



systems

Special Issue Reprint

Modelling and Simulation of Transportation Systems

Edited by
Eugen Rosca, Valentin Carlan and Florin Valentin Rusca

mdpi.com/journal/systems



Modelling and Simulation of Transportation Systems

Modelling and Simulation of Transportation Systems

Guest Editors

Eugen Rosca

Valentin Carlan

Florin Valentin Rusca



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors

Eugen Rosca
Transport Faculty
National University of
Science and Technology
Politehnica Bucharest
Bucharest
Romania

Valentin Carlan
Department of Transport and
Regional Economics
University of Antwerp
Antwerp
Belgium

Florin Valentin Rusca
Transport Faculty
National University for
Science and Technology
Politehnica Bucharest
Bucharest
Romania

Editorial Office

MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Systems* (ISSN 2079-8954), freely accessible at: https://www.mdpi.com/journal/systems/special_issues/O3LH7GN9RA.

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

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

ISBN 978-3-7258-8111-6 (Hbk)

ISBN 978-3-7258-8112-3 (PDF)

<https://doi.org/10.3390/books978-3-7258-8112-3>

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

About the Editors	vii
Preface	ix
Marcin Czaban, Eldar Sultanow, Alina Chircu, Christian Czarnecki, Joachim Riedl and Stefan Wengler	
User Interaction with Digital Twins for Driving: How Comparable Are Simulated and Real Trajectories?	
Reprinted from: <i>Systems</i> 2026 , <i>14</i> , 144, https://doi.org/10.3390/systems14020144	1
Anna Dolinayová, Zdenka Bulková, Jozef Gašparík and Igor Dömény	
Modelling and Forecasting Passenger Rail Demand in Slovakia Under Crisis Conditions with NARX Neural Networks	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 881, https://doi.org/10.3390/systems13100881	22
Vasile Dragu, Eugenia Alina Roman, Mircea Augustin Roșca, Floriana Cristina Oprea, Andrei-Bogdan Mironescu and Oana Maria Dinu	
Integrated Information System for Parking Facilities Operations and Management	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 769, https://doi.org/10.3390/systems13090769	57
Jinsong Zhang, Junyi Sha, Chunyu Zhang and Yijin Zhang	
A CNN-LSTM-GRU Hybrid Model for Spatiotemporal Highway Traffic Flow Prediction	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 765, https://doi.org/10.3390/systems13090765	83
Mircea Augustin Rosca, Floriana Cristina Oprea, Vasile Dragu, Oana Maria Dinu, Ilona Costea and Stefan Burciu	
Evaluating the Impact of Automated Vehicle Penetration on Intersection Traffic Flow: A Microsimulation-Based Approach	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 751, https://doi.org/10.3390/systems13090751	104
Duygu Yilmaz Eroglu	
Tariff-Sensitive Global Supply Chains: Semi-Markov Decision Approach with Reinforcement Learning	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 645, https://doi.org/10.3390/systems13080645	125
Eugen Rosca, Floriana Cristina Oprea, Anamaria Ilie, Stefan Burciu and Florin Rusca	
Automated Parcel Locker Configuration Using Discrete Event Simulation	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 613, https://doi.org/10.3390/systems13070613	151
Wentao Zhang and Taesung Hwang	
UAM Vertiport Network Design Considering Connectivity	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 607, https://doi.org/10.3390/systems13070607	173
Poorendra Ramlall, Ethan Jones and Subhradeep Roy	
Development of a Networked Multi-Participant Driving Simulator with Synchronized EEG and Telemetry for Traffic Research	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 564, https://doi.org/10.3390/systems13070564	197
Aixiu Hu, Ruiyun Huang, Yanqun Yang, Ibrahim El-Dimeery and Said M. Easa	
Length Requirements for Urban Expressway Work Zones' Warning and Transition Areas Based on Driving Safety and Comfort	
Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 525, https://doi.org/10.3390/systems13070525	216

Rongjun Cheng, Xudong An, Weiqi Fan and Dan Zhao Micro-Driving Behavior Analysis of Drivers in Congested and Conflict Environments Using Graph Theory Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 491, https://doi.org/10.3390/systems13060491	242
Seung-Wan Cho, Yeong-Hyun Lim, Seong-Hyeon Ju and Kyung-Min Seo Analyzing New Operation Strategy of Demand-Responsive Transports Using Discrete-Event Simulation Framework Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 303, https://doi.org/10.3390/systems13040303	259
Sergio Rojas-Blanco, Alberto Cerezo-Narváez, Sol Sáez-Martínez and Manuel Otero-Mateo Algorithmic Identification of Conflicting Traffic Lights: A Large-Scale Approach with a Network Conflict Matrix Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 290, https://doi.org/10.3390/systems13040290	289
Bojan Rupnik, Yuhong Wang and Tomaž Kramberger Hybrid Model for Motorway EV Fast-Charging Demand Analysis Based on Traffic Volume Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 272, https://doi.org/10.3390/systems13040272	311
Eugen Rosca, Florin Rusca, Valentin Carlan, Ovidiu Stefanov, Oana Dinu and Aura Rusca Assessing the Influence of Equipment Reliability over the Activity Inside Maritime Container Terminals Through Discrete-Event Simulation Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 213, https://doi.org/10.3390/systems13030213	330
Wenhan Lin, Congying Li, Weibin Zheng, Linwei Wang and Yanqun Yang Riding Risk: Factors Shaping Helmet Use Among Two-Wheeled Electric Vehicle Riders in Fuzhou, China Reprinted from: <i>Systems</i> 2025 , <i>13</i> , 171, https://doi.org/10.3390/systems13030171	352

About the Editors

Eugen Rosca

Eugen Rosca is full professor at National University of Science and Technology Politehnica Bucharest and director of the Transport, Traffic and Logistics Department. He works at the intersection of transportation engineering, logistics, econometrics, and computational decision support. His research focuses on applying advanced modelling and simulation methods to improve the efficiency, and resilience of transportation systems, with contributions spanning both freight operations and urban mobility performance analysis. Through his work, he supports the development of actionable modelling frameworks that translate computational experiments into practical insights for researchers, planners, and transport and logistics practitioners.

Valentin Carlan

Valentin Carlan is a guest professor of digitalization and innovation in transport and logistics at the University of Antwerp, Faculty of Business and Economics, Department of Transport and Regional Economics. His work is situated at the intersection of transport economics, logistics management, and digital innovation, with a strong focus on the economic evaluation, governance, and implementation of advanced digital solutions, including artificial intelligence, data platforms, and decision-support systems in freight transport and port operations. His research critically examines how digital technologies create value, reduce inefficiencies, and support sustainable transformation across transport and logistics systems. Through his academic work and close collaboration with industry and public stakeholders, he contributes to the development of analytically robust and practice-oriented frameworks that inform research agendas, policy design, and managerial decision-making in transport and logistics.

Florin Valentin Rusca

Florin Valentin Rusca is full professor and PhD supervisor at the National University of Science and Technology Politehnica Bucharest. His research activity is situated in the field of transportation systems engineering, with a strong focus on traffic flow theory, modelling and simulation of transport processes, and intelligent transport systems. He addresses transport systems as complex socio-technical systems, analyzing performance, reliability, sustainability, and resilience under varying demand and operational conditions. His work supports system-level assessment, data-driven decision support, and integrated approaches for the design and management of intelligent and sustainable mobility systems in urban and interurban networks.

Preface

The reprint, entitled “Modelling and Simulation of Transportation Systems,” brings together a collection of advanced research contributions that demonstrate how computational modelling, simulation, and data-driven analytics can be used to investigate transportation problems across multiple modes and scales. The subject of the volume is the development and application of modern modelling approaches capable of representing transport systems realistically under uncertainty, supporting decisions that improve safety, efficiency, resilience, and sustainability.

The volume spans key areas of contemporary transportation research and practice, including traffic safety and human behaviour, urban traffic operations and management, freight logistics and supply chain performance, infrastructure planning for electrification and automation, and emerging mobility systems. The papers collectively reflect the diversity of modern transport challenges, while also emphasising a unifying principle: transport systems cannot be effectively managed with isolated indicators or static assumptions. Instead, they require integrated models that capture interactions between infrastructure, users, operational rules, demand patterns, and technology adoption processes. The purpose is twofold: first, to advance scientific understanding through robust modelling formulations and evaluation strategies; second, to support practical transport decisions through measurable outcomes such as safety performance, waiting time reduction, flow stability, resource utilisation, resilience under disruption, and improved forecasting accuracy.

One core thematic group focuses on traffic safety and the behavioural mechanisms that shape risk in urban and highway environments. By applying large-scale behavioural analysis, simulator-based testing and digital-twin technology, these works contribute to the design of more targeted and context-aware safety measures. In parallel, microscopic behaviour modelling frameworks capture how risky driving patterns unfold over time, allowing conflicts to be interpreted as structured processes rather than random events. This type of modelling is particularly relevant for the development of predictive safety systems and advanced warning strategies that can anticipate hazardous trajectories before they result in collisions.

A complementary set of contributions addresses the design of safer road environments through simulation-based evaluation of infrastructure configurations. Work zones on high-speed urban corridors, for example, remain critical points of elevated risk because drivers must process information and adapt to changing geometry under time pressure. Simulation-based experimentation enables the testing of alternative warning and transition layouts without exposing road users to real danger.

Another major theme of the volume concerns urban traffic operations and the management of complex networks. Modern cities increasingly require scalable tools capable of supporting large traffic control systems, coordinating signalized intersections, and improving corridor-level performance. Within this scope, algorithmic modelling approaches are introduced to structure and automate tasks such as identifying conflicting signal movements at scale, enabling more efficient optimisation and more effective use of computational resources. The reprint also explores simulation-based assessment of mobility services that respond to changing travel demand. Simulation frameworks allow authorities and operators to evaluate alternative service designs and operational rules under different demand and fleet conditions, supporting the development of adaptive services that can meet diverse travel needs while maintaining acceptable service quality. Parking management has measurable impacts on congestion, fuel consumption, and emissions due to internal circulation and unnecessary vehicle movements. Simulation-based evaluation of alternative

allocation rules provides a structured way to compare strategies and improve facility operations.

Logistics and freight transport represent a further critical dimension of the reprint. Supply chains depend on reliability, yet they are vulnerable to equipment failures, bottlenecks, and operational variability. Discrete-event simulation emerges across the volume as a powerful tool for representing freight operations realistically, capturing resource constraints, stochastic disruptions, and process interdependencies. By quantifying how failures and delays influence system performance, simulation supports proactive strategies such as preventive maintenance planning, improved resource allocation, and operational resilience enhancement. Last-mile distribution challenges are also addressed through modelling and simulation of automated delivery infrastructure, where service quality and investment efficiency must be balanced. Simulation-based optimisation enables operators to evaluate automated locker sizing and operational trade-offs, including the less intuitive effects of flexibility and interdependence between locker units. Such research supports city logistics planning and contributes to more sustainable and efficient parcel delivery networks. The reprint further extends the logistics perspective to global supply chain decision making under policy and market uncertainty. International sourcing and inventory choices are shaped by external conditions such as tariffs, currency variability, and shifting trade environments. Reinforcement-learning-based frameworks provide adaptive decision tools that can respond dynamically to such volatility, demonstrating how transport and logistics modelling increasingly overlaps with advanced computational intelligence.

An essential motivation for the volume is the ongoing transition toward electrification and automation, which creates urgent planning challenges for infrastructure and traffic management. Electric vehicle adoption, in particular, places new demands on charging networks, requiring robust estimation of charging needs and operational performance under varying traffic intensity and adoption scenarios. Hybrid modelling approaches enable detailed analysis of charger utilisation, queue formation, and energy demand. This supports strategic infrastructure investment and provides a foundation for future integration into digital-twin systems capable of assisting real-time management.

A further methodological direction in the volume is the use of deep learning and advanced predictive analytics for traffic flow forecasting. Accurate prediction of spatiotemporal traffic patterns is fundamental for proactive congestion management, dynamic control strategies, and informed infrastructure planning. Hybrid architectures that integrate spatial correlations between detectors with temporal dependencies offer improved forecasting performance, especially when external factors such as meteorological conditions are incorporated. Such research highlights how modelling and simulation increasingly operate alongside machine learning to improve transport intelligence and operational readiness.

The scope of the reprint also includes emerging transport systems such as urban air mobility. The planning of vertiport networks introduces new modelling needs where infrastructure location decisions must satisfy cost considerations while ensuring operational connectivity and network viability. Optimisation frameworks designed for such problems illustrate how the next generation of transport systems will require advanced decision-support methods from the earliest planning stages, long before widespread deployment occurs.

Finally, the volume recognises that transportation systems must be resilient not only to routine variability, but also to crisis conditions that disrupt demand patterns and operational assumptions. Modelling and forecasting approaches applied to passenger rail demand under external restrictions demonstrate how neural network models with exogenous inputs can capture sharp ridership changes and support dynamic capacity adjustment strategies. This crisis-oriented modelling perspective

strengthens the role of simulation and forecasting as tools for adaptive planning, allowing transport operators to respond more efficiently and cost-effectively to sudden disruptions.

This reprint is addressed to a broad community of readers who engage with transportation systems from complementary angles. It is intended for researchers seeking validated modelling methods and new interdisciplinary directions; for engineers and practitioners responsible for traffic operations, infrastructure design, and logistics performance; for planners and policymakers designing regulations and investment strategies under uncertainty; and for technology developers working on intelligent transport systems, digital twins, and next-generation mobility services. The volume offers both methodological depth and application-driven insights, supporting the transfer of scientific results into operational and policy practice.

As Guest Editors, we thank all the contributors for their dedication in writing valuable papers that made this Special Issue possible. We also thank the large number of reviewers who dedicated their time and competence to improve the manuscripts. Their work increased the overall quality of the volume. We extend special thanks to the whole MDPI editorial team for their thorough work and commitment that made this reprint possible.

Eugen Rosca, Valentin Carlan, and Florin Valentin Rusca

Guest Editors

Article

User Interaction with Digital Twins for Driving: How Comparable Are Simulated and Real Trajectories?

Marcin Czaban ^{1,2,*}, Eldar Sultanow ³, Alina Chircu ⁴, Christian Czarnecki ⁵, Joachim Riedl ¹ and Stefan Wengler ¹

¹ Institute of Information Systems, University of Applied Sciences Hof, 95028 Hof, Germany

² Marketing and Innovation, University of Bayreuth, 95447 Bayreuth, Germany

³ Capgemini Deutschland GmbH, 81829 München, Germany

⁴ Anderson School of Management, University of New Mexico, Albuquerque, NM 87131, USA

⁵ Fachbereich 5—Elektrotechnik und Informationstechnik, FH Aachen-University of Applied Sciences, 52066 Aachen, Germany

* Correspondence: marcin.czaban.2@hof-university.de

Abstract

This paper investigates the physiological responses of individuals driving both on a real route and within a vehicle simulator designed as a digital twin of that route. The analysis of observed data patterns in stress response bio signals provides sufficient evidence of similarity to validating the driving simulation digital twin as a reliable replacement for real-world experiences in controlled and consistent settings, or when overall trends of physiological variables, rather than specific variable levels, are of interest. The findings also stress the need for optimizing the precision of digital twins in complex settings. This study introduces a time-series-based validation approach for driving digital twins by comparing continuous physiological trajectories between real and simulated driving.

Keywords: digital twins; driving simulation; physiological measurements; temporal comparison; time series; stress response

1. Introduction

Digital twins (DTs) require not only structural and functional similarity but also equivalence in human–system interaction dynamics. For driving digital twins, this implies that not only average behavioral or physiological responses must be comparable, but that the temporal evolution of these responses must follow similar trajectories in real and simulated environments. As a digital representation of a physical system, a digital twin enables users to interact with a virtual counterpart and to simulate real-world dynamics with increasing fidelity [1]. This technology offers clear benefits—reducing cost, risk, and development time while enabling scenario testing that would be impractical in the physical world [2,3]. Despite its growing relevance, an important question remains: to what extent can user interaction with a digital twin replicate the physiological responses of real-world interaction, including the variability, sequence and timing of these responses during the entire driving episode? Addressing this question is critical for validating the realism and applicability of digital twins in practice [4,5].

In this study, we focus on driving digital twins as a representative application area [6]. Driving simulators are frequently used in vehicle development because they offer advantages over real-world tests in terms of cost and environmental control, providing a standardized and safe testing environment. For a simulator and its corresponding route

to function as a true digital twin, the physiological responses of a person while driving in the simulator must correspond to those elicited during real driving. Only under such conditions can driving DTs be considered valid, as this comparability enables the use of DTs as reliable tools for driver training, safety testing, and vehicle development under realistic yet controlled conditions. A key differentiator of advanced DTs compared to traditional simulators is their focus on continuous, real-time, and often bidirectional data integration between the physical and virtual entity. Understanding if the digital twin reproduces the sequence and timing of physiological responses as the driver encounters specific events (e.g., intersections, lane changes, traffic density changes) is essential in validating simulations used in driving DTs. Thus, the following research question arises:

RQ: To what extent do time-resolved physiological trajectories during real-world driving and during interaction with a driving simulation digital twin exhibit temporal equivalence, and under which conditions can such a digital twin replace real-world driving tests?

To address this question, we compare participants' physiological responses while driving along a real route and within a simulator designed as its digital twin. Previous studies have examined this simulator–reality comparison in limited settings, for example on short route segments [7]. This study extends the primary analysis by Czaban and Himmels [8], which examined aggregated experimental data from a driving simulation using Bayesian dependent *t*-tests and repeated-measures ANOVAs to assess simulator validity using driving segment-wise means of physiological and cognitive indicators. The present paper constitutes a secondary re-analysis that differs fundamentally in its theoretical framing (DTs) and analytical approach (time series analysis of granular experimental data). This paper examines the role of driving simulators in DTs by comparing time-resolved physiological trajectories that are essential for DT functionality, rather than segment-wise means (which collapse rich temporal structures into single aggregate statistics). By analyzing normalized time series across entire drives, we preserve temporal dynamics and assess temporal equivalence, a requirement for real-time digital twin interaction. Taken together, these two papers address distinct but complementary research questions, employ different data structures, and contribute to related but distinct scientific debates.

Our findings indicate a certain degree of physiological equivalence between real and simulated driving, suggesting that, under certain conditions, interactions with a digital twin can reproduce key aspects of real-world experiences. The results contribute to understanding the physiological realism required for DTs to serve as valid substitutes in domains where real-world testing is costly, time-consuming, or unsafe [9].

2. Related Work

2.1. Digital Twins and Their Applications

Digital twins are virtual representations of physical entities—objects, processes, or people—that mirror their state and behavior across the lifecycle through data-driven models [10,11]. A defining characteristic of advanced DTs is continuous, dynamic, and bidirectional data exchange between the physical and virtual systems, enabling real-time synchronization, learning, and adaptation [1,10]. Through integration of real-time and historical data, DTs support continuous monitoring, adaptive interventions, and predictive modelling.

DT maturity levels range from digital models (offline simulations), to digital shadows (unidirectional real-time data updates), to true digital twins with bidirectional data flows and closed-loop control [11,12]. True DTs can simulate and predict future states, test interventions virtually, and implement optimized actions in the physical system. Their ability to simulate rare or unsafe scenarios makes them particularly valuable in domains where real-world experimentation is costly or infeasible [13,14]. DT technology enables the de-

sign, operation, and maintenance of physical assets across many domains: manufacturing, healthcare, agriculture, aerospace, construction, transportation, etc. [15–23].

DTs are increasingly used for city management, infrastructure planning, and user acceptance of smart city technologies [1,24–28]. DTs enable continuous monitoring of urban dynamics (e.g., traffic, energy use), adaptive interventions such as traffic re-routing or infrastructure adjustments, and predictive modeling to assess long-term impacts of policy or design decisions.

DTs support precision agriculture by integrating IoT sensor data with system-level models [29] to enable continuous monitoring of crops, livestock, and environmental conditions, adaptive interventions such as irrigation or fertilization adjustments, and predictive modeling of yields, disease risk, and sustainability outcomes. This supports data-driven decision-making under variable environmental conditions.

In healthcare, DTs can represent patients, diseases, or treatments to support personalized medicine [1,9,10,15,17]. Such DTs enable continuous monitoring of disease progression, adaptive treatment optimization based on patient response, and predictive modeling of future disease trajectories and treatment outcomes, improving shared decision-making and care quality.

Automotive DTs are widely used for vehicle design, traffic management, and autonomous driving validation [12,22,30]. DT-based simulations allow continuous monitoring of vehicle and driver states, adaptive interventions such as driver assistance or control strategy adjustments, and predictive modeling under rare or hazardous scenarios [31]. Driving simulator DTs enable prediction of driver distraction and risk by replicating real-world scenarios [32].

2.2. Simulation for Driving Digital Twins

Simulations, which encompass virtual environments (VEs), virtual reality (VR), and augmented reality (AR), have become indispensable tools for replicating real-world scenarios within controlled virtual settings. Although the technical definitions of these terms diverge slightly, they are unified by a common objective: the virtual recreation of real-world situations to investigate the extent to which human behavior in virtual and physical environments aligns or diverges. As a result of developments in computing power and rendering technologies, simulations have become an increasingly prominent feature of a diverse range of disciplines, including psychology, training and design [33]. These tools permit the isolated and controlled study of the impacts of spatial environments on human perception and behavior, in accordance with the principle of behavioral realism, which seeks to evoke responses in virtual settings that mirror those in real-world environments [34,35].

The value of simulations lies not only in their capacity to approximate the intricacies of the real world, but also in their utility in addressing practical concerns. They have been employed for the evaluation of design alternatives, the enhancement of training environments in areas such as medicine and the military, and the assurance of safety in hazardous scenarios [33,36]. A significant benefit of simulations is their capacity to integrate the ecological validity of field studies with the experimental control of laboratory studies, providing a reproducible framework for experimental scenarios that is often unfeasible in the physical world [37,38]. Nevertheless, ensuring the transferability of insights from virtual to real-world contexts remains a challenge. For example, while physiological responses in simulations frequently mirror those observed in real-world scenarios, psychological reactions can display significant discrepancies. This underscores the necessity of continuous validation and optimization of simulation models and methods [39,40].

Across domains, immersive simulations have demonstrated strong benefits for training, design, and behavioral research. In sports, adaptive virtual environments improve

real-world skill acquisition and long-term performance by tailoring practice to individual ability [41]. In architecture and environmental design, immersive simulations support controlled evaluation of cognitive and physiological responses to built environments, enabling evidence-based design decisions [42]. In medical, technical, and ergonomic training, VR and AR provide safe, repeatable, and high-fidelity practice of complex skills that outperform traditional methods [43,44]. Finally, in safety-critical domains, simulations offer realistic yet risk-free environments for studying behavior and training first responders, enabling experimentation that would be impractical or unsafe in real-world settings [45,46].

Driving simulators are a key enabling technology for driving digital twins because they allow controlled, repeatable, and safe replication of driving scenarios while capturing human responses under standardized conditions [6,47]. For a driving simulator to function as a meaningful component of a DT, its validity must be assessed against real driving in terms of subjective experience, behavior, and, crucially for this paper, physiological responses.

