair, T.; Ou, S.; Huff, S.; Wu, G.; Hao, P.; Boriboonsomsin, B.; Barth, M. Evaluation of electric vehicle component performance over eco-driving cycles. *Energy* **2019**, *172*, 823–839. [CrossRef]
101. Han, U.; Kang, H.; Song, J.; Oh, J.; Lee, H. Development of dynamic battery thermal model integrated with driving cycles for EV applications. *Energy Convers. Manag.* **2021**, *250*, 114882. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Applied Sciences Editorial Office
E-mail: applsci@mdpi.com
www.mdpi.com/journal/applsci



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel: +41 61 683 77 34
www.mdpi.com



ISBN 978-3-0365-4858-6