Driving simulator validity is typically examined through (i) subjective measures (e.g., perceived realism or stress), (ii) physiological measures (e.g., heart rate and electrodermal activity, while accounting for confounds such as simulator sickness), and (iii) objective behavioral measures (e.g., speed regulation and lane keeping) [48–51]. Validity is often discussed as absolute (numerical equivalence) versus relative (similar trends), with the latter frequently sufficient for research and design inference [52,53].

Driving simulators can reliably reproduce diverse road types and scenarios, enabling the study of driving dynamics such as speed, lane keeping, and errors [54,55]. Physiological and behavioral validation studies show strong similarities between real and simulated driving, including comparable heart rate, brain activity, and gaze patterns, supporting simulator presence, cognitive engagement, and both relative and absolute validity [7,56,57]. Nonetheless, simulator findings must be complemented with on-road validation to ensure generalizability, robustness, and safety in real-world applications [47,58,59].

Recent simulator validity reviews emphasize inferential statistics and correlation-based approaches but do not include time-series trajectory comparisons as a primary validation strategy [51]. This matters for DT contexts because DT functionality is inherently temporal: the intended use is typically not a single averaged state but an evolving state over time. Accordingly, this study evaluates temporal equivalence by comparing time-resolved physiological trajectories between real driving and the driving DT, complementing segment-wise aggregation with trajectory-level similarity assessment.

3. Materials and Methods

3.1. Context

To understand if a driving simulation DT can replicate a physical experience, we analyze data from an experimental study comparing the responses of participants in a simulated versus a real driving environment. The study focuses on measuring and comparing physiological stress measurements between the two experimental environments. The physiological stress response is believed to be related to performance [60], and therefore similar stress responses in the two driving environments will signal similar performance levels.

An initial simulation validation analysis from the experimental study was reported in [8]. In that work, the authors briefly describe the experiment and then use Bayesian comparisons of segment-wise means of physiological and cognitive measures to compare the simulation and real driving environments in aggregate. This approach provides evidence for absolute and relative validity at the level of aggregated segments, but it does not analyze the continuous temporal evolution of the physiological signals. Unlike the segment-based analysis in [8], this study focuses on continuous physiological trajectories, enabling time-resolved comparison of real and simulated driving dynamics. This allows

evaluation of temporal equivalence rather than aggregated mean differences, which is essential for validating driving digital twins in real-time interaction settings.

3.2. Driving Simulation DT

The simulated driving DT was designed using a physical driving simulator (see Figure 1) that closely resembles a real vehicle and a virtual route that replicates a real-world driving experience, ensuring the simulator's validity [61].



Figure 1. The Driving Simulator and Golf 8 Cockpit View from Participant's Perspective.

First, the driving simulator was built to closely replicate the vehicle used in the real-world test. The simulator uses components from a Volkswagen Golf 7, while the real vehicle is a Volkswagen Golf 8, which has similar dimensions, steering and pedal functionality as the previous model. The driving simulator is a mid-fidelity simulator [50] based on a D-Box motion system with three degrees of freedom and equipped with an original vehicle (Sparco) driver's seat. The simulator has a Golf 7 force-feedback steering wheel and pedal set, providing realistic steering and ground feedback. The visual system consists of three 55-inch LCD screens, creating a 180° horizontal field of view to enhance participant immersion. The speedometer is digitally displayed on a screen. A high level of similarity between the simulator and real vehicle exists in terms of ergonomics and control elements (steering wheel, turn signal levers, other buttons, and pedal system) because the two Volkswagen models are based on similar platforms. The steering wheels of both models share a similar shape and tactile feel, including multifunction buttons, ensuring a consistent user interface. The pedal system is also comparable in terms of arrangement and resistance, allowing for an authentic transfer of driving inputs between the simulator and the real vehicle. Despite technological advancements in the Golf 8, such as a more digitized instrument cluster, the fundamental control concept remains similar.

Second, the simulated driving route was designed to closely replicate the real-world driving route (23 km, including urban, rural, and highway driving). Both the starting and ending points of the simulated route were identical to those of the real-world drive, ensuring high fidelity in route replication. This minimizes potential discrepancies in user experience due to hardware differences, ensuring the validity of the physiological and behavioral data collected.

The driving route was designed using the Silab (V7.2) software, developed by the Würzburg Institute for Traffic Sciences. This software allows the integration of OpenStreetMap data as a foundation, incorporating the topographical features of the real-world environment. Since OpenStreetMap provides not only road layouts but also precise street widths and lengths, the route could be reconstructed with a high degree of accuracy. The surrounding environment was designed by selecting pre-modeled buildings, with particular attention given to choosing structures that most closely resembled their real-world counterparts. Elements such as trees and other landscape features were added accordingly to create a realistic overall representation. For prominent landmarks along the route, especially the starting point, characteristic buildings were meticulously recreated using Blender and integrated into the simulation. This ensured that key orientation points from the real-world environment were recognizable in the virtual setting. Traffic signs, road markings, lights, and special road surfaces were replicated at identical locations in the simulator, ensuring high consistency and realistic visual perception relative to the real-world route. Figure 2 shows several sections of the real and the simulated routes.



Figure 2. Comparison of Driving Route Sections (Real and Simulated).

3.3. Experimental Procedure

The experimental procedure was divided into several consecutive phases to ensure a systematic implementation. Participants were welcomed and introduced to the study protocol before completing a pre-study survey to collect demographic information and initial subjective assessments. After the pre-study survey, physiological sensors were attached and the participants proceeded to the real driving phase. During this phase, continuous physiological measurements were recorded. Subsequently, participants returned to the lab to transition into the simulation phase. To minimize the likelihood of simulator sickness and allow for adaptation, participants completed three short warm-up drives on standardized routes [62]. After this acclimatization period, participants drove a simulated route on the DT of the previously driven real-world route, with identical physiological parameters recorded. For practical reasons, the drives were conducted in a fixed order: the real drive always preceded the simulator drive. This simplified the experimental procedure and prevented potential safety risks in real traffic due to possible simulator sickness. The experiment concluded with a final debrief and participant farewell. Thus, both the real

and simulated driving conditions were tested under comparable circumstances, enabling robust analyses and meaningful comparisons.

3.4. Measurements

To understand the effect of the real and simulated ride experience on study participants, physiological measurements of their stress response were collected using electrocardiogram (ECG) and galvanic skin response (GSR) devices. ECG and GSR can measure several biological markers that are widely recognized as reliable indicators of autonomic nervous system (ANS) activation and regulation—which determines the physiological stress responses [60]. It is important to note that these physiological markers are mainly involuntary, and not easily manipulated or hidden, like psychological or behavioral metrics can be. Therefore, physiological markers can offer robust insights into the amount of stress experienced during various activities, and the associated expected performance level (which tends to follow an inverted U curve) [60]. Inverted U-shaped relationships have been found for similar variables as well, for example, between task demand (workload) and performance [63], in agreement with stress–performance results. Both ECG and GSR indicators can be used to predict if a driver has situational awareness or if he is distracted or fatigued [64,65]. Figure 3 depicts the placement of the ECG and GSR sensors.

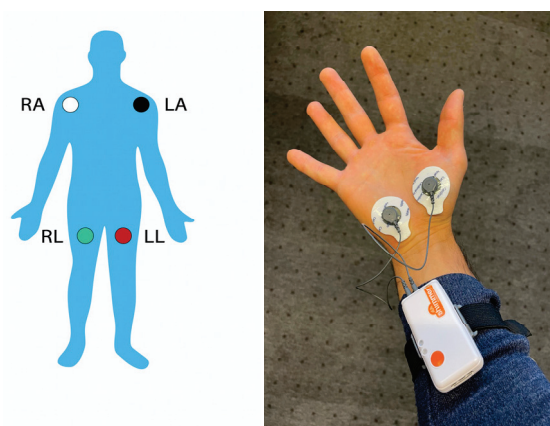


Figure 3. Placement of ECG and GSR sensors.

ECG markers included the heart rate (HR) and heart rate variability (HRV) measures—RR interval (RR-Int), root mean square of successive differences (RMSSD), and standard deviation of NN intervals (SDNN). HR, measured in beats per minute (bpm), exhibits a significant increase during stress conditions, reflecting the activation of the sympathetic nervous system (SNS). This increase indicates the body's heightened metabolic demand as it prepares for a “fight-or-flight” response to perceived threats. Elevated heart rate under stress is a well-documented phenomenon [66,67]. The RR interval, defined as the time between consecutive heartbeats, demonstrates an inverse relationship with heart rate, decreasing during stress. The Root Mean Square of Successive Differences (RMSSD) serves as a measure of short-term heart rate variability and typically decreases under stress. Similarly, the Standard Deviation of NN Intervals (SDNN) also decreases [68,69]. ECG measurements were taken using the Shimmer3 EXG sensor (Shimmer, Dublin, Ireland), with electrodes placed in a four-point configuration. These were positioned below the left and right clavicles, above the right iliac crest, and on the left iliac crest (serving as the reference electrode). The Shimmer3 EXG sensor has been validated in several studies and proven reliable for physiological measurements [70–72].

GSR markers included the skin conductance response (SCR), peak amplitude (PA), and skin conductance level (SCL), which are electrodermal activity measures that increase

significantly during stress due to the activation of sweat glands [60]. The GSR data were recorded using the Shimmer3 GSR+ sensor, which operates exosomatically with direct current [73]. Electrodes were placed on the palms of the participants' non-dominant hands to ensure they could comfortably grip the steering wheel in both the real vehicle and the simulator. The Shimmer3 GSR+ sensor is frequently used in research and has provided valid results in similar contexts [74–76].

One of the challenges in data collection was synchronizing the data streams from the measurement devices to a common starting and ending point. To achieve this, iMotions Software 9.4 was utilized. The raw data collected during the experiment were processed using the integrated R algorithms of the iMotions software, which enabled both noise filtering and data segmentation. This processing included the calculation of heart rate (HR), heart rate variability metrics (HRV) such as the RR interval, RMSSD, and SDNN, as well as the segmentation of tonic and phasic GSR values (SCR, PA, SCL). These calculations and data preparations served as the foundation for further analysis of the participants' physiological responses in both the real and simulated driving environments.

Figure 4 summarizes the data collection and preparation procedures, as well as the data analysis steps.

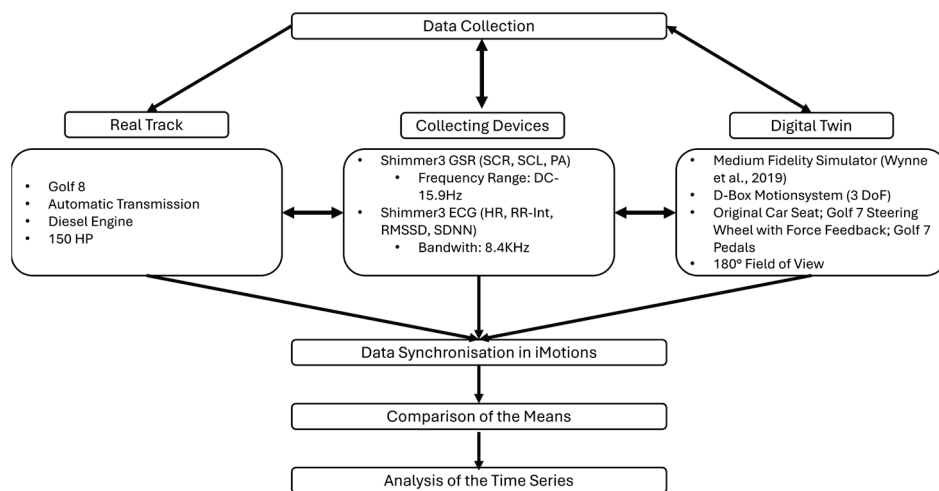


Figure 4. Data Collection and Analysis [50].

3.5. Participants and Sampling

The data was collected during the third quarter of 2023 in a small sized town in Germany. A convenience sampling method was used, with quotas set for age and gender to ensure a balanced sample. A total of 72 participants initially took part in the experiment; however, due to equipment failures and incomplete data for several experimental observations, only data from 68 participants is included in the final analysis. Participants were incentivized with €30 for their participation in the study. The final sample consists of 37 women (54.4%) and 31 men (45.6%), aged between 18 and 63 years ($M = 30.07$, $SD = 11.58$). Regarding educational background, most participants indicated they have either completed or are pursuing higher education (85.3%), while 8.8% have a high school diploma, and 5.9% have completed secondary school qualifications. In terms of regional origin, 41.2% identify as living in rural areas, 13.2% describe themselves as urban dwellers, and 45.6% categorize themselves as residents of small to medium-sized towns. Furthermore, none of the participants had any prior experience with a driving simulator.

The Ethics Committee of the University of Bayreuth approved the study. All participants took part voluntarily, and each participant was informed that they could withdraw from the study at any time without providing a reason.

3.6. Route Description

The driving route (both real and simulated) was approximately 23 km long. The route was divided into seven segments to encompass a wide range of driving scenarios and environmental conditions, intended to impose different cognitive demands on the driver and influence the driver's physiological response. These segments included urban driving (5.3 km), rural roads (9.7 km), and highway driving (8 km). The division of the route into seven segments was based on the actual course of the selected circuit. The route was designed as a loop starting and ending at the university, ensuring coverage of all three driving environments (urban, rural, and highway). The highway section formed one continuous segment, while the urban and rural sections were naturally interspersed due to the loop's geographical layout (e.g., rural stretches interrupted by short urban passages). This reflected the practical nature of the chosen route and allowed for a realistic representation of transitions between different driving contexts.

To mitigate the risk that physiological stress responses in simulated driving are weakly differentiated under low task demands [77], the study combined three different driving environments (urban, rural and highway), thereby increasing situational variability and supporting analysis across differing cognitive and environmental demands. This segmentation ensured that the driving simulator could realistically replicate diverse real-world scenarios, validating its versatility and applicability as a research tool. It should be noted that the segments were presented in a mixed order. Each driving environment was characterized by distinct attributes:

- **Urban Driving:** This environment had high stimulus density from crossings, intersections, signals, and congestion, requiring frequent speed changes and rapid decisions, and increasing mental workload and stress.
- **Rural Roads:** This environment involved longer stretches at steady speeds with occasional curves and overtaking and navigating narrow roads, producing a moderate cognitive and physiological load.
- **Highway Driving:** This environment involved high-speed driving under consistent conditions, with challenges such as merging, lane changes, and vehicle interactions requiring sustained attention and anticipation, despite some overall monotony.

4. Results

This section presents the empirical results of the study. First, descriptive statistics of the physiological stress indicators are reported for both the real and simulated drives across all route segments. Subsequently, the time series analysis is introduced to examine the similarity between simulator and real-world conditions in greater temporal detail. Finally, the main findings are summarized and discussed with respect to physiological stress patterns observed in both environments.

4.1. Descriptive Statistics

Tables 1–7 report the minimum, maximum, mean, and standard deviation for each variable (SCR, SCL, PA, HR, RR-interval, RMSSD, and SDNN), both for the real drive (Real Ride) and the simulated drive (Sim. Ride), and separately for each one of the seven driving route segments. Driving route segments 1, 3, and 5 are classified as rural traffic, segments 2, 6, and 7 as urban traffic, and segment 4 as highway traffic. Paired *t*-tests ($N = 68$; SPSS v27; two-tailed; $\alpha = 0.05$; pairwise deletion) are conducted to examine whether there are any significant differences between the two driving environments, and the corresponding *p*-values are also included in Tables 1–7.

Table 1. Descriptives for Skin Conductance Response in real and simulated ride; 7 segments.

		SCR				
		Min.	Max.	Mean	SD	<i>p</i> (sig.)
Segment 1	Real Ride	6.38	15.35	10.25	2.16	<0.001
	Sim. Ride	5.18	15.54	10.10	2.06	
Segment 2	Real Ride	5.27	17.80	9.80	2.91	0.008
	Sim. Ride	4.78	16.72	9.94	2.65	
Segment 3	Real Ride	2.89	15.64	9.73	2.73	<0.001
	Sim. Ride	4.31	16.00	9.68	2.47	
Segment 4	Real Ride	4.06	14.34	9.26	2.21	<0.001
	Sim. Ride	5.02	14.63	9.55	2.31	
Segment 5	Real Ride	3.32	14.99	9.15	2.22	<0.001
	Sim. Ride	4.01	14.12	9.04	2.48	
Segment 6	Real Ride	0.00	15.45	8.79	3.12	<0.001
	Sim. Ride	1.61	14.36	8.43	2.77	
Segment 7	Real Ride	1.99	14.06	9.49	2.31	<0.001
	Sim. Ride	5.32	14.25	9.97	2.30	

Table 2. Descriptives for Skin Conductance Level in real and simulated ride; 7 segments.

		SCL				
		Min.	Max.	Mean	SD	<i>p</i> (sig.)
Segment 1	Real Ride	3.10	46.50	13.07	5.73	<0.001
	Sim. Ride	5.92	39.76	14.46	5.17	
Segment 2	Real Ride	3.20	45.22	12.47	5.56	<0.001
	Sim. Ride	5.00	44.40	13.70	5.64	
Segment 3	Real Ride	3.19	52.33	13.05	6.38	<0.001
	Sim. Ride	4.33	52.20	14.17	6.42	
Segment 4	Real Ride	3.21	55.85	13.09	6.49	<0.001
	Sim. Ride	3.58	68.67	14.57	8.04	
Segment 5	Real Ride	2.88	57.18	12.72	6.76	<0.001
	Sim. Ride	4.27	88.13	14.38	10.15	
Segment 6	Real Ride	2.70	56.38	12.29	6.85	<0.001
	Sim. Ride	2.32	93.98	14.17	10.97	
Segment 7	Real Ride	2.32	51.84	13.45	6.50	<0.001
	Sim. Ride	4.26	102.55	15.85	11.67	

Data were screened for outliers; extreme values ($>3 \times \text{IQR}$) were winsorized to the nearest non-extreme bound to limit their influence while retaining observations [78]. The normality assumption was evaluated on the difference scores (Real – Sim) for each segment and measure using Q–Q plots, histograms, and Shapiro–Wilk/Kolmogorov–Smirnov tests. While the normality tests frequently rejected perfect normality—consistent with the behavior of physiological measures that often exhibit mild asymmetry [79]—visual inspection indicated only slight to moderate skewness without substantial anomalies.

Table 3. Descriptives for Peak Amplitude in real and simulated ride; 7 segments.

		PA				<i>p</i> (sig.)
		Min.	Max.	Mean	SD	
Segment 1	Real Ride	0.03	0.75	0.30	0.17	<0.001
	Sim. Ride	0.06	0.84	0.36	0.19	
Segment 2	Real Ride	0.03	0.79	0.27	0.19	<0.001
	Sim. Ride	0.04	0.95	0.34	0.21	
Segment 3	Real Ride	0.03	0.84	0.30	0.19	<0.001
	Sim. Ride	0.04	1.00	0.39	0.23	
Segment 4	Real Ride	0.03	0.79	0.28	0.17	<0.001
	Sim. Ride	0.04	0.93	0.39	0.23	
Segment 5	Real Ride	0.02	0.84	0.26	0.17	<0.001
	Sim. Ride	0.05	0.99	0.48	0.24	
Segment 6	Real Ride	0.00	0.84	0.24	0.19	0.007
	Sim. Ride	0.03	1.05	0.36	0.29	
Segment 7	Real Ride	0.02	0.81	0.35	0.21	<0.001
	Sim. Ride	0.05	0.92	0.38	0.23	

Table 4. Descriptives for Heart Rate in real and simulated ride; 7 segments.

		HR				<i>p</i> (sig.)
		Min.	Max.	Mean	SD	
Segment 1	Real Ride	57.56	112.86	83.26	13.55	<0.001
	Sim. Ride	54.47	105.00	75.99	11.28	
Segment 2	Real Ride	55.92	114.24	85.44	13.61	<0.001
	Sim. Ride	48.31	105.00	77.03	12.13	
Segment 3	Real Ride	51.52	110.63	80.69	12.65	<0.001
	Sim. Ride	51.86	105.00	76.70	11.29	
Segment 4	Real Ride	56.58	112.48	83.17	12.87	<0.001
	Sim. Ride	57.08	106.50	77.35	10.68	
Segment 5	Real Ride	53.81	110.98	81.53	11.91	<0.001
	Sim. Ride	54.50	105.03	76.83	10.69	
Segment 6	Real Ride	60.43	113.37	83.83	13.70	<0.001
	Sim. Ride	53.92	104.00	77.27	11.23	
Segment 7	Real Ride	58.94	113.71	81.61	11.76	<0.001
	Sim. Ride	55.13	106.21	77.97	11.01	

Given the sample size ($N = 68$), the paired-sample t -test is robust to moderate deviations from normality of the difference scores [80], and by the Central Limit Theorem, the sampling distribution of mean differences is approximately normal for $N \geq 30$, particularly when underlying departures are mild [81]. No correction for multiple comparisons was applied because the seven route segments represent a priori distinct driving contexts rather than repeated tests of the same effect, and the six physiological measures capture different constructs. Note that, in a within-subject design, the paired analysis reduces error variance by accounting for inter-individual variability. Consequently, the standard deviation of the difference scores is typically smaller than the raw standard deviations of each condition, which can yield relatively large t -values even when mean differences appear numerically small.

Table 5. Descriptives for RR-Interval in real and simulated ride; 7 segments.

		RR-Int				<i>p</i> (sig.)
		Min.	Max.	Mean	SD	
Segment 1	Real Ride	530.12	1011.90	734.98	125.15	<0.001
	Sim. Ride	546.58	1115.29	831.99	143.39	
Segment 2	Real Ride	533.26	1208.01	766.17	149.83	<0.001
	Sim. Ride	534.41	1231.49	810.22	140.54	
Segment 3	Real Ride	531.86	1155.43	748.03	141.88	<0.001
	Sim. Ride	550.14	1226.95	819.02	138.34	
Segment 4	Real Ride	537.11	1100.00	754.84	131.09	<0.001
	Sim. Ride	565.66	1135.38	811.49	126.15	
Segment 5	Real Ride	531.68	1100.00	765.74	119.55	<0.001
	Sim. Ride	572.79	1157.47	814.64	126.37	
Segment 6	Real Ride	546.15	1100.00	775.33	120.06	<0.001
	Sim. Ride	554.22	1181.40	805.72	132.77	
Segment 7	Real Ride	546.34	1051.04	764.43	113.81	<0.001
	Sim. Ride	567.92	1093.19	806.02	127.44	

Table 6. Descriptives for RMSSD in real and simulated ride; 7 segments.

		RMSSD				<i>p</i> (sig.)
		Min.	Max.	Mean	SD	
Segment 1	Real Ride	12.44	425.90	126.04	116.88	<0.001
	Sim. Ride	10.78	672.23	145.24	157.19	
Segment 2	Real Ride	5.85	603.15	113.46	123.32	0.005
	Sim. Ride	8.34	766.49	135.42	170.59	
Segment 3	Real Ride	7.70	562.59	106.95	125.29	<0.001
	Sim. Ride	18.43	638.60	122.87	136.31	
Segment 4	Real Ride	7.67	875.76	104.53	142.85	<0.001
	Sim. Ride	15.03	519.35	119.51	132.21	
Segment 5	Real Ride	6.37	581.00	98.77	124.18	0.003
	Sim. Ride	16.13	464.68	101.52	120.55	
Segment 6	Real Ride	6.17	559.57	83.85	111.17	0.002
	Sim. Ride	5.46	585.25	82.35	115.12	
Segment 7	Real Ride	7.91	444.34	102.86	97.31	0.004
	Sim. Ride	17.47	458.33	111.65	117.92	

Table 7. Descriptives for SDNN in real and simulated ride; 7 segments.

		SDNN				<i>p</i> (sig.)
		Min.	Max.	Mean	SD	
Segment 1	Real Ride	14.97	424.78	127.01	120.47	<0.001
	Sim. Ride	17.88	532.63	143.35	152.58	
Segment 2	Real Ride	10.70	702.26	118.69	135.49	0.024
	Sim. Ride	16.24	621.76	127.57	149.21	
Segment 3	Real Ride	14.52	558.48	115.38	135.92	0.001
	Sim. Ride	24.03	582.51	120.67	128.48	

Table 7. Cont.

		SDNN				<i>p</i> (sig.)
		Min.	Max.	Mean	SD	
Segment 4	Real Ride	14.64	878.39	112.01	151.76	<0.001
	Sim. Ride	18.98	551.75	118.39	131.65	
Segment 5	Real Ride	13.91	617.05	103.27	132.79	0.010
	Sim. Ride	21.88	480.04	104.18	121.64	
Segment 6	Real Ride	11.78	450.57	80.17	100.16	0.008
	Sim. Ride	6.57	271.74	70.80	54.38	
Segment 7	Real Ride	17.06	485.72	105.70	100.30	0.008
	Sim. Ride	25.67	449.61	114.18	119.15	

4.2. Time Series Analysis

By analyzing time-resolved physiological trajectories, we can assess whether the digital twin tracks the real-time evolution of stress responses in a way that is compatible with its intended use: continuous monitoring, adaptive interventions, and predictive modelling. To achieve this, the dataset was restructured into a long-format time series representation to enable trajectory-based analysis across the full driving duration. Through this process, columns are transformed into rows, converting a dataset from a wide format to a long format [82]. In our case, each segment of the route (column) becomes a new observation (row). This allows us to calculate averages for the entire duration of the drive rather than focusing solely on individual segments. To compare the physiological data between the entire real driving route and its DT, across participants, we first calculated the mean values of the physiological parameters and then performed a paired *t*-test.

The analysis of skin conductance showed that for the real drive, the Skin Conductance Response (SCR) (Table 1) had a mean value of 9.49 (range: 0.00–17.80; SD = 2.57), which was almost identical to the mean value of 9.53 (range: 1.61–16.72; SD = 2.49) for the simulated drive ($p = 0.829$). However, for the tonic Skin Conductance Level (SCL) (Table 2), a highly significant difference was observed: In the real drive, the mean was 12.87 (range: 2.32–57.18; SD = 6.31), while in the simulated drive, the mean was significantly higher at 14.47 (range: 2.32–102.55; SD = 8.62; $p < 0.001$). The Peak Amplitude (PA) (Table 3) also showed significant differences, with a mean of 0.29 (range: 0.00–0.84; SD = 0.18) in the real drive compared to 0.39 (range: 0.03–1.05; SD = 0.24) in the simulation ($p < 0.001$).

Significant differences were also found in heart rate (HR) (Table 4) between the real and simulated drives. The mean HR during the real drive was 82.79 bpm (range: 51.52–114.24; SD = 12.89), which was higher than in the simulated drive (mean = 77.02 bpm; range = 48.31–106.50; SD = 11.14; $p < 0.001$). The average RR interval (Table 5) was 758.50 ms (range: 530.12–1208.01; SD = 129.13) in the real drive, significantly shorter than 814.15 ms (range: 534.41–1231.49; SD = 133.13) in the simulated drive ($p < 0.001$). Regarding heart rate variability, represented by RMSSD (Table 6) and SDNN (Table 7), no significant differences were observed between the two conditions (RMSSD: $p = 0.165$; SDNN: $p = 0.524$).

Our analysis uses the continuous time series of physiological recordings. After normalizing timestamps to a common [0,1] scale, we compute Pearson correlations of the full physiological trajectories and visually inspect their temporal evolution.

The analysis focuses on normalized physiological time series, enabling direct comparison of temporal signal dynamics between real and simulated driving across the full interaction period.

Although averaging gives an insight into the similarity of the driving behavior of drivers in the simulator and on the real road, this is only a rough indication. Peaks and extreme values, which indicate intense or stress-related responses, are smoothed out by averaging, resulting in a loss of valuable information [83]. Mean values obscure important characteristics of time series data by failing to capture variability and response stability, which are critical for interpretation [84]. Context-specific effects, such as those arising during cornering or emergency braking, are also lost through averaging, masking the true nature of responses [85]. Moreover, averaging hinders analysis of adaptation and learning over time and prevents examination of dynamic interactions between variables [53]. For these reasons, we want to go one step further and investigate the time series similarity in more detail. To achieve this, we use the Pearson Correlation Coefficient on normalized time series. This linear correlation between two variables X and Y ranges from -1 and 1 , where 1 is a perfect positive linear correlation, -1 is a perfect negative linear correlation, and 0 is no linear correlation. We also normalize the time series to a common scale, enabling a meaningful comparison independent of absolute values or differing measurement durations.

Conducting this time-series analysis required addressing several technical challenges that go beyond the procedures described in [8] and constitutes an analytical innovation for physiological DT validation:

- (1) Synchronizing real and simulated data streams with differing durations;
- (2) Normalizing the time axis to a dimensionless scale to allow comparison of drives with different lengths;
- (3) Restructuring the dataset via unpivoting so that each route segment becomes a separate observation in order to increase the effective sample size per segment and enable calculation of mean values over the entire drive instead of only segment-level averages;
- (4) Selecting similarity metrics that are both robust to noise and interpretable for practitioners (i.e., combining Pearson Correlation Coefficients of normalized trajectories with visual inspection of the curves, which jointly reveal pattern similarity and deviations that are completely invisible in segment-wise means).

To ensure a meaningful comparison between the time series from the simulator and real-world driving, we normalized the time axis for both datasets to a common scale of $[0,1]$. This normalization adjusts the timestamps of each dataset so that their relative progression is comparable, irrespective of their differing durations or absolute time values. The normalization is performed using the following formula, where Timestamp is the original time value for each data point, $\text{Timestamp}_{\min}$ and $\text{Timestamp}_{\max}$ are the minimum and maximum time values within the dataset:

$$\text{Normalized Time} = \frac{\text{Timestamp} - \text{Timestamp}_{\min}}{\text{Timestamp}_{\max} - \text{Timestamp}_{\min}}$$

This formula ensures that the earliest timestamp maps to 0 and the latest timestamp maps to 1 , creating a dimensionless time scale that is independent of the original measurement duration. By normalizing both the simulator and real-world datasets, we aligned their time axes, enabling direct visual comparison of the signal dynamics.

TimeStamp denotes the original time value of a measurement, while $\text{TimeStamp}_{\min}$ and $\text{TimeStamp}_{\max}$ represent the first and last recorded time values of the respective dataset. Through this normalization, the starting point of every drive is mapped to 0 and the end to 1 , creating a dimensionless and aligned time scale. This ensures that the temporal progression of the biosignals can be directly compared between real-world and simulated conditions, independent of absolute drive duration. In simple terms, this normalization ensures that all drives “start at 0 and end at 1 ”, no matter how long they actually lasted in minutes or seconds. Without this step, two time series with slightly different durations

could not be compared fairly, because one would have more data points or a longer timeline than the other. By converting all timestamps into the same normalized scale, we align both curves so that the same relative moments of the drive can be compared—for example, the beginning (0.0), the middle (0.5), or the final part of the drive (1.0). We then use this NormalizedTime as the common time axis to visually overlay both time series and to calculate correlation values, ensuring a fair, consistent, and meaningful comparison of signal patterns between the real and simulated driving conditions. This step is critical for comparing datasets with different lengths or temporal resolutions, as it eliminates the impact of absolute time differences. The normalized time series were then plotted to visually assess their similarity.

Figure 5 illustrates a biosignal time series for a participant driving the urban and rural segments, comparing data collected in the simulator and real-world conditions. The Pearson Correlation Coefficients for the urban drive ($r = 0.31$) and the rural drive ($r = 0.34$) suggest a moderate level of linear similarity between the simulator and real-world measurements. The normalization revealed strong parallels in the overall patterns of the tonic signal across both environments. This visual alignment, combined with the Pearson Correlation Coefficients, provides robust evidence for the comparability of the simulator and real-world measurements, supporting the validity of using simulators in such studies.

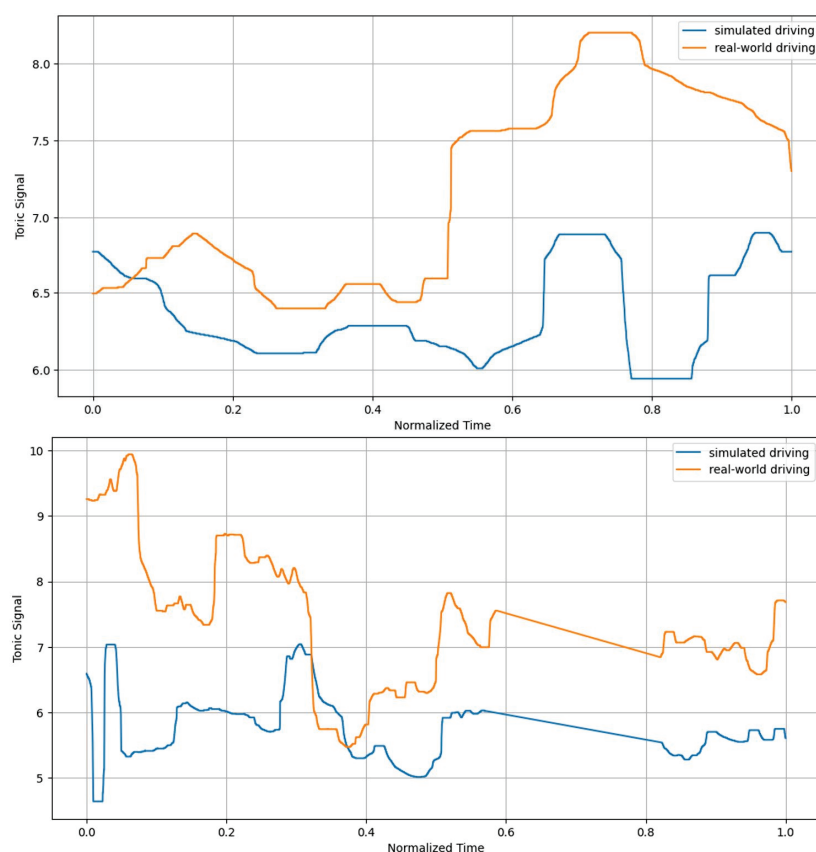


Figure 5. Comparison of Two Time Series (for a participant driving the urban and rural segments, real and simulated; see code on GitHub [86]).

4.3. Discussion of Results

Analysis of the biosignal data reveals several notable findings regarding the physiological stress responses of the participants. In terms of cardiovascular responses, the real ride generally produces higher heart rate values, indicating greater physical exertion and more intense emotional experiences. In contrast, the simulated ride produces a less pronounced

heart rate response, suggesting a less physically demanding experience. The longer RR intervals and higher RMSSD values during the simulated ride indicate a more relaxed heart rate variability, reflecting a less stressful physiological response. The real ride, with shorter RR intervals and higher SDNN values, results in greater cardiovascular stress and more intense responses. In addition, in terms of skin responses, the simulated ride tended to elicit greater sympathetic activation, as reflected by increased skin conductance response (SCR) and greater variability, particularly in the later stages. This suggests that the simulated experience may be perceived as more emotionally arousing or stressful, despite potentially being less physically challenging than the real ride. This trend is further supported by higher peak values and greater variability in skin conductance during the simulated rides.

The Skin Conductance Response (SCR) generally indicates higher sympathetic activation during the simulated ride in most segments, as reflected by somewhat higher mean values and greater fluctuations. The maximum values are also higher during the simulated ride, suggesting stronger autonomic responses.

Overall, the simulated ride leads to a stronger and more variable activation of the sympathetic nervous system compared to the real ride, which may indicate higher emotional stress or arousal during the simulation. In all seven segments, there is a general trend that the simulated rides exhibit higher Skin Conductance Level (SCL) than the real rides. This could suggest that the simulated experiences were either perceived as more intense or that the reactions were more strongly stimulated, even though they may offer fewer physical challenges compared to the real ride. The standard deviations are also generally higher in the simulated ride, which points to greater variability in skin conductance values, suggesting a more diverse range of reactions among participants.

The simulated rides exhibit higher peak amplitudes (PAs) in all segments compared to the real rides, suggesting that the simulated experiences generally elicit stronger peak responses. Particularly in later segments, the simulated rides show intense reactions and greater variability, possibly reflecting a broader emotional range.

The heart rate (HR) during the real ride is higher in all segments compared to the simulated ride, suggesting greater physical load. The higher maximum heart rate values during the real rides may reflect stronger peak reactions due to physical exertion or intense emotional experiences. In contrast, the simulated ride shows lower mean and maximum values, suggesting that it was less physically demanding or less exciting.

The RR intervals (RR-Int) are longer in all segments during the simulated versus the real ride, suggesting less physiological strain and more relaxed heart rate variability.

The mean RMSSD is higher during the simulated ride, indicating stronger parasympathetic activity and greater relaxation. Similarly, SDNN values tend to be higher, reflecting greater heart rate variability during simulation.

Overall, the analysis shows that the real ride generally produces higher heart rate values, indicating greater physical exertion, while the simulated ride elicits greater sympathetic activation in skin responses, pointing to higher emotional arousal.

Together, these findings suggest that simulated driving induces greater variability in physiological responses—especially sympathetic activation and heart rate variability—whereas real driving produces stronger, more consistent physical stress responses. This highlights the different physiological characteristics of real versus simulated experiences.

In summary, physiological data show complementary stress patterns in real and simulated driving: real driving induces stronger cardiovascular activation reflecting higher physical load, whereas simulation induces greater sympathetic responses reflecting higher emotional arousal. This supports the validity of driving simulators for studying stress while highlighting inherent physiological differences from real-world driving.

5. Summary, Limitations and Outlook

This study examines the similarity of physiological responses in real and simulated driving to assess the validity of simulators as proxies for real-world experience. The results show both meaningful similarities and systematic differences, indicating that while some physiological aspects are well replicated, others diverge due to inherent contextual and perceptual limitations of virtual environments.

The mean values comparison across route segments identifies similarities and differences between real and simulated driving. Higher skin conductance measures in the simulator suggest increased arousal potentially due to unfamiliarity or simulator configuration. Real-world driving elicited higher heart rates and shorter RR intervals, indicating greater physical and emotional strain. Notably, no significant differences were found in RMSSD and SDNN, suggesting comparable autonomic responses across environments.

The time series analysis shows higher correlation for rural driving, suggesting simulators better replicate less complex environments, while lower urban correlations reflect greater behavioral variability not fully captured by simulation. Despite only moderate correlations, the overall temporal trends are similar, indicating that key aspects of driver behavior are reproduced. These results highlight the value of time series analysis for studying comparability and support the use of simulators in time-dependent driving research.

Taken together, the mean and time-series analyses suggest that the driving simulation DT can serve as a useful tool for studying general physiological response patterns under controlled conditions. However, it does not fully replace real-world driving tests, particularly in complex or high-variability environments.

We believe the combination of DT framing and physiological time series analysis we present in this paper is novel and fills a gap at the intersection of DT research and real-time user interaction. We acknowledge there is a small but growing body of research that employs time-series techniques in driving simulation studies (i.e., dynamic time warping and similarity metrics linking ECG/EMG signals to risk patterns in intersection scenarios, as well as Pearson correlations of driving behavior time series) to quantify consistency between real and simulated driving. However, to the best of our knowledge, no prior study (1) conceptualizes a driving simulator explicitly as a digital twin, (2) evaluates physiological time series trajectories as the primary comparison object, and (3) links trajectory similarity to the conditions under which a digital twin may replace real-world testing.

These findings indicate a need to improve DT precision in complex, particularly urban, driving contexts. Future work should better differentiate driving scenarios, examine links between physiological stress and driving performance, and address limitations in ecological validity such as reduced motion cues, environmental variability, and perceived risk. Future studies should also mitigate potential order effects through randomized designs, and employ adaptive simulation algorithms that enable real-time bidirectional data integration, better capture real-world variability, and improve alignment with actual driving behavior. Furthermore, researchers could explore DT interface designs that enhance realism and physiological engagement to improve user interaction and performance.

The insights from this study contribute valuable knowledge to both theory and practical applications of digital twins. From a theoretical perspective, our findings enhance the understanding of user interaction with DTs, especially regarding how physiological equivalence between simulation and reality can serve as an indicator of accuracy. This offers a basis for future research exploring the psychological and perceptual alignment between DTs and their physical counterparts, further developing criteria for evaluating DT effectiveness. From a practical perspective, our findings make an essential contribution to the evaluation of DTs' ability to replicate real-world environments in general, and to the specific area of vehicle simulation in particular. The observed similarities suggest

that certain physiological and behavioral dynamics of real driving can be approximated within the simulator, though substantial contextual and sensory differences remain. These limitations emphasize that simulation-based DTs should currently be viewed as complementary, rather than equivalent, to real-world testing. This insight is of significant interest for industries where real-world experiences are costly, time-consuming, impractical, dangerous, or unethical. The results suggest that DTs can be a valuable tool for studying general physiological trends under controlled and consistent conditions, but that further optimizations are necessary for increasing accuracy for complex environments.

Author Contributions: Conceptualization, M.C., E.S., A.C., C.C., J.R. and S.W.; methodology, M.C., E.S., C.C., J.R. and S.W.; software, M.C. and E.S.; validation: M.C.; formal analysis, M.C., E.S., C.C., J.R. and S.W.; investigation, M.C.; resources, M.C., J.R. and S.W.; data curation, M.C.; writing—original draft preparation, M.C., E.S., A.C., C.C., J.R. and S.W.; writing—review and editing, M.C., E.S., A.C., C.C., J.R. and S.W.; visualization, E.S.; supervision, C.C., E.S., J.R. and S.W.; project administration, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of University of Bayreuth (application number: 25-008 and date 30 January 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study is available on request from the corresponding author.

Conflicts of Interest: Author Eldar Sultanow was employed by the company Capgemini Deutschland GmbH. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Semeraro, C.; Lezoche, M.; Panetto, H.; Dassisti, M. Digital twin paradigm: A systematic literature review. *Comput. Ind.* **2021**, *130*, 103469. [CrossRef]
2. Attaran, M.; Attaran, S.; Celik, B.G. The impact of digital twins on the evolution of intelligent manufacturing and industry 4.0. *Adv. Comput. Intell.* **2023**, *3*, 11. [CrossRef]
3. Singh, M.; Srivastava, R.; Fuenmayor, E.; Kuts, V.; Qiao, Y.; Murray, N.; Devine, D. Applications of digital twin across industries: A review. *Appl. Sci.* **2022**, *12*, 5727. [CrossRef]
4. Atalay, M.; Murat, U.; Oksuz, B.; Parlaktuna, A.M.; Pisirir, E.; Testik, M.C. Digital twins in manufacturing: Systematic literature review for physical–digital layer categorization and future research directions. *Int. J. Comput. Integr. Manuf.* **2022**, *35*, 679–705. [CrossRef]
5. Lo, C.K.; Chen, C.H.; Zhong, R.Y. A review of digital twin in product design and development. *Adv. Eng. Inform.* **2021**, *48*, 101297. [CrossRef]
6. Piromalis, D.; Kantaros, A. Digital twins in the automotive industry: The road toward physical-digital convergence. *Appl. Syst. Innov.* **2022**, *5*, 65. [CrossRef]
7. Johnson, M.J.; Chahal, T.; Stinchcombe, A.; Mullen, N.; Weaver, B.; Bédard, M. Physiological responses to simulated and on-road driving. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **2011**, *81*, 203–208. [CrossRef] [PubMed]
8. Czaban, M.; Himmels, C. Investigating simulator validity by using physiological and cognitive stress indicators. *Transp. Res. Part F Traffic Psychol. Behav.* **2025**, *114*, 831–851. [CrossRef]
9. Voigt, I.; Inojosa, H.; Dillenseger, A.; Haase, R.; Akgün, K.; Ziemssen, T. Digital twins for multiple sclerosis. *Front. Immunol.* **2021**, *12*, 669811. [CrossRef]
10. Barricelli, B.R.; Casiraghi, E.; Fogli, D. A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE Access* **2019**, *7*, 167653–167671. [CrossRef]
11. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital twin: Enabling technologies, challenges and open research. *IEEE Access* **2020**, *8*, 108952–108971. [CrossRef]

12. Botín-Sanabria, D.M.; Mihaita, A.-S.; Peimbert-García, R.E.; Ramírez-Moreno, M.A.; Ramírez-Mendoza, R.A.; Lozoya-Santos, J.D.J. Digital twin technology challenges and applications: A comprehensive review. *Remote Sens.* **2022**, *14*, 1335. [CrossRef]
13. Chircu, A.; Czarnecki, C.; Friedmann, D.; Pomaskow, J.; Sultanow, E. Towards a digital twin of society. In Proceedings of the 56th Hawaii International Conference on Systems Sciences, Maui, HI, USA, 3–6 January 2023.
14. Mihai, S.; Yaqoob, M.; Hung, D.V.; Davis, W.; Towakel, P.; Raza, M.; Karamanoglu, M.; Barn, B.; Shetve, D.; Prasad, R.V.; et al. Digital twins: A survey on enabling technologies, challenges, trends and future prospects. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 2255–2291. [CrossRef]
15. Biller-Andorno, N.; Christen, M.; Krauthammer, M.; Witt, C. *Artificial Intelligence in Medicine—Objectives and Recommendations for the Responsible Use of DIGITAL Twins*; Position paper; University of Zurich: Zurich, Switzerland, 2023.
16. Iranshahi, K.; Brun, J.; Arnold, T.; Sergi, T.; Müller, U.C. Digital twins: Recent advances and future directions in engineering fields. *Intell. Syst. Appl.* **2025**, *26*, 200516. [CrossRef]
17. Kanaga Priya, P.; Reethika, A. *A Review of Digital Twin Applications in Various Sectors*; Springer Nature: Cham, Switzerland, 2024; pp. 239–258.
18. Liu, M.; Fang, S.; Dong, H.; Xu, C. Review of digital twin about concepts, technologies, and industrial applications. *J. Manuf. Syst.* **2021**, *58*, 346–361. [CrossRef]
19. Moenck, K.; Rath, J.-E.; Koch, J.; Wendt, A.; Kalscheuer, F.; Schüppstuhl, T.; Schoepflin, D. Digital twins in aircraft production and MRO: Challenges and opportunities. *CEAS Aeronaut. J.* **2020**, *15*, 1051–1067. [CrossRef]
20. Rasheed, A.; San, O.; Kvamsdal, T. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access* **2020**, *8*, 21980–22012. [CrossRef]
21. VanDerHorn, E.; Mahadevan, S. Digital twin: Generalization, characterization and implementation. *Decis. Support Syst.* **2021**, *145*, 113524. [CrossRef]
22. Wang, B.; Zhang, C.; Zhang, M.; Liu, C.; Xie, Z.; Zhang, H. Digital twin analysis for driving risks based on virtual physical simulation technology. *IEEE J. Radio Freq. Identif.* **2022**, *6*, 938–942. [CrossRef]
23. Xie, M.; Pan, W. Opportunities and challenges of digital twin applications in modular integrated construction. In *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC)*; Osumi, H., Furuya, H., Tateyama, K., Eds.; International Association for Automation and Robotics in Construction (IAARC): Kitakyushu, Japan, 2020; pp. 278–284.
24. Deng, T.; Zhang, K.; Shen, Z.-J.M. A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *J. Manag. Sci. Eng.* **2021**, *6*, 125–134. [CrossRef]
25. Jiang, F.; Ma, L.; Broyd, T.; Chen, W.; Luo, H. Digital twin enabled sustainable urban road planning. *Sustain. Cities Soc.* **2022**, *78*, 103645. [CrossRef]
26. Lu, Q.; Parlikad, A.K.; Woodall, P.; Don Ranasinghe, G.; Xie, X.; Liang, Z.; Konstantinou, E.; Heaton, J.; Schooling, J. Developing a digital twin at building and city levels: Case study of West Cambridge campus. *J. Manag. Eng.* **2020**, *36*, 05020004. [CrossRef]
27. Michalik, D.; Kohl, P.; Kummert, A. Smart cities and innovations: Addressing user acceptance with virtual reality and digital twin city. *IET Smart Cities* **2022**, *4*, 292–307. [CrossRef]
28. Schrotter, G.; Hürzeler, C. The digital twin of the city of Zurich for urban planning. *PFG—J. Photogramm. Remote Sens. Geoinf. Sci.* **2020**, *88*, 99–112. [CrossRef]
29. Verdouw, C.; Tekinerdogan, B.; Beulens, A.; Wolfert, S. Digital twins in smart farming. *Agric. Syst.* **2021**, *189*, 103046. [CrossRef]
30. Deng, S.; Ling, L.; Zhang, C.; Li, C.; Zeng, T.; Zhang, K.; Guo, G. A systematic review on the current research of digital twin in automotive application. *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 180–191. [CrossRef]
31. Shoukat, M.U.; Yan, L.; Yan, Y.; Zhang, F.; Zhai, Y.; Han, P.; Nawaz, S.A.; Raza, M.A.; Akbar, M.W.; Hussain, A. Autonomous driving test system under hybrid reality: The role of digital twin technology. *Internet Things* **2024**, *27*, 101301. [CrossRef]
32. Ma, Y.; Du, R.; Abdelraouf, A.; Han, K.; Gupta, R.; Wang, Z. Driver digital twin for online recognition of distracted driving behaviors. *IEEE Trans. Intell. Veh.* **2024**, *9*, 3168–3180. [CrossRef]
33. Bishop, I.; Rohrmann, B. Subjective responses to simulated and real environments: A comparison. *Landsc. Urban Plan.* **2003**, *65*, 261–277. [CrossRef]
34. Freeman, J.; Avons, S.E.; Meddis, R.; Pearson, D.E.; Ijsselstein, W. Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence Teleoperators Virtual Environ.* **2000**, *9*, 149–164. [CrossRef]
35. Ijsselstein, W.A.; de Ridder, H.; Freeman, J.; Avons, S.E. Presence: Concept, determinants, and measurement. In *Human Vision and Electronic Imaging V*; Rogowitz, B.E., Pappas, T.N., Eds.; SPIE Proceedings: San Jose, CA, USA, 2000; pp. 520–529.
36. Kyrilitsias, C.; Michael-Grigoriou, D. Social interaction with agents and avatars in immersive virtual environments: A survey. *Front. Virtual Real.* **2022**, *2*, 786665. [CrossRef]
37. Loomis, J.M.; Blascovich, J.J.; Beall, A.C. Immersive virtual environment technology as a basic research tool in psychology. *Behav. Res. Methods Instrum. Comput.* **1999**, *31*, 557–564. [CrossRef] [PubMed]
38. Weibel, R.P.; Grübel, J.; Zhao, H.; Thrash, T.; Meloni, D.; Hölscher, C.; Schinazi, V.R. Virtual reality experiments with physiological measures. *J. Vis. Exp.* **2018**, *138*, 58318.

39. Higuera-Trujillo, J.L.; López-Tarruella Maldonado, J.; Llinares Millán, C. Psychological and physiological human responses to simulated and real environments: A comparison between photographs, 360° panoramas, and virtual reality. *Appl. Ergon.* **2017**, *65*, 398–409. [CrossRef]
40. Hu, B.; Ma, L.; Zhang, W.; Salvendy, G.; Chablat, D.; Bennis, F. Predicting real-world ergonomic measurements by simulation in a virtual environment. *Int. J. Ind. Ergon.* **2011**, *41*, 64–71. [CrossRef]
41. Gray, R. Transfer of training from virtual to real baseball batting. *Front. Psychol.* **2017**, *8*, 2183. [CrossRef] [PubMed]
42. Kalantari, S.; Rounds, J.D.; Kan, J.; Tripathi, V.; Cruz-Garza, J.G. Comparing physiological responses during cognitive tests in virtual environments vs. in identical real-world environments. *Sci. Rep.* **2021**, *11*, 10227. [CrossRef]
43. Botden, S.M.B.I.; Jakimowicz, J.J. What is going on in augmented reality simulation in laparoscopic surgery? *Surg. Endosc.* **2009**, *23*, 1693–1700. [CrossRef]
44. Carlson, K.J.; Gagnon, D.J. Augmented reality integrated simulation education in health care. *Clin. Simul. Nurs.* **2016**, *12*, 123–127. [CrossRef]
45. Bhagavathula, R.; Williams, B.; Owens, J.; Gibbons, R. The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2018**, *62*, 2056–2060. [CrossRef]
46. Narciso, D.; Melo, M.; Rodrigues, S.; Dias, D.; Cunha, J.; Vasconcelos-Raposo, J.; Bessa, M. Assessing the perceptual equivalence of a firefighting training exercise across virtual and real environments. *Virtual Real.* **2024**, *28*, 14. [CrossRef]
47. Mueller, J.A. Driving in a Simulator Versus On-Road: The Effect of Increased Mental Effort While Driving on Real Roads and a Driving Simulator. Ph.D. Thesis, Montana State University, Bozeman, Montana, 2015.
48. Blaauw, G.J. Driving experience and task demands in simulator and instrumented car: A validation study. *Hum. Factors J. Hum. Factors Ergon. Soc.* **1982**, *24*, 473–486. [CrossRef]
49. Dużmańska, N.; Strojny, P.; Strojny, A. Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Front. Psychol.* **2018**, *9*, 2132. [CrossRef] [PubMed]
50. Wynne, R.A.; Beanland, V.; Salmon, P.M. Systematic review of driving simulator validation studies. *Saf. Sci.* **2019**, *117*, 138–151. [CrossRef]
51. Zhang, S.; Zhao, C.; Zhang, Z.; Lv, Y. Driving simulator validation studies: A systematic review. *Simul. Model. Pract. Theory* **2025**, *138*, 103020. [CrossRef]
52. Pawar, N.M.; Velaga, N.R.; Sharmila, R.B. Exploring behavioral validity of driving simulator under time pressure driving conditions of professional drivers. *Transp. Res. Part F Traffic Psychol. Behav.* **2022**, *89*, 29–52. [CrossRef]
53. Törnros, J. Driving behavior in a real and a simulated road tunnel—A validation study. *Accid. Anal. Prev.* **1998**, *30*, 497–503. [CrossRef]
54. Bella, F. Validation of a driving simulator for work zone design. *Transp. Res. Rec. J. Transp. Res. Board* **2005**, *1937*, 136–144. [CrossRef]
55. Veldstra, J.L.; Bosker, W.M.; de Waard, D.; Ramaekers, J.G.; Brookhuis, K.A. Comparing treatment effects of oral THC on simulated and on-the-road driving performance: Testing the validity of driving simulator drug research. *Psychopharmacology* **2015**, *232*, 2911–2919. [CrossRef]
56. Carter, C.J.; Laya, O. Drivers visual search in a field situation and in a driving simulator. In *Vision in Vehicles VI*; Gale, A.G., Brown, I.D., Haselgrave, C.M., Taylor, S.P., Eds.; Elsevier Science Ltd.: Kidlington, UK, 1998.
57. Li, J.; Zhao, X.; Xu, S.; Ma, J.; Rong, J. The study of driving simulator validation for physiological signal measures. *Procedia-Soc. Behav. Sci.* **2013**, *96*, 2572–2583. [CrossRef]
58. Milleville-Pennel, I.; Charron, C. Driving for real or on a fixed-base simulator: Is it so different? An explorative study. *Presence* **2015**, *24*, 74–91. [CrossRef]
59. Stocco, A.; Pulfer, B.; Tonella, P. Mind the gap! a study on the transferability of virtual versus physical-world testing of autonomous driving systems. *IEEE Trans. Softw. Eng.* **2022**, *49*, 1928–1940. [CrossRef]
60. Giannakakis, G.; Grigoriadis, D.; Giannakaki, K.; Simantiraki, O.; Roniotis, A.; Tsiknakis, M. Review on psychological stress detection using biosignals. *IEEE Trans. Affect. Comput.* **2022**, *13*, 440–460. [CrossRef]
61. Klüver, M.; Herrigel, C.; Heinrich, C.; Schöner, H.-P.; Hecht, H. The behavioral validity of dual-task driving performance in fixed and moving base driving simulators. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *37*, 78–96. [CrossRef]
62. Hoffmann, S.; Krüger, H.-P.; Buld, S. *Avoidance of Simulator Sickness by Training the Adaptation to the Driving Simulation*; VDI Berichte: Düsseldorf, Germany, 2003; pp. 385–404.
63. Ma, S.; Zhang, W.; Yang, Z.; Kang, C.; Wu, C.; Chai, C.; Shi, J.; Li, H. Promote or inhibit: An inverted U-shaped effect of workload on driver takeover performance. *Traffic Inj. Prev.* **2020**, *21*, 482–487. [CrossRef] [PubMed]
64. Avetisyan, L.; Yang, X.J.; Zhou, F. Towards context-aware modeling of situation awareness in conditionally automated driving. *Int. J. Hum.-Comput. Interact.* **2025**, *42*, 291–306. [CrossRef]
65. Guan, D.; Wang, Q.; Wang, K.; Song, X. Fatigue and distracted driving recognition method based on multimodal information fusion. *IEEE Access* **2025**, *13*, 83815–83827. [CrossRef]

66. Reinhardt, T.; Schmahl, C.; Wüst, S.; Bohus, M. Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the Mannheim multicomponent stress test (mmst). *Psychiatry Res.* **2012**, *198*, 106–111. [CrossRef]
67. Schubert, C.; Lambertz, M.; Nelesen, R.; Bardwell, W.; Choi, J.-B.; Dimsdale, J. Effects of stress on heart rate complexity—A comparison between short-term and chronic stress. *Biol. Psychol.* **2009**, *80*, 325–332. [CrossRef]
68. Bernardi, L.; Wdowczyk-Szulc, J.; Valenti, C.; Castoldi, S.; Passino, C.; Spadacini, G.; Sleight, P. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *J. Am. Coll. Cardiol.* **2000**, *35*, 1462–1469. [CrossRef]
69. Clays, E.; De Bacquer, D.; Crasset, V.; Kittel, F.; De Smet, P.; Kornitzer, M.; Karasek, R.; De Backer, G. The perception of work stressors is related to reduced parasympathetic activity. *Int. Arch. Occup. Environ. Health* **2011**, *84*, 185–191. [CrossRef] [PubMed]
70. Matthiesen, A.; Majgaard, G.; Scirea, M. The process of creating interactive 360-degree VR with biofeedback. In *HCI International 2023 Posters*; Stephanidis, C., Antona, M., Ntoa, S., Salvendy, G., Eds.; volume 1836 of Communications in Computer and Information Science; Springer Nature: Cham, Switzerland, 2023; pp. 261–268.
71. Minen, M.; Gregoret, L.; Seernani, D.; Wilson, J. Systematic evaluation of driver's behavior: A multimodal biometric study. In *HCI International 2023 Posters*; Stephanidis, C., Antona, M., Ntoa, S., Salvendy, G., Eds.; volume 1836 of Communications in Computer and Information Science; Springer Nature: Cham, Switzerland, 2023; pp. 57–64.
72. Tavakoli, A.; Lai, N.; Balali, V.; Heydarian, A. How are drivers' stress levels and emotions associated with the driving context? a naturalistic study. *J. Transp. Health* **2023**, *31*, 101649. [CrossRef]
73. Boucsein, W. *Electrodermal Activity*; Springer: Boston, MA, USA, 2012.
74. Dey, A.; Phoon, J.; Saha, S.; Dobbins, C.; Billingham, M. A neurophysiological approach for measuring presence in immersive virtual environments. In *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*; IEEE: New York, NY, USA, 2020; pp. 474–485.
75. Liu, B.; Kralj, A.; Moyle, B.; Li, Y. Developing 360-degree stimuli for virtual tourism research: A five-step mixed measures procedure. *Inf. Technol. Tour.* **2024**, *26*, 485–519. [CrossRef]
76. Llinares, C.; Higuera-Trujillo, J.L.; Montañana, A. A comparative study of real and virtual environment via psychological and physiological responses. *Appl. Sci.* **2024**, *14*, 232. [CrossRef]
77. Liebherr, M.; Mueller, S.M.; Schweig, S.; Maas, N.; Schramm, D.; Brand, M. Stress and simulated environments: Insights from physiological marker. *Front. Virtual Real.* **2021**, *2*, 618855. [CrossRef]
78. Nyitrai, T.; Virág, M. The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Econ. Plan. Sci.* **2019**, *67*, 34–42. [CrossRef]
79. Ghasemi, A.; Zahediasl, S. Normality tests for statistical analysis: A guide for non-statisticians. *Int. J. Endocrinol. Metab.* **2012**, *10*, 486. [CrossRef] [PubMed]
80. Lumley, T.; Diehr, P.; Emerson, S.; Chen, L. The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health* **2002**, *23*, 151–169. [CrossRef] [PubMed]
81. Kwak, S.G.; Kim, J.H. Central limit theorem: The cornerstone of modern statistics. *Korean J. Anesthesiol.* **2017**, *70*, 144. [CrossRef]
82. Berg Hansen, K. Unpivoting columns to rows. In *Practical Oracle SQL: Mastering the Full Power of Oracle Database*; Apress: Berkeley, CA, USA, 2020; pp. 107–129.
83. Terumitsu, H.; Tetsuo, Y.; Tsuyoshi, T. Development of the driving simulation system Movic-t4 and its validation using field driving data. *Tsinghua Sci. Technol.* **2007**, *12*, 141–150. [CrossRef]
84. Kaptein, N.; Theeuwes, J.; van der Horst, R. Driving simulator validity: Some considerations. *Transp. Res. Rec. J. Transp. Res. Board* **1996**, *1550*, 30–36. [CrossRef]
85. Godley, S.T.; Triggs, T.J.; Fildes, B.N. Driving simulator validation for speed research. *Accid. Anal. Prev.* **2002**, *34*, 589–600. [CrossRef] [PubMed]
86. Sultanow, E. Simulator vs. Real World. 2024. Available online: https://github.com/Sultanow/fh-hof/tree/main/simulator_vs_realworld (accessed on 15 January 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Modelling and Forecasting Passenger Rail Demand in Slovakia Under Crisis Conditions with NARX Neural Networks

Anna Dolinayová ¹, Zdenka Bulková ^{1,*}, Jozef Gašparík ¹ and Igor Dömény ²

¹ Department of Railway Transport, Faculty of Operation and Economics of Transport and Communications, University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia; anna.dolinayova@uniza.sk (A.D.); jozef.gasparik@uniza.sk (J.G.)

² Faculty of Logistics and Crisis Management, Tomas Bata University, Studentské nám. 1532, 686 01 Uherské Hradiště, Czech Republic; domeny@utb.cz

* Correspondence: zdenka.bulkova@uniza.sk; Tel.: +421-41-513-34-11

Abstract: Transportation systems are particularly vulnerable to disruptions such as pandemics, which create significant challenges for maintaining efficiency, safety, and service quality. This study focuses on rail passenger transport in the Slovak Republic and develops a simulation framework to evaluate system performance under crisis conditions. Weekly data from the national rail operator for the period 2019–2021 were combined with information on governmental restrictions, standardized into a five-level framework. A non-linear autoregressive model with exogenous inputs (NARX), implemented and validated in MATLAB R2021b (MathWorks, Natick, MA, USA), was applied to simulate the impact of restrictive measures on passenger demand. The results revealed a strong relationship between the severity of measures and ridership levels, with the most significant effects observed in education, workplace access, movement limitations, and retail. For instance, during complete school closures, passenger volumes declined by up to 75% relative to the pre-pandemic baseline. Based on the simulation outcomes, recommendations were formulated for adapting railway operations, including dynamic adjustments of transport capacity (10–40%) according to restriction levels. The proposed modelling and simulation approach offers transport authorities a cost-effective tool for scenario testing, disruption management, and the design of resilient passenger rail systems capable of adapting to crises and uncertainties.

Keywords: railway passenger transport; pandemic; disruption management; NARX neural network; mobility modelling; sustainable and resilient transportation

1. Introduction

Population mobility is one of the key determinants of economic and social development. In a globalized world, where frequent movement of people at regional, national, and international levels is expected, societies have also become increasingly vulnerable to various crisis situations [1]. Among the most significant of these was the COVID-19 pandemic, which profoundly affected the functioning of transport systems and exposed their weaknesses in terms of flexibility and preparedness to respond to unforeseen circumstances [2].

Given today's high levels of mobility, it is reasonable to expect that crisis situations caused by epidemics will continue to emerge [3,4], whether at global, continental, national, or regional scales [5,6]. It is therefore crucial for transport authorities to respond effectively to ensure safe and high-quality mobility for the population, while at the same time

considering the economic sustainability of transport services [7,8]. A sound forecast of transport demand development is fundamental for such decision-making. In this research, data from the COVID-19 pandemic were used to develop a model for predicting mobility in rail transport, with a specific focus on the regional dimension of epidemic situations.

In rail passenger transport, which plays a crucial role in ensuring the daily mobility of the population, pandemic restrictions resulted in a dramatic decline in travel demand [9–12]. Measures affecting education, workplace access, free movement, and retail substantially reduced passenger numbers, with significant economic consequences for railway operators as well as for the state [13,14]. At the same time, it became evident that traditional approaches to transport planning and organization are insufficient to cope with crisis situations of this scale. Although numerous studies have investigated the impact of the COVID-19 pandemic on transport demand, most focus on countries with railway networks, mobility patterns, and policy frameworks that differ from those of Central and Eastern Europe. For the Slovak Republic, where long-distance commuting and a strong regional dimension are key features, no comprehensive, data-driven simulation of passenger rail demand under crisis conditions has yet been published. This study addresses this gap by developing and validating a neural network-based NARX model tailored to Slovak regional data and pandemic measures. The approach not only quantifies the relationship between restrictions and rail ridership but also provides practical recommendations for adaptive capacity management—a perspective missing in existing European research.

Accurate forecasting of transport demand during a pandemic or other extraordinary situation is therefore an essential prerequisite for effective decision-making by transport authorities. Modelling population mobility while accounting for the dynamics of restrictive measures not only provides deeper insights into passenger behaviour [15] but also enables the design of interventions aimed at reducing social costs and optimizing the operation of the transport system [16,17].

Particular attention in this study is devoted to the regional perspective, as movements between neighbouring regions and districts were heavily restricted during the pandemic. The objective is to ensure that the proposed model can also be applied in crisis situations of a more local character.

The development of the rail transport model was divided into several partial steps:

- A review of the scientific literature addressing the modelling of pandemic impacts on mobility (passenger numbers, capacity utilization, etc.);
- An analysis of adopted pandemic measures and their transformation into a form suitable for model development;
- Collection and processing of passenger mobility data in rail transport;
- Model creation;
- Evaluation of model results.

Based on the constructed data-driven model, which predicts the evolution of rail passenger mobility according to the severity of implemented measures, recommendations were subsequently formulated for adjusting technological processes in rail transport and for adapting railway operations to regional, national, and international conditions.

The aim of this study is therefore to analyse the demand for rail passenger transport in the Slovak Republic during the COVID-19 pandemic and based on empirical data, to develop a modelling framework capable of predicting mobility dynamics under different levels of restrictions. Although the primary analysis uses national-level patterns, the approach is designed to capture potential regional differences and remains applicable to region-specific crisis scenarios. The findings are intended to contribute to more effective transport policy planning, improved capacity management for rail operators, and enhanced resilience of the transport system to future crises.

This study contributes to the broader field of transportation system modelling and simulation by demonstrating how machine learning techniques can be applied to evaluate system performance under disruption. The proposed framework not only captures the nonlinear relationship between restrictive measures and passenger demand but also enables scenario testing, supporting decision-makers in planning adaptive and cost-effective transport operations. In line with the objectives of this Special Issue, the findings highlight how simulation-based approaches can enhance resilience, optimize resource allocation, and reduce the vulnerability of transportation systems to future crises.

2. Literature Review

The issue of modelling the impact of the COVID-19 pandemic on population mobility has been the subject of intensive scientific research since 2020. Recent research confirms that modelling and simulation have become indispensable tools for assessing the resilience and performance of transportation systems under disruptive conditions such as the COVID-19 pandemic. Beyond quantifying passenger losses, these approaches enable scenario testing, provide insights into behavioural responses, and help decision-makers design adaptive strategies for crisis management. Against this backdrop, the present literature review summarizes the diversity of methods employed internationally and highlights the gap for studies focused on rail passenger transport in Central and Eastern Europe, particularly under regional conditions. Table 1 provides a concise overview of several studies [18–37] of the modelling approaches used to examine the effects of the COVID-19 pandemic on mobility, categorized by author and the country for which the model was developed.

Table 1. Modelling the impact of COVID-19 on transport—literature review.

Author	Case Study	Domain	Dependent Variable	Explanatory Variable	Modelling Approach
Singh et al. (2023) [18]	Netherlands	Rail passenger travel	Number of train passengers, overcrowding level, delay schedule, departure time	Overcrowding, occupancy, delay, fare discount	Latent class cluster model
Deplomo et al. (2021) [19]	Philippines	Passenger transport	Number of passengers	–	Polynomial regression
Lei et al. (2020) [20]	China	Rail passenger travel	Number of train passengers	–	Spreading model
Elias and Zatmeh-Kanj (2021) [21]	Israel	Travel risk and decisions	Travel risk and decision to use rail	Frequency, socio-economic characteristics, demographic factors	Structural equation model
Ferreira et al. (2022) [22]	Germany	Travel demand	Number of trips	Frequency	Conditional logit model
Zhang et al. (2021) [23]	Australia, Canada, Japan, New Zealand, UK, USA	Policy impacts on mobility	Political measures, passenger movements	–	Dynamic Bayesian multilevel generalized structural equation model
Guo et al. (2021) [24]	Honolulu (Hawaii)	Public transport ridership	Number of passengers	Frequency	Negative binomial models, Pearson correlation
Bian et al. (2021) [25]	New York, Seattle (USA)	Policy delays and mobility	Infections, vehicle volume, passenger volume, bicycle counts	Policy delay variables	Odds ratio, regression structure, Bayesian change-point detection

Table 1. Cont.

Author	Case Study	Domain	Dependent Variable	Explanatory Variable	Modelling Approach
Tan and Ma (2020) [26]	China	Selected rail transport	Travel characteristics	Fare attributes, COVID-19 perception	Logit model, binary logistic regression
Molin and Kroesen (2022) [27]	Netherlands	Travel behaviour and safety	Infection risk perception, referendum voting	Political measures	Latent class choice model, mediation model, linear regression, binary logistic regression
Ding and Zhang (2021) [28]	Japan	Policy making	Behavioural changes of respondents	–	Dynamic structural equation model
Carteni et al. (2021) [29]	Italy	Accessibility	Transport accessibility	Socio-economic, territorial and environmental factors	Multiplicative linear regression
Coppola and De Fabiis (2021) [30]	Italy	Rail demand	Boarding and alighting	Timetables, vehicles, onboard layout	Demand forecasting model
Ding et al. (2023) [31]	Shanghai (China)	Passenger experience	Regular services, pandemic prevention, psychological distancing, safety perception, satisfaction	–	Structural equation model
Haque and Hamid (2022) [32]	India	Capacity allocation	Seat assignment	Capacity utilization	Mixed-integer linear programming
Vichiensan et al. (2021) [33]	Bangkok (Thailand)	Trust and awareness	Trust, awareness, distancing, health measures, contact-tracing app	–	Factor analysis, structural equation model
Navarrete-Hernandez et al. (2023) [34]	London (UK), Milan (Italy), Santiago (Chile)	Perceptions of infection risk	Perceived risk	Government measures	Photo-simulation approach in randomized controlled trial
Thaithatkul et al. (2023) [35]	Bangkok (Thailand)	Travel frequency	Frequency of trips	Online activities, socio-economic factors, COVID-19-related behaviours and attitudes	Random effects models, WBRE
Haque and Hamid (2023) [36]	India	Train capacity planning	Distance, urban planning, boarding and alighting	Train capacity	Mixed-integer programming (MIP), integrated MILP
Abdullah et al. (2021) [37]	Pakistan	Mode choice	Mode preference	Mode of transport	Exploratory factor analysis (EFA)

While several studies have used machine learning to analyse pandemic-related changes in travel demand, most focused on aggregated urban mobility or road traffic and rarely addressed national or regional passenger rail. Only a few works experimented with neural networks, but they typically used simple architectures (feedforward or LSTM) without explicit integration of government intervention indicators. Our approach differs by (i) applying a nonlinear autoregressive model with exogenous inputs (NARX) tailored to rail demand, (ii) directly encoding the intensity of restriction measures as explanatory variables, and (iii) testing applicability in a small, highly centralized railway system such as Slovakia. This extends previous machine-learning applications beyond descriptive analysis and towards actionable demand forecasting under crisis conditions. The literature review (Table 1) shows that most studies focused on quantifying the decline in passenger numbers in rail or urban public transport in relation to the introduction of pandemic restrictions. A significant share of research analysed the impact of these measures on travel behaviour, trip frequency, or passengers' risk perception. Studies such as [18,27,38] examined factors

influencing passenger decision-making, highlighting the key role of overcrowding and the subjective perception of safety. Other studies [19,20,39,40] concentrated on predicting passenger volumes through regression or spreading models.

A specific strand of research has been the evaluation of the effectiveness of policy and organizational measures. Studies such as [23,34] point out that the type and timing of restrictions significantly shape mobility levels, with public trust playing an important role. Similarly, the studies [29,41] emphasized that transport accessibility and territorial characteristics are decisive factors in both the spread of the virus and the subsequent decline in mobility.

The literature also demonstrates that the pandemic's impact was not homogeneous. Regional differences, as documented in studies such as [24] for Honolulu or [25] for New York and Seattle, indicate that local conditions and transport infrastructure strongly influence passenger behaviour. Equally important is the psychological dimension, addressed by [28], who argue that changes in population behaviour are crucial in shaping transport policy during crises.

As shown in Table 1, existing studies predominantly analyse pandemic-related changes in passenger transport at a national or aggregated urban level and rarely address the specific conditions of Central and Eastern Europe. None of the reviewed works explicitly modelled the Slovak passenger rail system, which features a strongly centralized network and high regional commuting flows, nor did they integrate governmental restriction measures into a nonlinear predictive framework. Our study addresses this gap by developing a regionally applicable, data-driven NARX neural network model tailored to Slovak pandemic conditions and evaluating its potential for adaptive railway capacity management.

Previous studies have thus provided a wide range of methodological approaches and empirical evidence on the impact of the pandemic on mobility. However, most of this research has been carried out in countries with transport systems and mobility structures that differ from those of the Slovak Republic. This highlights the need to develop a model reflecting the specific regional and national conditions, one that can not only analyse the impacts of the pandemic but also propose measures for more effective management of rail transport in future crisis situations.

3. Data and Research Methodology

A high-quality database is a fundamental prerequisite for developing a reliable model of rail passenger transport demand under crisis conditions. In our research, we worked with two main types of data:

- Qualitative data, representing government measures introduced to contain the spread of COVID-19;
- Quantitative data, capturing passenger mobility in rail transport.

These two sources of information had to be aligned to analyse the relationship between the dynamics of implemented measures and changes in passenger volumes. At the same time, it was necessary to transform and harmonize the data into a unified structure suitable for processing in the modelling environment.

We worked with weekly data, which represent an appropriate compromise between detail and statistical stability. While daily data would provide higher resolution, they are characterized by substantial variability caused by seasonality and short-term fluctuations. Monthly data, on the other hand, would not allow us to capture passengers' rapid responses to changing pandemic measures.

On the quantitative side, we used a comprehensive database provided by the national rail operator (ZSSK), which accounts for more than 97% of passenger rail services in the Slovak Republic. This ensures that the data are sufficiently representative for the entire

country. On the qualitative side, we drew from official government sources, primarily the COVID automat, which specified the type and severity of restrictions during different periods. The resulting integrated database provides the foundation for linking adopted measures with changes in rail mobility, thus forming the basis of the modelling framework.

The methodological approach of this study integrates the systematic collection and transformation of data with advanced time-series modelling techniques. First, qualitative and quantitative data on rail passenger mobility and pandemic measures were gathered and harmonized into a unified database. Then, seasonality was analysed and removed to obtain stable input variables. These prepared datasets were subsequently applied to a Nonlinear Autoregressive model with Exogenous Inputs (NARX), a neural network-based technique well-suited for forecasting non-stationary and nonlinear time series. This process ensured that the modelling framework is robust, data-driven, and capable of reflecting the dynamic effects of government restrictions on passenger demand.

3.1. COVID Automat Data—Slovak Republic

During the COVID-19 pandemic, several nationwide measures were introduced in the Slovak Republic to limit the spread of the virus. Since 2021, the so-called COVID automat (it is a government traffic-light system defining the intensity of restrictions based on the epidemiological situation in individual districts) was applied. This tool represented the primary source of qualitative data for our research.

The COVID automat was modified several times during the observation period. From January to August 2021, it operated in a seven-level version, while from August until the end of November 2021, it was transformed into a five-level system. For modelling purposes, it was necessary to harmonize these differences; therefore, the original seven-level system was consolidated into a unified five-level framework, as shown in Table 2.

Table 2. COVID automat in the Slovak Republic.

7-Level COVID Automat	5-Level COVID Automat	Colour
Monitoring	Monitoring	Green
First level of vigilance	Vigilance	Orange
Second level of vigilance		
First level of warning	First level of threat	Red
Second level of warning		
Third level of warning	Second level of threat	Burgundy
Fourth level of warning	Third level of threat	Black

This transformation ensured data consistency and enabled comparability across different periods. The framework incorporated measures in key areas that most significantly influenced passenger mobility, in particular:

- Restriction of movement;
- Entry into employment;
- Limiting the operating hours of non-essential operations and services;
- Education;
- Mass events;
- Shops;
- Services;
- Shopping centres;
- Restaurants;
- Accommodation (hotels, pensions, etc.);
- Taxi services.

The data was processed on a weekly basis for all measures using MS Excel spreadsheets. An example of data processing for schools is illustrated in Figure 1.

Region	Number of the week in 2021										
	37	38	39	40	41	42	43	44	45	46	47
Bratislava	0	1	1	1	1	2	2	3	3	3	4
Tnava	2	3	3	3	3	3	3	3	3	3	4
Trenčín	2	2	2	2	2	2	2	3	3	3	4
Nitra	1	2	2	2	2	2	2	4	4	4	4
Žilina	2	2	2	3	3	4	4	4	4	4	4
Banská Bystrica	2	3	3	3	3	3	3	4	4	4	4
Prešov	2	3	3	3	3	4	4	4	4	4	4
Košice	2	3	3	3	3	3	4	4	4	4	4
Daily cases	351	548	888	1207	1607	1942	2438	3646	5302	9103	10,780
Hospitalized patients	126	203	329	453	708	824	966	1223	1774	2904	3188
Revenues to hospitals	77	141	249	376	544	675	736	854	1454	x	x

Figure 1. Illustration processing of measures to prevent the spread of the COVID virus for schools in the Slovak Republic.

The data were processed in weekly intervals for all regions of Slovakia. Sources included the Epidemiological Information System (EPIS) and records from the Public Health Authority of the Slovak Republic, which provided continuous updates on the development of the pandemic situation at the district level [42].

The processing of data involved coding the measures into numerical values according to their severity. This step was necessary to enable their integration as input variables into the modelling framework. A graphical representation of the processing (for the education sector—Figure 1) illustrates how the measures were systematically transformed into the database. The resulting set of qualitative data forms the foundation for analysing the relationship between the intensity of measures and the number of passengers transported in individual regions of the Slovak Republic.

3.2. Data on the Mobility of Rail Passenger Transport

Many authors addressing similar issues have relied on freely available data sources (for example Google Mobility Reports). While such datasets provide useful insights into general mobility trends, they do not allow for identifying which specific mode of transport was used. Since the objective of our research is to propose technological and policy measures in the transport sector for future epidemics or pandemics, mode-specific data were required.

On the quantitative side, we used a database provided by the national rail operator, Železničná spoločnosť Slovensko (ZSSK), which accounts for more than 97% of passenger rail services in the country. For this reason, the dataset is sufficiently representative both at the national and regional levels.

The data were supplied in weekly intervals for the period 2019–2021 and structured as follows:

- The start of validity of the travel document (week of the given year);
- Origin region;
- Destination region.

Thus, we had at our disposal a comprehensive database covering passenger transport between regions as well as within individual regions. The territory of the Slovak Republic is divided into eight higher territorial units/self-governing regions (HTUs). Figure 2 shows the territorial division of the Slovak Republic together with railway lines.



Figure 2. Territorial division of the Slovak Republic.

The national rail passenger carrier provided a database covering three years, with 52 weeks recorded in 2019 and 2021 and 53 weeks in 2020. We had at our disposal a dataset capturing passenger transport flows from each region to every other region, as well as within individual regions. Each week, it was possible to monitor 64 combinations (variations with repetition) between the regions. The resulting database contained more than 10,000 records. An example of the database structure is presented in Table 3.

Table 3. Structure of the database of the number of passengers transported.

Starting Region	Beginning of the Document Validity		Destination Region	Number of Passengers
	Week	Year		
Bratislava	1	2019	Košice	...
Bratislava	2	2019	Košice	...
Bratislava	...	2019	Košice	...
Bratislava	52	2019	Košice	...
Bratislava	1	2019	Prešov	...
Bratislava	2	2019	Prešov	...
Bratislava	...	2019	Prešov	...
Bratislava	52	2019	Prešov	...
...	...	...	...	...

From a mathematical perspective, these combinations represent variations with repetition, which can be expressed by the Formula (1):

$$V'_k(n) = n^k \quad (1)$$

where:

n number of elements (number of HTUs),

k k -element variation (in this case, from one HTU to another; $k = 2$).

The number of data entries for one week is 64, which means that we had a total database of 10,048 records available. When processing the data, it was necessary to consider the actual geography of the routes. For example, if a journey occurred between the Bratislava and Košice regions, we assumed the use of the main railway corridor via Trnava,

Trenčín, Žilina, and Prešov. This approach was based on empirical data from ZSSK, which showed that the alternative route via Nitra and Banská Bystrica accounted for less than 5% of all trips. Other combinations between regions were processed in an analogous way, with the assignment methodology ensuring that each trip was allocated to all regions it passed through.

The results of data processing showed that the development of passenger volumes was similar in most regions. This trend was related to the fact that, during 2020, measures were implemented predominantly at the national level. Regional differences began to appear only in 2021, when the COVID automat introduced differentiated measures according to the epidemiological situation in individual districts. However, differences between regions remained relatively small. The resulting database thus provided a robust foundation for further processing and subsequent modelling of the relationship between pandemic measures and passenger mobility in rail transport. The database in Excel is available from the authors. Since the dataset is subject to trade secrecy, it will be provided upon request only for the purpose of verifying the validity of the data, while ensuring the integrity of the trade secret.

3.3. Adaptation of the Data for the Model

Considering that our goal was to examine how measures to prevent the spread of the virus on mobility are reflected in individual regions, we had to adjust the data obtained to know the mobility in each region by individual weeks. For example, if we had data on traffic between the Bratislava and Žilina regions (Figure 2)—we had to assign these passengers to the Trnava and Trenčín regions as well. To address this issue, we applied a methodology of allocating passengers to all regions that the journey would realistically pass through from a geographical perspective. This ensured that not only the origin and destination were considered, but also the intermediate regions traversed by the route. For instance, in the case of the Trenčín region, the weekly data included the following combinations of starting and destination regions, where Trenčín was either the origin, destination, or a transit region along the route.

- Starting region—Bratislava; destination regions—Trnava, Žilina, Prešov, Košice;
- Starting region—Trnava; destination regions—Trenčín, Žilina, Prešov, Košice;
- Starting region—Trenčín; destination regions—Bratislava, Trnava, Nitra, Trenčín, Banská Bystrica, Žilina, Prešov, Košice;
- Starting region—Nitra; destination regions—Trenčín, Žilina;
- Starting region—Banská Bystrica; destination regions—Trenčín, Trnava;
- Starting region—Prešov; destination regions—Trenčín, Trnava, Bratislava;
- Starting region—Košice; destination regions—Trenčín, Trnava, Bratislava.

If the passenger flow occurred between the Bratislava and Košice regions, or in the opposite direction, we assumed that most passengers would travel along the main corridor passing through Trnava, Trenčín, Žilina, and Prešov. This assumption was based on empirical data from ZSSK, which confirmed that the alternative route through Nitra and Banská Bystrica accounted for less than 5% of total trips.

This methodological adjustment ensured that the resulting database reflects not only the origin and destination of trips, but also the intermediate regions realistically traversed by passengers. Such an approach was essential for making the model usable at the regional level, and not merely at the national scale. The development of passenger volumes across individual regions is shown in Figure 3.

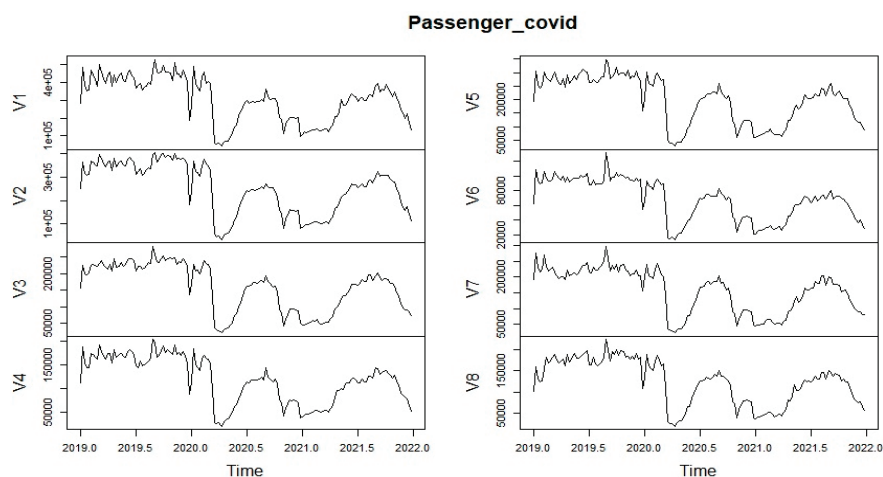


Figure 3. Number of passengers by region.

The number of passengers showed a similar trend across individual regions (V1—Bratislava, V2—Trnava, V3—Trenčín, V4—Nitra, V5—Žilina, V6—Banská Bystrica, V7—Prešov, V8—Košice). This was primarily because, from the beginning of the pandemic, identical measures were introduced nationwide. Only from 2021 onward were measures applied at the regional level, and even then, the differences between regions remained very small. In this form, the adapted database provided a robust foundation for modelling the relationship between pandemic-related measures and the development of mobility in rail passenger transport.

4. Modelling Framework

The objective of the modelling was to investigate the relationship between the number of passengers transported by rail and the measures introduced during the COVID-19 pandemic. Since the data were available on a weekly basis, it was necessary to apply time series methods, eliminate seasonal effects, and subsequently propose an appropriate model. Traditional econometric approaches such as regression or autoregressive integrated moving average (ARIMA) models are often suitable for capturing linear dependencies in transport data. However, the extraordinary circumstances of the pandemic introduced strong nonlinearities and abrupt shifts in passenger behaviour, caused by the dynamic implementation of government measures. These conditions required a more flexible modelling framework capable of capturing complex relationships and delayed effects. For this reason, we applied a nonlinear autoregressive model with exogenous inputs (NARX) implemented through a neural network architecture. The NARX approach has proven effective in previous mobility studies, as it allows the incorporation of external inputs (in this case, the severity of pandemic measures) and is well-suited for forecasting nonlinear and nonstationary time series. This choice ensured that the model could realistically reflect the interplay between policy interventions and passenger demand, while maintaining sufficient accuracy for scenario testing and policy evaluation.

4.1. Decomposition of Passenger Transportation Time Series

The evolution of the number of passengers revealed the presence of seasonal components that could distort the modelling results. Therefore, the original time series were decomposed into three parts: trend, seasonal, and random components. A simple seasonal variation model was applied to eliminate the seasonal effect.

The use of ARMA/ARIMA models was not feasible, as the stationarity test (Figure 4) confirmed the non-stationary character of the data. Specifically, the null hypothesis of

stationarity was rejected ($p < 0.05$), which indicated that the dataset contained significant structural changes that standard autoregressive models could not properly capture.

```
> tseries::adf.test(Passenger_covid[,1])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 1]
Dickey-Fuller = -3.0116, Lag order = 5, p-value = 0.1546
alternative hypothesis: stationary

> tseries::adf.test(Passenger_covid[,2])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 2]
Dickey-Fuller = -2.7334, Lag order = 5, p-value = 0.2706
alternative hypothesis: stationary

> tseries::adf.test(Passenger_covid[,3])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 3]
Dickey-Fuller = -2.6809, Lag order = 5, p-value = 0.2925
alternative hypothesis: stationary

> tseries::adf.test(Passenger_covid[,4])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 4]
Dickey-Fuller = -2.7423, Lag order = 5, p-value = 0.2669
alternative hypothesis: stationary

> tseries::adf.test(Passenger_covid[,5])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 5]
Dickey-Fuller = -2.8704, Lag order = 5, p-value = 0.2135
alternative hypothesis: stationary

> tseries::adf.test(Passenger_covid[,6])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 6]
Dickey-Fuller = -2.7329, Lag order = 5, p-value = 0.2708
alternative hypothesis: stationary

> tseries::adf.test(Passenger_covid[,7])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 7]
Dickey-Fuller = -2.9363, Lag order = 5, p-value = 0.186
alternative hypothesis: stationary

> tseries::adf.test(Passenger_covid[,8])
Augmented Dickey-Fuller Test
data: Passenger_covid[, 8]
Dickey-Fuller = -2.9545, Lag order = 5, p-value = 0.1785
alternative hypothesis: stationary
```

Figure 4. Test of stationarity of the passenger transport time series by individual region.

As shown by the results of the stationarity test, $p < 0.05$, and therefore the null hypothesis was rejected (H_0 : the time series is stationary). This confirmed the non-stationary nature of the dataset, further justifying the need for decomposition prior to modelling. Figure 5 presents the decomposition of the passenger transport time series into its individual components for the Bratislava region, illustrating the separation into trend, seasonal, and random parts.

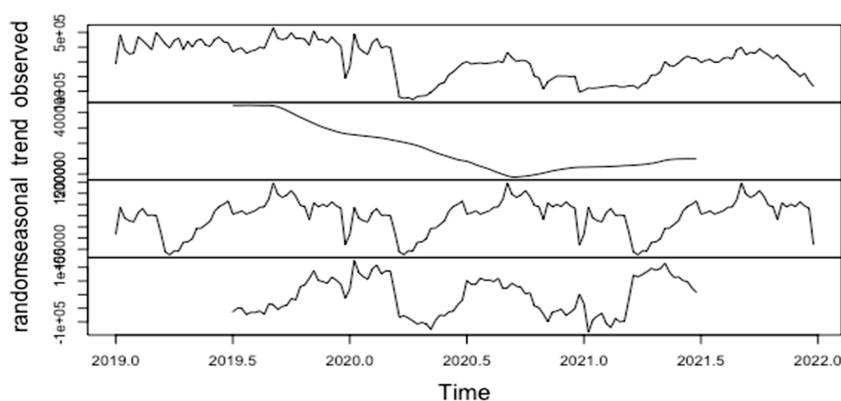


Figure 5. Decomposition of passenger transport time series—Bratislava region.

Figure 5 shows the components of the time series from top to bottom—original data, trend, seasonal component, and random component. Analogous decompositions of the time series were carried out for all regions. Figure 6 presents the seasonally adjusted passenger transport data, which remove recurring seasonal effects and allow a clearer comparison of mobility trends across regions.

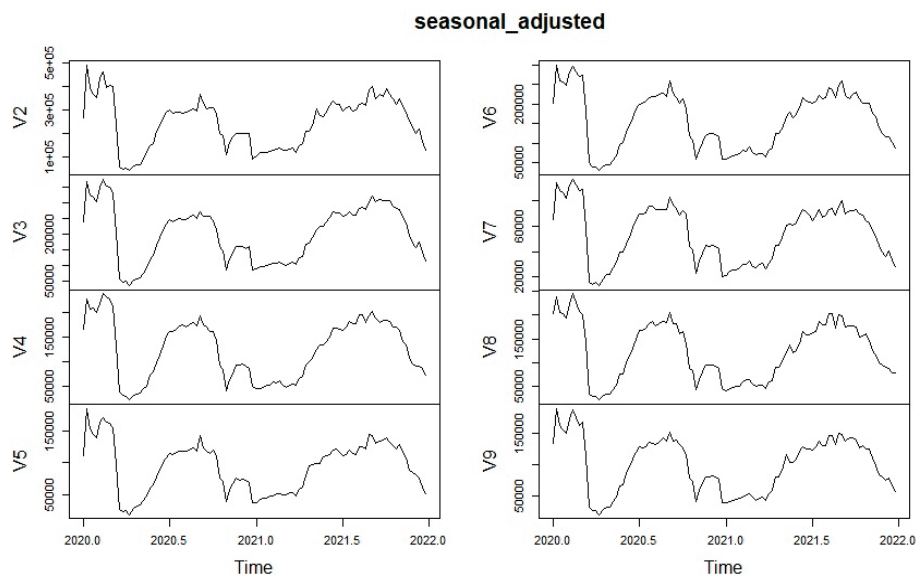


Figure 6. Seasonally adjusted data on mobility in rail passenger transport.

As illustrated in Figure 6, the trend in the development of mobility in rail passenger transport during the COVID-19 pandemic (2020–2021) was very similar across individual regions of the Slovak Republic (V2—Bratislava, V3—Trnava, V4—Trenčín, V5—Nitra, V6—Žilina, V7—Banská Bystrica, V8—Prešov, V9—Košice).

4.2. Model Creation

In mathematics and statistics, the Levenberg–Marquardt algorithm (LMA or simply LM), also known as damped least squares (DLS), is widely used to solve nonlinear least squares problems [43]. This algorithm was selected for training the neural network due to its efficiency, robustness, and widespread application in nonlinear regression tasks. The general equation of the neuron model can be expressed as Formula (2):

$$y(t) = f(x(t-1), \dots, x(t-d), y(t-1), \dots, y(t-d)) \quad (2)$$

To construct the model, we employed a Nonlinear Autoregressive model with Exogenous Inputs (NARX) neural network, as this framework allows the capture of nonlinear dependencies between past values of the output variable and external input factors. Such an approach is particularly well suited for transport demand data, which are strongly affected by sudden external shocks such as pandemic measures.

The input to the model consisted of seasonally adjusted passenger transport data (Figure 6) and the set of measures introduced in individual policy areas. Since these measures (education, mass events, entry into employment, etc.) differed not only in time (nationwide measures versus the COVID automat), but also in severity across areas, it was necessary to transform them into a binary variable for modelling purposes, as follows:

- Variable “0” if there was no change in the measure in the given week compared to the previous week, or if the measure was relaxed;
- Variable “1” if the degree of the measure was tightened according to the COVID automat in the given week compared to the previous week.

A model was developed in MATLAB using a neural network based on the seasonally adjusted time series of the number of passengers transported in individual regions. The NARX neural network was implemented in MATLAB using the following steps:

- Import time series of weekly passenger counts and exogenous inputs representing government measures.

- Pre-process data by removing seasonality and normalizing values.
- Define a NARX architecture with one hidden layer of 10 neurons and time delays of 1–4 weeks.
- Split the dataset into training (70%), validation (15%), and testing (15%) sets.
- Train the network using the Levenberg–Marquardt backpropagation algorithm.
- Evaluate performance using R and MSE on validation and test sets.

In this study, model performance was evaluated using the coefficient of determination (R) and mean squared error (MSE) for training, validation, and test sets. While these metrics provide a first indication of predictive quality, further robustness checks such as k-fold cross-validation, sensitivity analysis to input variables, and benchmarking against simpler models (for example ARIMA, multiple regression) were not performed in the current version. This will be an important part of future research to better substantiate the selection of NARX for this type of demand modelling. Figure 7 shows workflow of the NARX modelling process.



Figure 7. Workflow of the NARX modelling process.

The diagram summarizes the main steps used to build and validate the nonlinear autoregressive network with exogenous inputs (NARX) for forecasting passenger rail demand. Raw weekly passenger counts, and government restriction indicators are first collected and pre-processed (seasonality removal, normalization). The dataset is then split into training, validation, and testing subsets. A NARX network architecture with selected time delays and hidden neurons is configured and trained using the Levenberg–Marquardt backpropagation algorithm. Finally, the model performance is evaluated on validation and test sets and applied to scenario-based demand forecasting.

This standard division of data ensured that the model was not only fitted to the training subset but also validated and tested for predictive accuracy, thereby increasing the reliability and generalizability of the results. Prior to the creation of the final model, we verified the distribution of random errors (Figure 8) and the autocorrelation of residuals (Figure 9), to confirm that the neural network’s outputs met the basic assumptions of randomness and independence of errors.

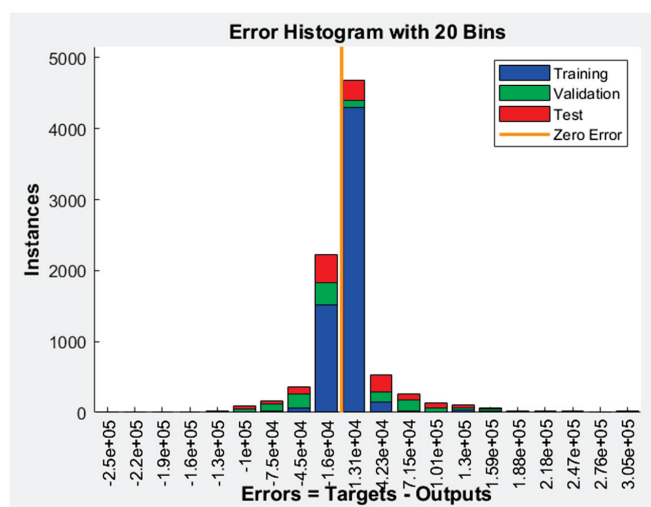


Figure 8. Randomness distribution of the errors.

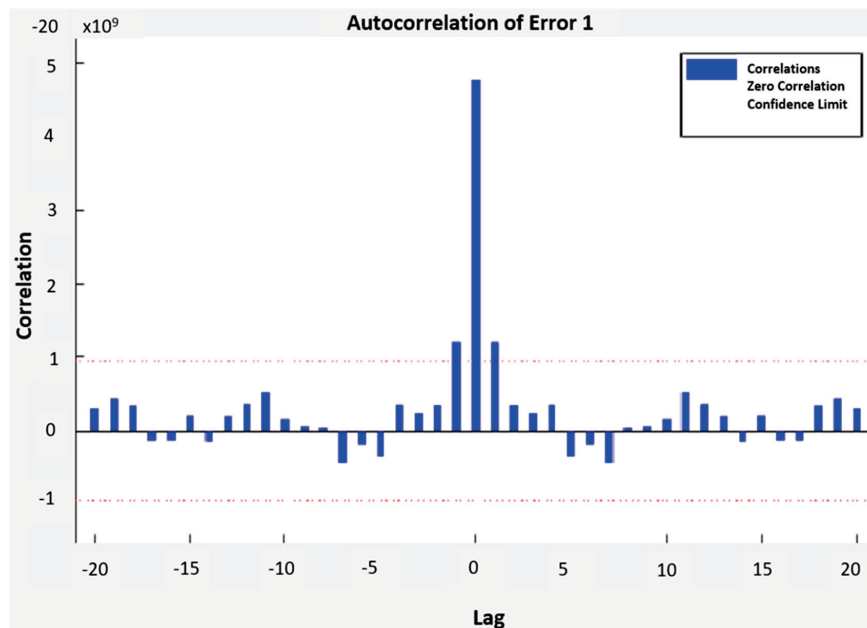


Figure 9. Autocorrelation of the random errors.

The distribution of residual errors corresponded approximately to a normal (Gaussian) probability distribution, and the autocorrelation values obtained from network training did not exceed the critical thresholds. These diagnostics confirmed that the proposed model can be reliably used to investigate the relationship between the number of transported passengers and the degree of measures introduced. The main advantage of the proposed NARX neural network model lies in its ability to flexibly adapt to nonlinear changes in passenger demand under varying levels of restrictions. Nevertheless, a limitation of this approach is the occasional occurrence of extreme predicted values, which naturally result from the neural network methodology when multiple measures are applied simultaneously.

5. Results

This section presents the outcomes of the data-driven analysis and the NARX neural network model applied to rail passenger transport in the Slovak Republic. First, the relationship between pandemic measures and passenger mobility is described and illustrated using comparative graphs for selected policy areas. Then, the predictive performance of the model is evaluated, and the regression outputs are presented. Finally, the results are broken down by individual regions and key measures (education, workplace entry, restrictions on movement, and retail), providing a basis for formulating recommendations for railway capacity management during crisis conditions.

5.1. Comparison of Measures and Mobility Data in Rail Passenger Transport

Based on the development of the number of passengers transported in individual regions, comparative graphs were constructed according to the levels of restrictions in areas assumed to influence rail passenger transport. These included measures related to:

- Schools (education sector);
- Entry into employment (workplace entry);
- Restriction of movement (limitation of movement);
- Shops;
- Accommodation facilities.

Figure 10 shows that education restrictions had the most significant impact on rail demand, with complete school closures reducing passenger numbers by about 75% compared to pre-crisis levels.

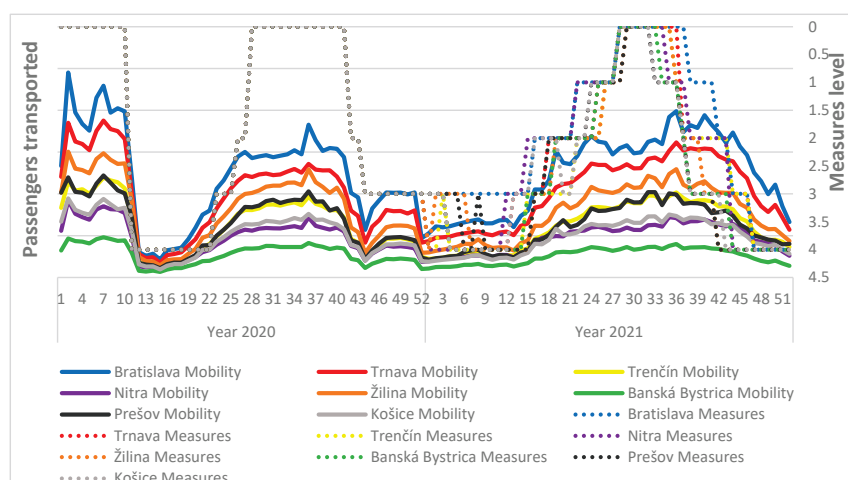


Figure 10. Comparison of rail passenger mobility according to the levels of restrictions in the education sector.

Although Figures 10–19 may appear visually similar at first glance, each plot represents a distinct government measure and its specific relationship to passenger rail demand. Subtle differences in slopes, turning points, and temporal responses reveal how individual interventions, such as school closures, workplace restrictions, retail limitations, and mobility bans, influenced demand dynamics. Education-related restrictions produced the steepest and most immediate declines in passenger volumes, confirming the major impact of student commuting on rail travel. Workplace access limitations caused a slower but persistent decrease in demand, consistent with the gradual adoption of remote work. Mobility bans triggered sharp drops followed by rapid rebounds once restrictions were lifted, indicating a high latent demand for travel. Retail and leisure closures had more moderate, short-term effects. These patterns suggest that future capacity planning must pay special attention to school and workplace policy signals, as they drive the most pronounced fluctuations in ridership.

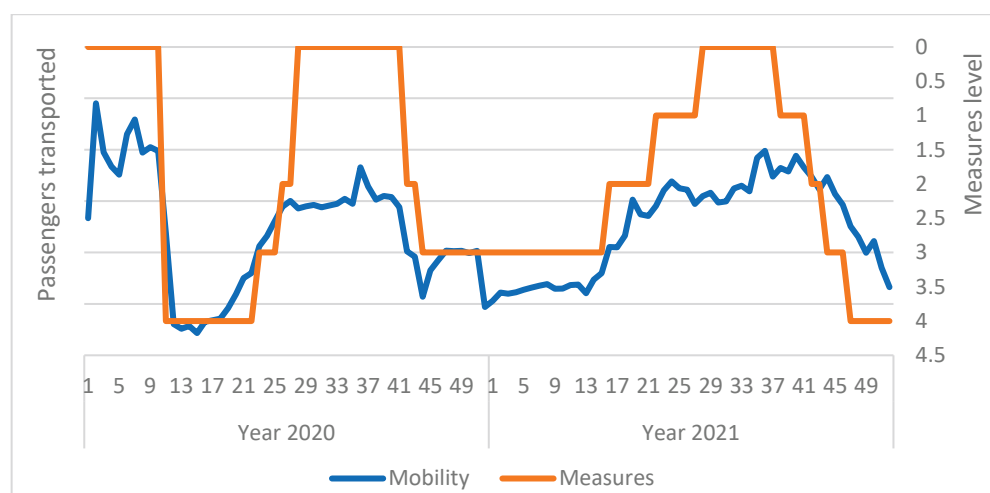


Figure 11. Comparison of rail passenger mobility and the levels of restrictions in the education sector for the Bratislava region.

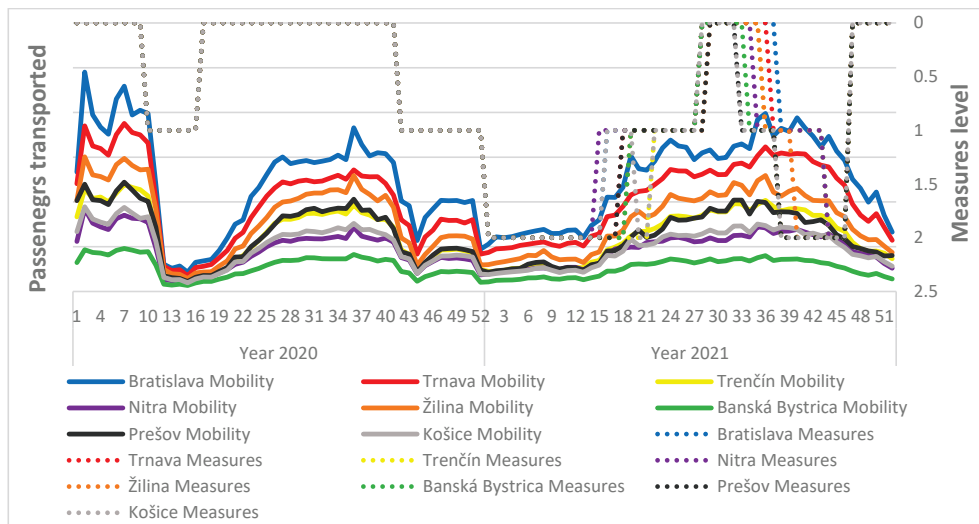


Figure 12. Comparison of rail passenger mobility according to the levels of restrictions on workplace entry.

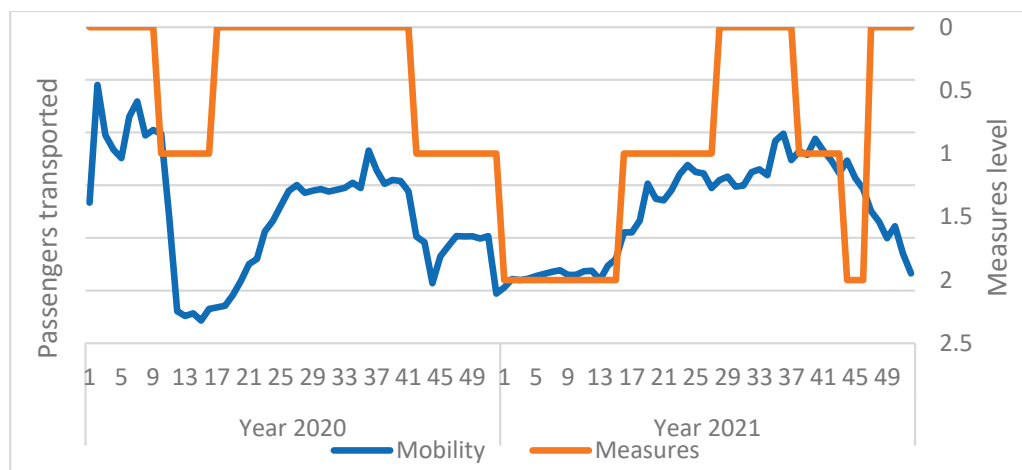


Figure 13. Comparison of rail passenger mobility and the levels of restrictions on workplace entry in the Bratislava region.

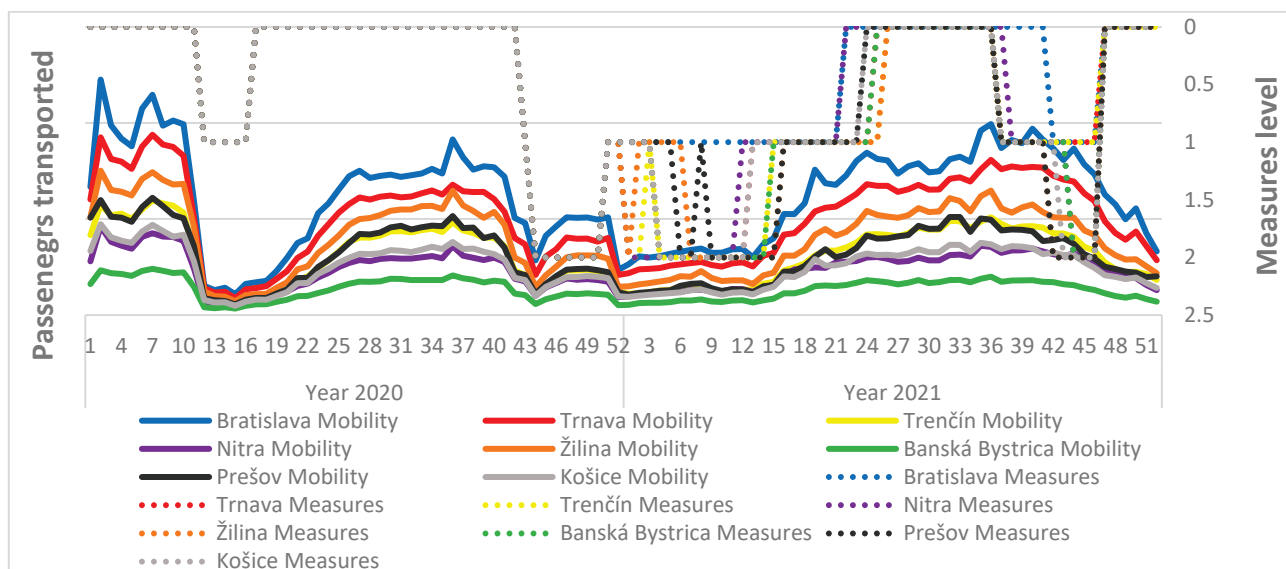


Figure 14. Comparison of rail passenger mobility according to the levels of restrictions on movement.

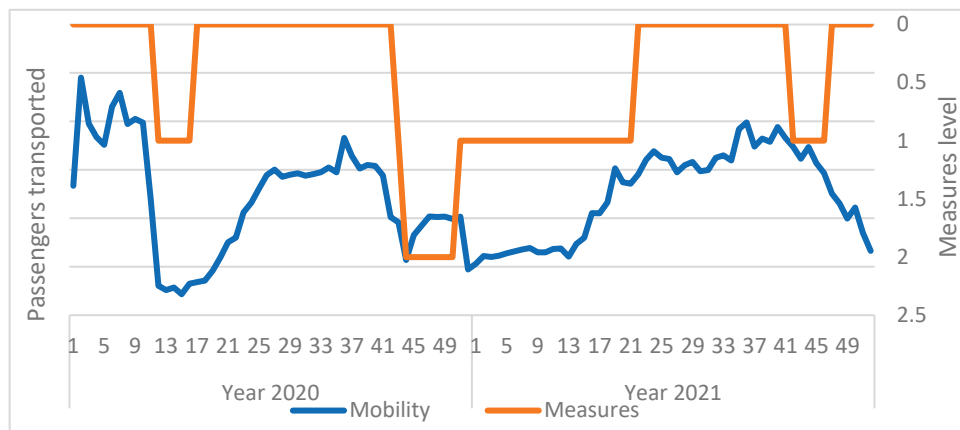


Figure 15. Comparison of rail passenger mobility and the levels of restrictions on movement in the Bratislava region.

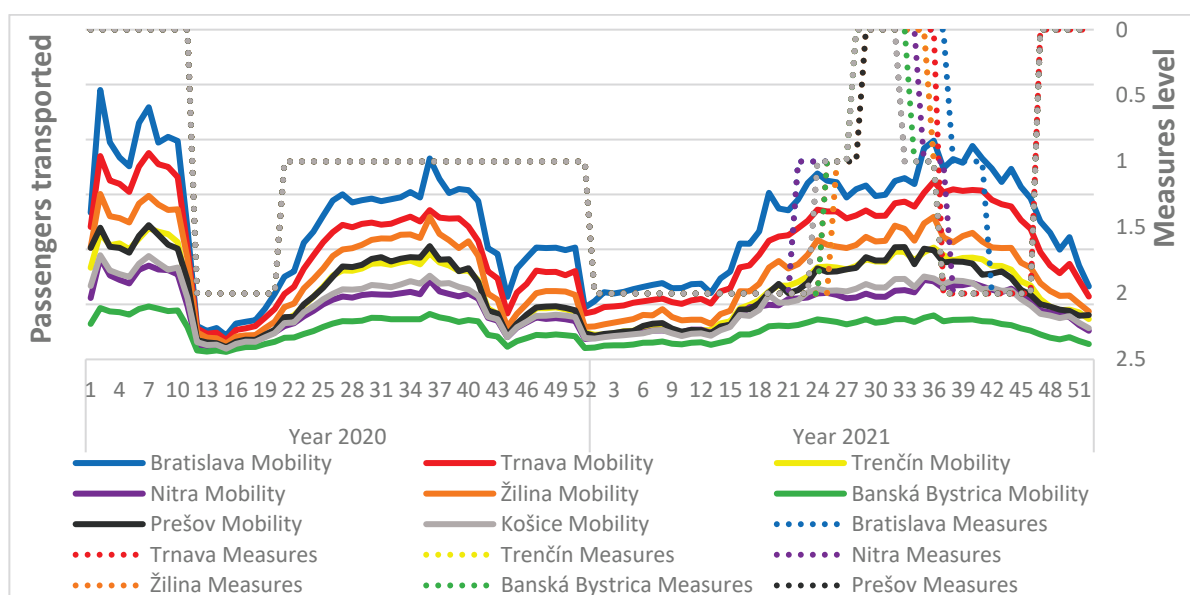


Figure 16. Comparison of rail passenger mobility according to the levels of restrictions on shops.

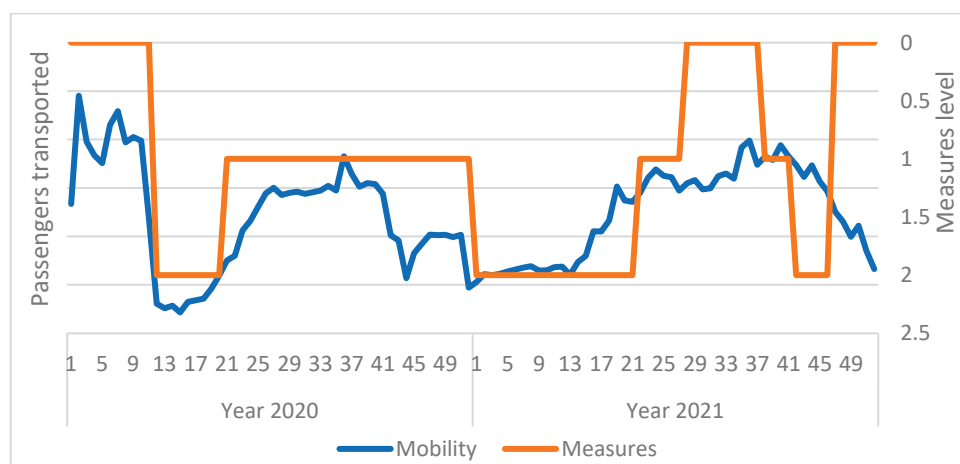


Figure 17. Comparison of rail passenger mobility and the levels of restrictions on shops in the Bratislava region.

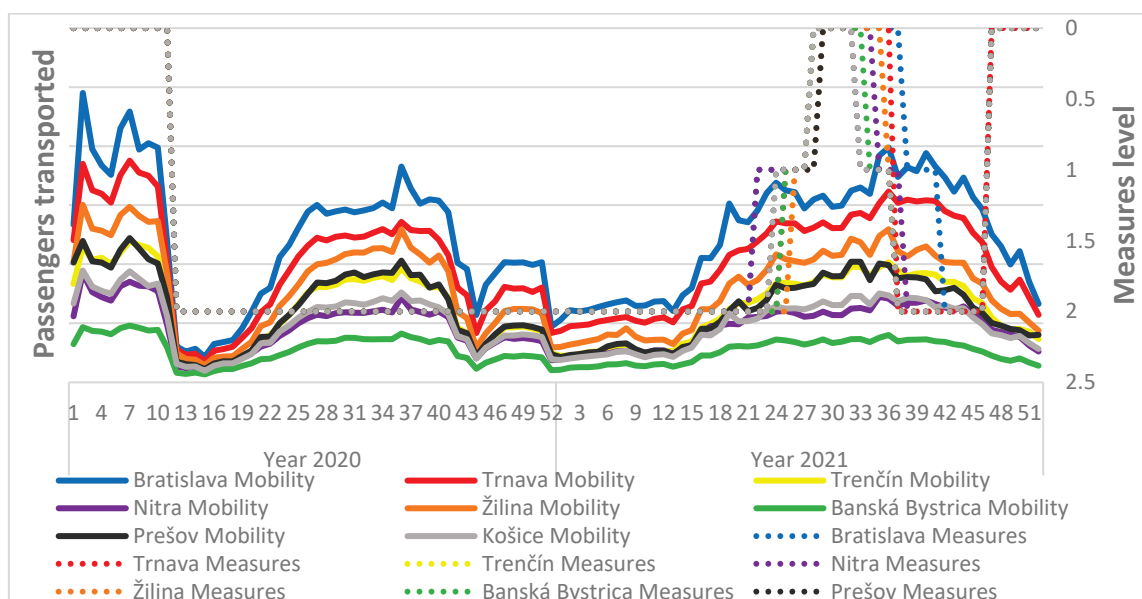


Figure 18. Comparison of rail passenger mobility according to the levels of restrictions in accommodation facilities.

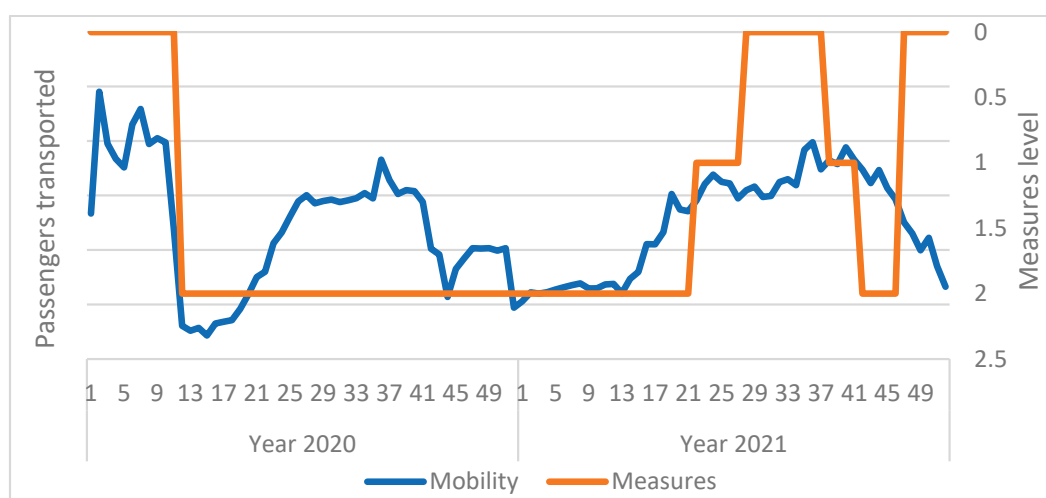


Figure 19. Comparison of rail passenger mobility and the levels of restrictions in accommodation facilities in the Bratislava region.

Restriction levels were nearly identical across regions in 2020 due to nationwide measures. The Bratislava region is shown in Figure 11 as a representative example.

Across all eight regions of Slovakia, a consistent pattern was observed between the number of passengers transported and the levels of restrictions. As the severity of measures increased, rail passenger mobility declined, and vice versa. The differences between regions were only minimal. This finding underlines the fact that, during the initial phase of the pandemic, the nationwide character of restrictions overshadowed regional specificities, which only began to emerge later with the introduction of the COVID automat. Passenger transport in railways was also strongly influenced by the daily commuting of residents to employment. Figure 12 illustrates the comparison between restrictions on entering workplaces and the number of passengers transported across all regions of Slovakia.

When comparing the development of rail passenger mobility and the levels of restrictions on employment, certain categories of measures were merged for analytical purposes. Specifically, the level of vigilance was combined with the first level of threat, and the second

and third levels of threat were also merged, since the imposed measures were identical. For example, at both the second and third levels of vigilance, home-office was mandated wherever possible, with exceptions allowed under the OTP (vaccinated, tested, recovered) regime. This reflects the gradual shift to remote work. Patterns across regions remained similar, again dominated by national policies (Figure 13).

Workplace-related restrictions showed similar nationwide patterns, with demand falling as measures tightened.

These deviations underscore the significant role of long-distance connections, which can introduce discrepancies into otherwise homogeneous regional trends. Another divergence was observed across all regions in the final weeks of 2021, primarily due to seasonal effects, such as the reduction in travel during the Christmas holidays. Figure 14 illustrates the development of rail passenger mobility and the corresponding levels of restrictions related to the limitation of movement.

Like the workplace entry measures, the levels of restrictions for movement were merged for analytical purposes. Specifically, monitoring and vigilance were combined, as well as the first and second levels of threat, since the applied measures were identical in these categories.

To provide a clearer picture of these relationships, Figure 15 presents the comparison of rail passenger mobility and the levels of movement restrictions in the Bratislava region.

In the case of movement restrictions, a trend analogous to workplace entry measures was observed. Mobility decreased with increasing severity of restrictions, and vice versa. In some regions, particularly Prešov and Košice, anomalies were again identified, caused primarily by long-distance commuting from other parts of the country. Restrictions on shops caused moderate, short-term decreases in ridership (Figures 16 and 17). Toward the end of 2021, seasonal effects such as the Christmas period partly masked policy impacts.

The development of passenger transport in relation to shop-related restrictions displayed a similar pattern across all regions. As the severity of restrictions increased, the number of passengers declined. However, an opposite trend was again recorded in the final weeks of 2021, primarily due to seasonal effects. These results confirm that rail passenger mobility is directly influenced by restrictions on economic activity and that the impact of seasonality may temporarily overshadow the effects of governmental measures.

Measures affecting accommodation produced minimal changes in rail demand (Figures 18 and 19), suggesting that recreational and leisure travel is mostly undertaken by individual transport rather than rail.

The comparison of rail passenger mobility and the restrictions introduced in accommodation facilities shows an almost identical pattern across all regions. Figures 18 and 19 indicate that, despite the introduction of stricter measures in accommodation facilities, there was no significant change in the number of passengers transported by rail. This finding suggests that travel related to recreational or sports activities is predominantly carried out by individual transport, which explains why the restrictions in accommodation facilities had only a limited impact on rail mobility.

Moreover, when examined at a higher level of detail (in our case, weekly data), the development of rail passenger transport clearly exhibits seasonal characteristics. Therefore, prior to the construction of the model, it was necessary to decompose the time series of the observed data into its individual components and to exclude the seasonal effect.

5.2. Model Performance

The outputs of the model are presented in Table 4 and Figure 20.

Table 4. Model outputs.

Calculation Process	Performance	Regression
Training	$6.557107629351653 \times 10^8$	0.9598
Validation	$5.067852971757998 \times 10^9$	0.6541
Testing	$4.393849062438726 \times 10^9$	0.7389

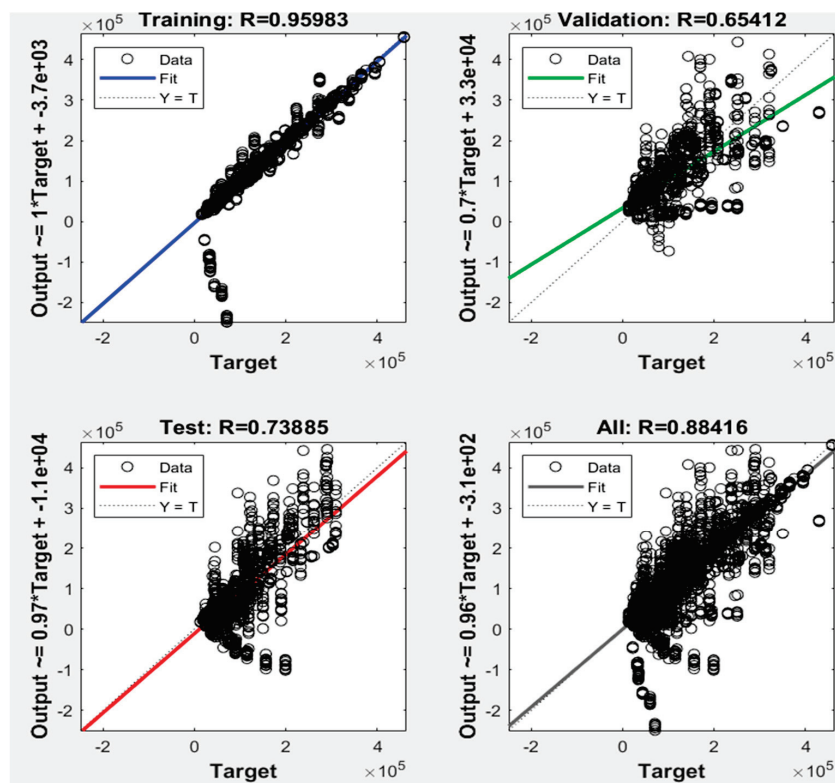


Figure 20. Regression results of the NARX model.

The results confirm the existence of regression for the training, validation, and testing phases, thereby demonstrating that the model adequately captured the relationship between the applied measures and passenger mobility. Figure 20 shows the regression results of the NARX model.

Figure 20 presents the regression results of the NARX model, including the functions and corresponding regression coefficients for training, validation, and testing. An important outcome is the overall regression coefficient ($R = 0.88416$), which confirms the dependence between the measures introduced to limit the spread of COVID-19 in the individual regions and rail passenger mobility.

The R value close to 0.96 for the training set confirms a very good model fit. The lower values for the validation set (0.65) and the testing set (0.74) indicate that, although the predictive ability of the model is satisfactory, there remains a degree of variability not fully captured. Nevertheless, the overall regression coefficient ($R = 0.88416$) demonstrates a strong correlation between the severity of restrictions and the development of passenger mobility. These results show that the proposed NARX model is capable of adequately capturing the relationships between the input variables (measures) and the output variable (mobility). The model thus provides a reliable basis for further analysis and for formulating recommendations for railway transport management during crisis situations.

The proposal of measures for rail transport is based on the constructed and validated model. The model was developed using all available data (passenger transport across all

regions and measures in all areas), while its results can also be applied individually to specific regions and particular types of measures. For this reason, this chapter is further divided into two parts: (i) the presentation of modelling results for individual regions, which illustrates regional specificities in mobility trends, and (ii) the proposal of measures for rail transport, which builds on the identified relationships and provides recommendations for practice.

5.2.1. Results of Modelling Data for Individual Regions

The analysis revealed that the main areas influencing the development of passenger transport in railways are education, entry into employment, restrictions on movement, and retail. For this reason, the model results are presented for these four areas (Figures 21–24). At the same time, the model enables the evaluation of changes in passenger transport according to changes in measures across all areas and within individual regions. Figure 21 shows the modelled time series (forecast) of changes in passenger mobility (number of passengers) in rail transport between individual weeks under the implemented measures.

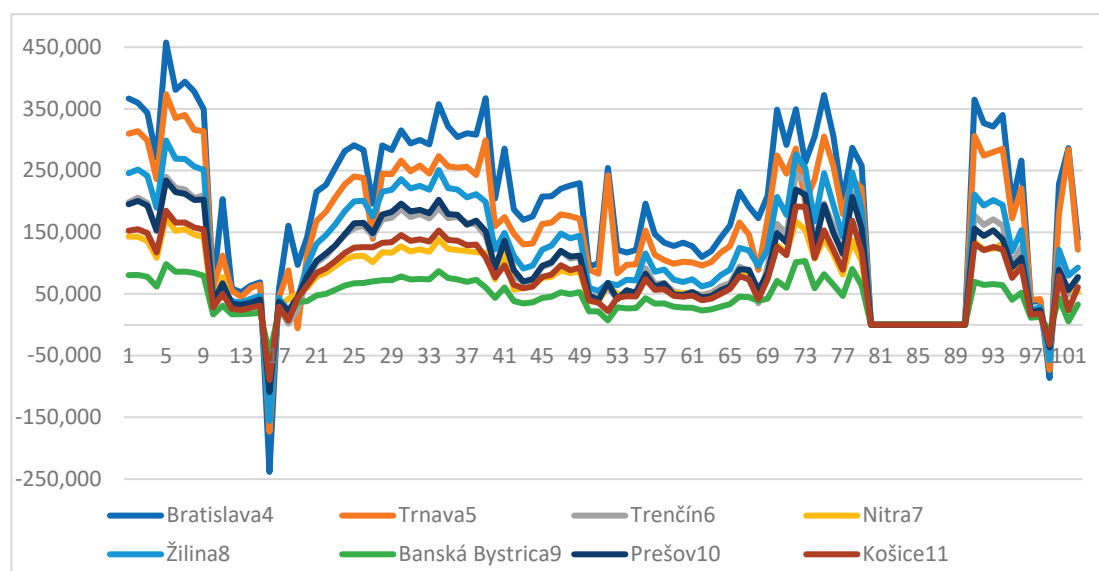


Figure 21. Model results for the measure Education sector.

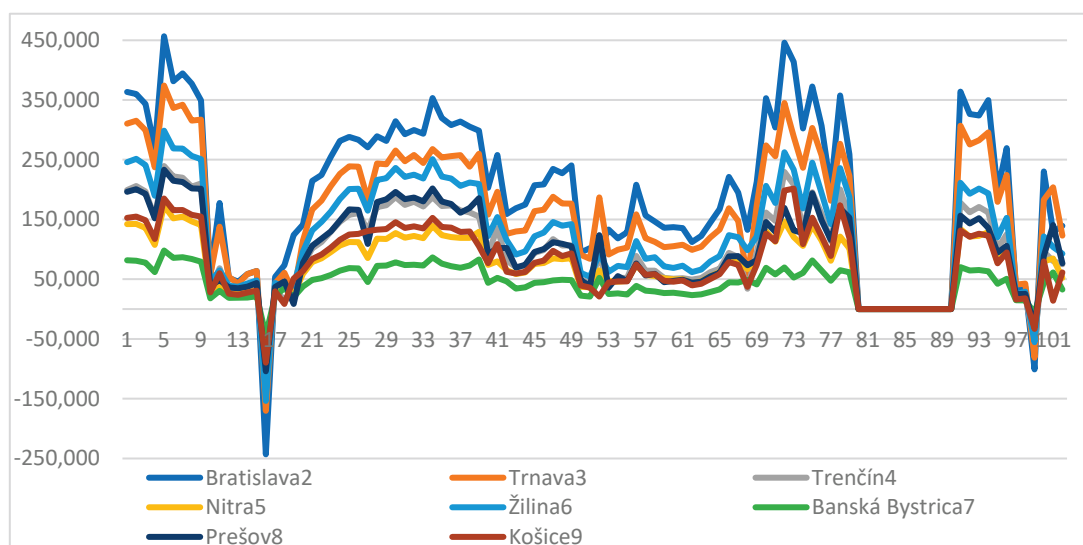


Figure 22. Model results for the measure Workplace entry.

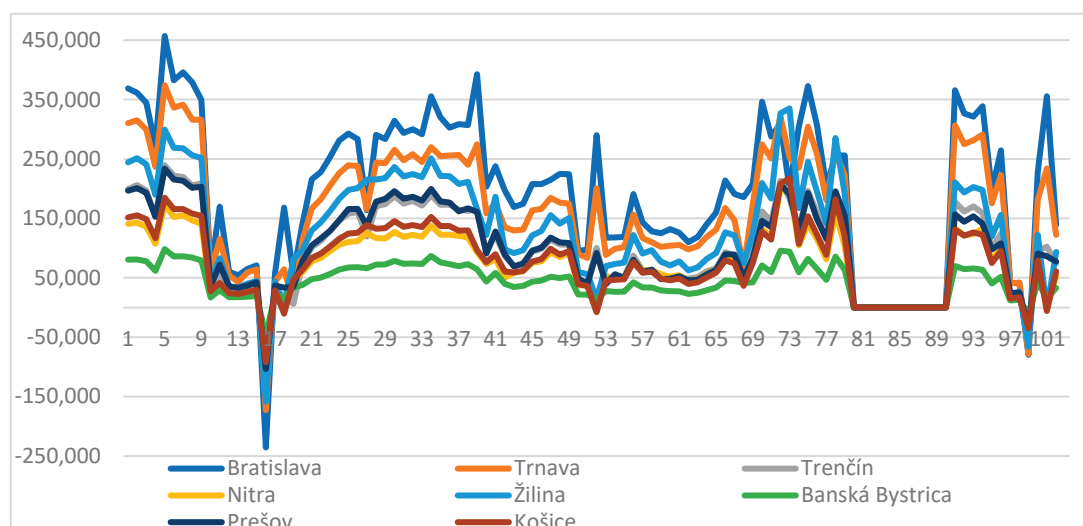


Figure 23. Model results for the measure Restriction of movement.

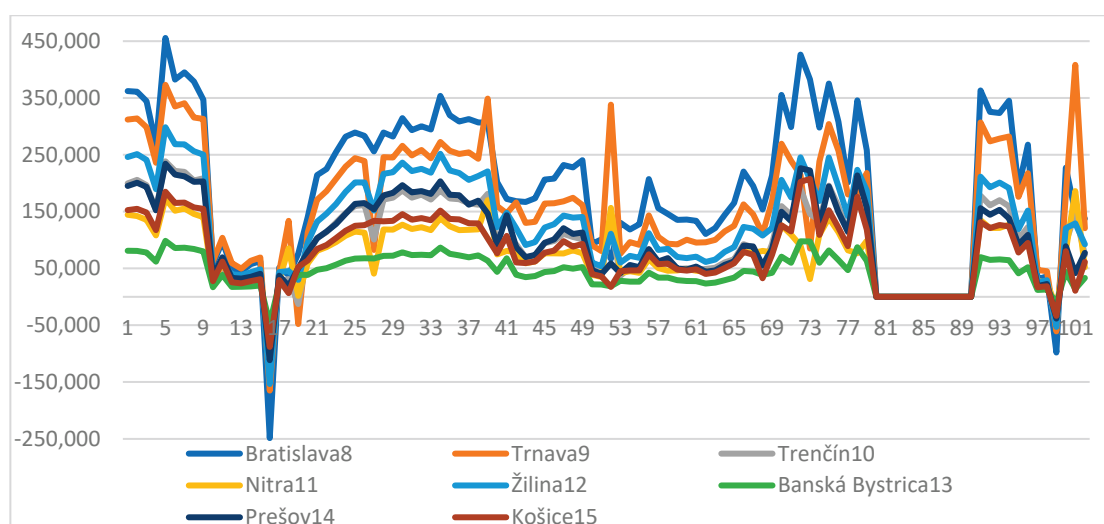


Figure 24. Model results for the measure Shops.

Due to the use of seasonally adjusted data, the forecast produced by the model covered a two-year period, using data from 2021 and 2022. By excluding the seasonal component, regularly recurring fluctuations (for example holidays, summer vacations) were eliminated, which increased the accuracy of the prediction. The modelled data across individual regions show a very similar development; the same applies to the individual measures, where deviations between the graphs are only minimal. Although the model was designed to allow region-level analysis, the empirical results confirm that regional differences in Slovakia during the pandemic were relatively minor. This can be attributed to the predominantly nationwide character of government restrictions in 2020 and the late introduction of regionally differentiated measures in 2021. Additionally, the strong centralization of the Slovak rail network along the Bratislava–Košice corridor reduces regional heterogeneity in travel patterns. Nevertheless, maintaining a regionally sensitive modelling framework is valuable, as it enables scenario testing and capacity planning for potential future events with stronger local impacts (district-level outbreaks or natural disruptions). Figure 22 illustrates the modelled time series (forecast) of changes in passenger mobility (number of passengers) in rail transport for the measure Entry into employment.

Figure 23 presents the modelled time series (forecast) of changes in passenger mobility (number of passengers) in rail transport for the measure Restriction of movement.

The modelled time series (forecast) of changes in passenger mobility (number of passengers) in rail transport for the measure Shops is shown in Figure 24.

When comparing these predicted (modelled) data with the actual development, it was confirmed that the model correctly captures the fundamental dynamics of mobility changes under the introduction of different levels of measures. This demonstrates its ability to reflect the relationships between restrictive interventions and passenger behaviour.

When comparing the predicted (modelled) data for a given region and measure using chain indices between weeks with different levels of restrictions, the correlation was also confirmed. Figure 25 shows the development of modelled data for education measures in the Bratislava region.



Figure 25. Model results for the measure Education sector in the Bratislava region.

Figure 26 presents the development of modelled data for the measure Workplace entry in the Bratislava region.

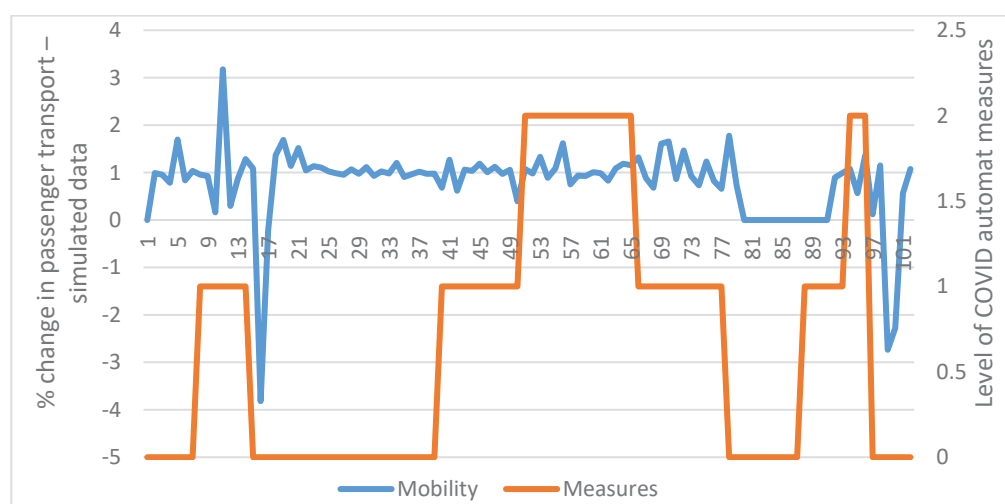


Figure 26. Model results for the measure Workplace entry in the Bratislava region.

The development of modelled data for the measure Restriction of movement in the Bratislava region is shown in Figure 27.

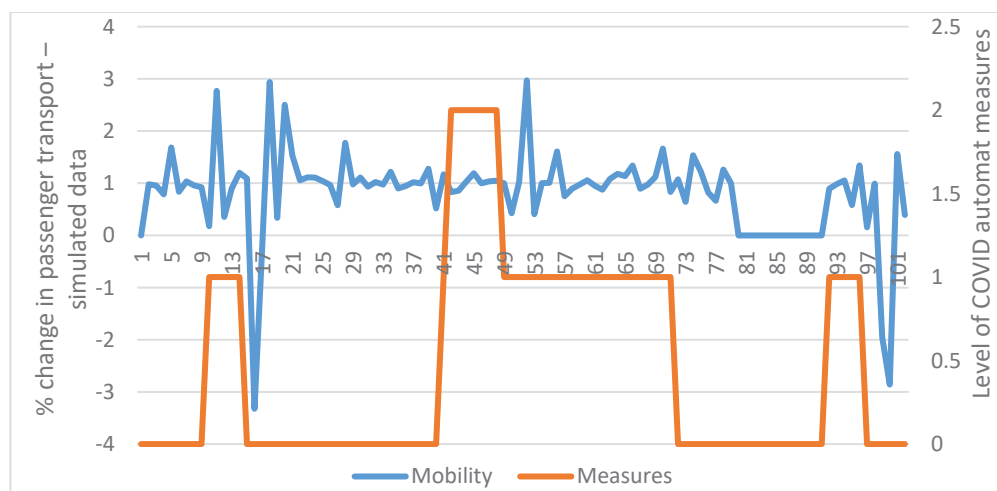


Figure 27. Model results for the measure Restriction of movement in the Bratislava region.

The last modelled measure in the Bratislava region was Shops, which is illustrated in Figure 28.

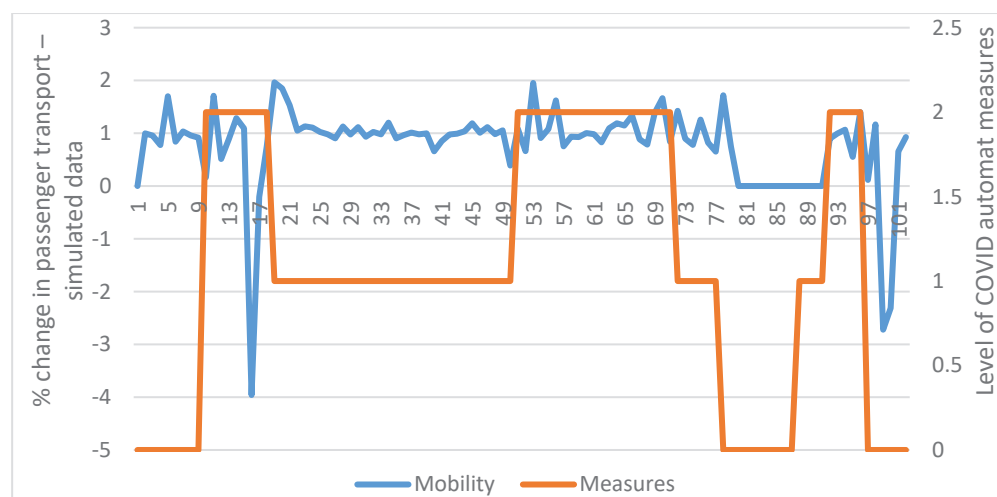


Figure 28. Model results for the measure Shops in the Bratislava region.

The modelled and predicted data are characterized by several extreme values resulting from the chosen method and the training of the neural network. The main cause of these extremes lies in the markedly different measures applied in individual areas at a given time. A neural network-based model cannot selectively isolate such phenomena, and at the same time none of them can be excluded during model creation (network training), since they act simultaneously on the development of rail passenger mobility. These extreme values should therefore be interpreted with caution—they do not represent an error of the model, but a natural consequence of its attempt to reconcile the conflicting effects of multiple measures implemented at the same time.

The prediction of passenger rail mobility based on the established model is possible both for Slovakia as a whole and for individual regions or specific measures. In this way, the model becomes a universal tool that can be applied during crisis situations of both national and regional scope. By entering the relevant measures, it is possible to simulate their expected impact on the demand for rail passenger transport.

5.2.2. Proposal for Measures for Rail Transport

Since the model was built using a binary variable, when proposing concrete measures (reduction of realized performance in train kilometres), we carried out a comparison of the number of passengers transported before and during the COVID-19 pandemic. Passenger transport in individual weeks of 2019 was taken as the baseline for comparison with 2020 and 2021. Such a comparison is justified, as the developed and validated model confirmed the existence of a correlation between the levels of restrictions in different areas and the number of passengers transported in rail transport.

Schools

Measures introduced to prevent the spread of coronavirus in the education sector were among the strictest and had a significant impact on changes in mobility in rail passenger transport. Figure 29 shows the comparison of the level of restrictions in education and the change in mobility in the Bratislava region.

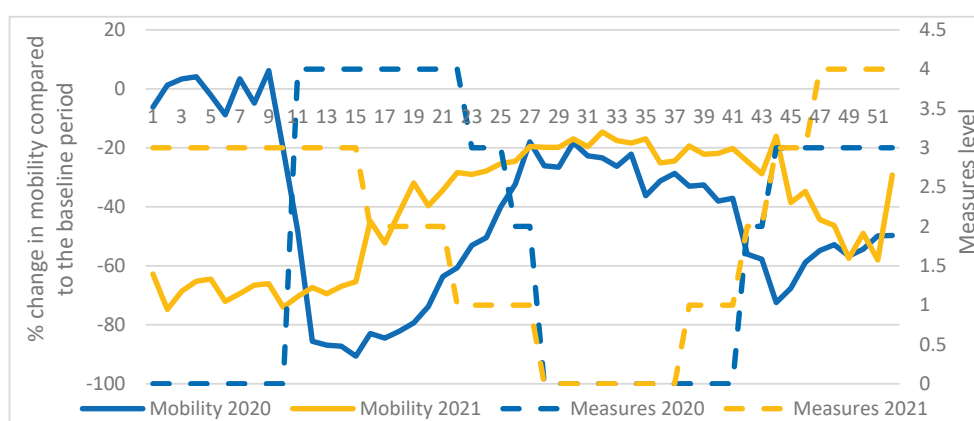


Figure 29. Development of measures in education sector and percentage change in rail passenger transport in the Bratislava region.

The change in the number of passengers transported in relation to the introduction of individual levels of measures in education followed a similar pattern across all eight regions. When the strictest measure, i.e., level 4, was introduced—when in-person teaching was replaced by distance learning—the average decrease in passenger transport reached 75%. At level 3 (in-person teaching with a limited number of pupils/students depending on vaccination status or compliance with the OTP regime—vaccinated, tested, recovered—or according to the “school COVID automat”), rail passenger mobility decreased on average by 65%. The second level of restrictions in education (increased in-person teaching capacity under the OTP regime) resulted in an average reduction of 40%. Under level 1 (at universities: 400 people for OTP, 1000 for fully vaccinated, while other schools operated in-person according to the school COVID automat), transport performance decreased on average by approximately 20%. These results confirm that education is among the key areas with the strongest impact on rail mobility, primarily due to the daily commuting of pupils and students.

Entry into Employment

Measures in all other examined areas had a roughly comparable impact on changes in rail passenger transport. Figure 30 illustrates the comparison of passenger transport changes in relation to the level of restrictions introduced for entry into employment in the Bratislava region.

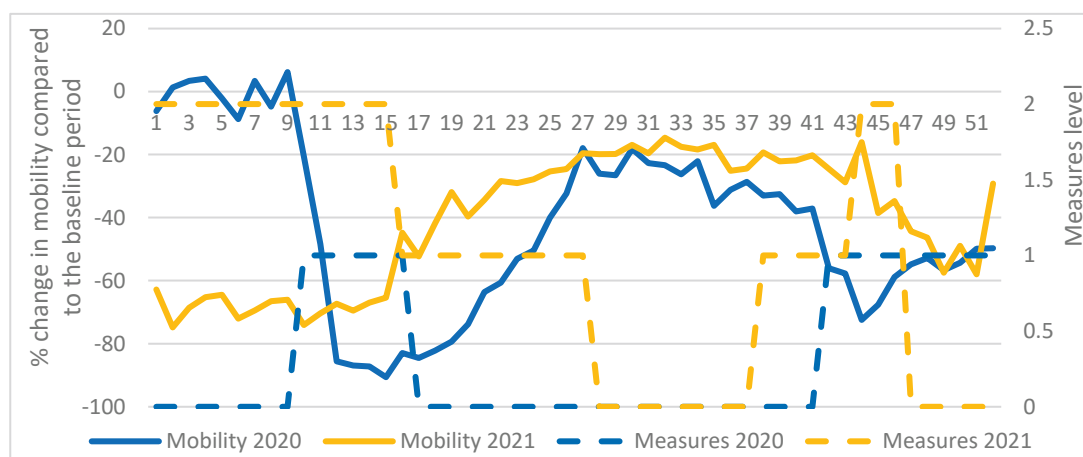


Figure 30. Development of measures related to entry into employment and the percentage change in rail passenger transport in the Bratislava region.

Restriction of Movement

Figure 31 presents a comparison of passenger transport changes in relation to the level of restrictions introduced for the limitation of movement in the Bratislava region.

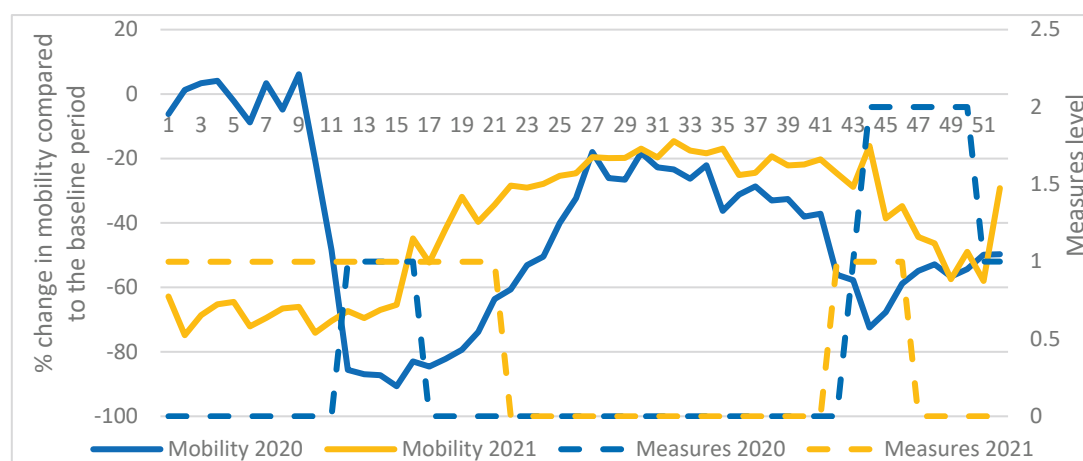


Figure 31. Development of measures related to the restriction of movement and the percentage change in rail passenger transport in the Bratislava region.

Measures associated with the restriction of movement were among those that had an immediate and direct impact on population mobility. Under the strictest levels of restrictions (curfew with exceptions, limitations on inter-district travel), the average decrease in the number of rail passengers reached approximately 60–70%. At lower levels of restrictions, when travel to work or school was still permitted under certain conditions, the decline in mobility was less pronounced—around 25–35%. It is important to emphasize that these “intermediate” measures reveal the capacity of rail transport to retain part of its ridership even during a crisis, particularly when no suitable alternative in the form of individual transport is available.

Shops

Figure 32 shows the comparison of changes in rail passenger transport with respect to the level of restrictions imposed on shops in the Bratislava Region. The analysis indicates that limiting the number of customers per unit area in retail establishments had a measurable impact on mobility, particularly during periods of stricter measures.

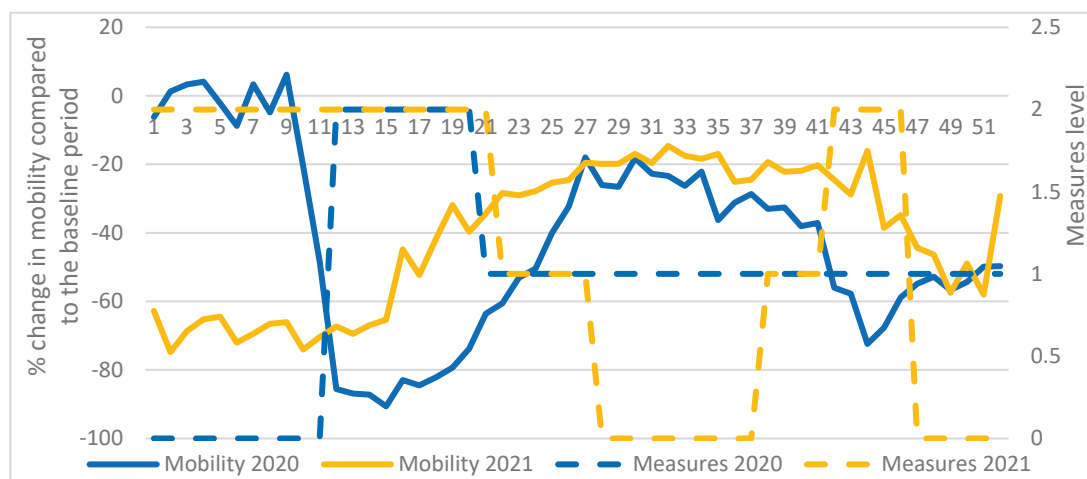


Figure 32. Development of measures in shops and the percentage change in rail passenger transport in the Bratislava Region.

The development of changes in rail passenger transport was influenced by the severity of individual measures, with notable differences observed between periods of nationwide restrictions in 2020 and the COVID automat system introduced in 2021. This distinction reflects the changing policy framework associated with the availability of COVID-19 vaccines and the corresponding adjustments in governmental responses. The results therefore demonstrate that identical measures could have had different effects in various periods due to the evolving social and epidemiological context. In general, strict measures (a limit of one customer per 1 m² of retail space, or restrictions on movement with only a few exceptions) led to an average decrease in rail passenger transport of approximately 70%. By contrast, less stringent measures (recommended but not mandatory home office arrangements) were associated with an average reduction of around 30%.

Based on the constructed and validated neural network model in MATLAB, the following conclusions can be drawn:

- Passenger rail transport was significantly influenced by measures implemented in the areas of education, workplace entry, restrictions on movement, and limits on the number of customers in shops. Other measures did not have a substantial impact on changes in passenger numbers.
- Although measures in 2021 were introduced at the regional level according to the evolving epidemiological situation, the same level of restrictions generally led to comparable percentage changes in the number of transported passengers across different regions.
- A comparison of transport capacity and passenger demand in rail transport revealed considerable over-dimensioning of rolling stock capacity under the strictest measures, particularly those related to education and workplace entry.

Based on these findings, we propose the following reductions in rail transport capacity (achieved either by reducing the number of coaches per train or by decreasing total train kilometres):

- Under the strictest measures—a reduction of capacity by 40%;
- Under medium-level measures—a reduction of capacity by 30%;
- Under mild measures—a reduction of capacity by 10%.

These recommendations are derived from the validated model and the analysis of passenger transport dynamics in relation to the introduction of specific measures. Importantly, stricter measures required higher capacity per train due to mandated physical distancing between passengers.

The application of demand-modelling approaches in the context of pandemics or other crisis situations can lead to significant reductions in social costs in passenger rail transport, as well as improvements in the management of technological processes and human resources. The implementation of these recommendations could therefore contribute to more efficient use of available resources, cost savings at the societal level, and more flexible management of railway operations during crisis conditions.

6. Discussion

The analysis and modelling results confirm that the most decisive factors influencing rail passenger mobility during the COVID-19 pandemic were measures related to schooling, access to workplaces, restrictions on movement, and retail activities. Among these, the closure of schools proved to be the strongest driver of mobility reduction, with an average decline of up to 75%. This underscores the crucial role of daily commuting of pupils and students in shaping passenger flows in the Slovak rail network.

The application of the NARX neural network model demonstrated its suitability for capturing nonlinear relationships between pandemic measures and passenger mobility. The model achieved a high overall correlation ($R = 0.88$) between the imposed measures and the number of passengers and provided a robust basis for simulating alternative scenarios. This makes it a valuable tool for transport authorities, enabling them to design adaptive strategies and implement flexible capacity planning during crises.

Despite its robustness, the model also has several limitations. Extreme predicted values were observed in certain cases, which can be attributed to the simultaneous implementation of multiple measures and the neural network's limited ability to disentangle their individual effects. Moreover, the analysis relied on aggregated regional-level data, which may obscure differences in mobility behaviour across socio-economic groups or trip purposes. Finally, the study focused in detail on the Bratislava region, which was chosen due to its role as the most significant commuting hub in Slovakia, with a high concentration of both regional and long-distance flows. While this focus allows for an in-depth analysis of the area most affected by restrictions, it also limits the generalizability of the findings to other regions with different demographic or infrastructural profiles.

When compared to international findings, the results are consistent with studies from neighbouring countries. It is important to note that, although the study set out to emphasize a regional perspective, the observed differences between Slovak regions proved modest. This outcome reflects both the uniform nationwide response during most of the pandemic and the high interconnectedness of the country's rail system. However, the modelling framework remains regionally applicable and can support targeted decision-making should future crises require geographically differentiated interventions. Research in the Czech Republic [44] using regression-based approaches confirmed that school closures were the strongest determinant of passenger demand declines, with reductions exceeding 70%. Similar outcomes were reported in Italy, where passenger numbers dropped by more than 70% during lockdowns, while studies applied discrete choice models and agent-based simulations to capture behavioural shifts [30,45–47]. These findings suggest that the Slovak results are in line with broader European experiences.

While the proposed NARX model achieved a high correlation between restrictions and passenger demand, several limitations and potential sources of bias should be acknowledged. First, the binary representation of tightening or relaxing measures cannot fully capture the heterogeneity of real restrictions, for example partial school closures or mixed home/office work regimes may have different impacts than assumed. Second, the analysis is based on aggregated weekly and regional data, which may mask short-term behavioural shifts and differences among socio-economic groups. Third, long-distance

commuting patterns, especially from Eastern Slovakia to Bratislava, may distort regional comparability and partly explain the anomalies observed in Prešov and Košice. Finally, the model's predictive ability decreases when multiple extreme restrictions are imposed simultaneously, as the neural network cannot always disentangle interacting effects. These aspects should be considered when interpreting the forecasts and applying the results to operational planning.

In addition, although the model suggests capacity reductions of up to 40%, practical implementation may be constrained by the availability of rolling stock and crew, contractual obligations with service providers, and the need to maintain minimum connectivity for remote regions.

Another important limitation lies in the binary coding of government measures. The model treats any tightening of restrictions as a single-step change, regardless of whether it represents a minor intervention (for example mandatory masks) or a major disruption (for example full lockdown). This simplification can reduce the model's ability to distinguish between moderate and severe policy impacts and may partly explain extreme predicted values when multiple measures were implemented simultaneously.

Another methodological limitation is the lack of a formal benchmark comparison with simpler time-series models such as ARIMA or regression. Although these approaches were considered and dismissed due to pronounced non-stationarity and nonlinear demand shocks, their explicit inclusion as baseline models would make the argument for NARX's superiority more robust. The notable drop in predictive accuracy from the training set ($R \approx 0.96$) to the validation and testing sets ($R \approx 0.65$ – 0.74) also indicates a potential risk of overfitting that should be examined more systematically.

A limitation of this study is the lack of extended validation procedures beyond R and MSE. The model has not yet been tested with cross-validation, sensitivity analysis, or compared with simpler baselines such as ARIMA or regression. Incorporating these robustness checks would help confirm the necessity and superiority of the NARX approach. A further limitation of the current validation is its reliance solely on standard goodness-of-fit statistics (R and MSE). Without scenario-based or sensitivity analyses, it remains unclear how stable the model predictions are under moderate perturbations of key input variables. For instance, it is unknown whether small changes in the stringency or timing of workplace restrictions produce proportionate effects on predicted ridership or cause abrupt, unrealistic swings. Incorporating such analyses would clarify model behaviour and strengthen its practical relevance.

Since the acute pandemic period, passenger rail transport in the Slovak Republic has shown gradual but incomplete recovery. Data from the national rail operator ZSSK indicates that ridership in 2022 and 2023 increased compared to 2021, as restrictions were lifted and students as well as employees partly returned to daily commuting. However, demand has not yet fully reached pre-2019 levels. A lasting change is the higher prevalence of hybrid and remote work, which continues to reduce peak weekday flows, while leisure and weekend travel rebounded faster. The pandemic accelerated several modernization steps in Slovak passenger rail, digital ticket sales and reservations became the dominant channel, real-time passenger counting systems are being expanded, and timetable adjustments are more dynamic than before. ZSSK also tested flexible train compositions and partial capacity reductions on low-demand routes. Nevertheless, the vulnerability of the system to sudden demand shocks remains, and there is still no fully institutionalized mechanism for rapid scaling of capacity during crises. These developments confirm the practical relevance of predictive approaches such as the NARX model, which can support more adaptive and resilient railway operations in the long term.

Taken together, the visual trends (Figures 9–18) highlight a strong policy–mobility coupling: education and workplace restrictions exert primary control over national passenger flows, while other measures contribute only secondary, short-lived effects. This implies that operators can anticipate major demand shocks by closely monitoring educational and work-related policy changes. It also underlines the resilience of leisure and shopping trips, which tend to recover quickly once restrictions ease, offering opportunities for targeted marketing and dynamic pricing.

Beyond Europe, mobility reductions were observed worldwide. The study [48] showed that a 20–40% decrease in mobility can reduce the effective reproduction rate of COVID-19 below 1, particularly when related to commuting, transport, and retail activities. In the United Kingdom, car use decreased by around 30% over 2020 compared to pre-pandemic levels, with a 50% reduction during the first lockdown. Rail passenger numbers fell by 76% between March and December 2020, and public bus ridership declined to only 15% of the previous year's level [49]. Comparable decreases were documented in Germany [50], Sweden [51,52], Switzerland [53], the Netherlands [54], Indonesia [55], Colombia [56], China [57–59], Australia [60] and Japan [61]. Compared to these international studies, the Slovak case highlights the added value of using neural networks, which are capable of capturing nonlinearities often overlooked by traditional econometric techniques.

The findings also have several implications for railway policy and operations. Adjusting capacity according to the severity of restrictions—for example, reducing train kilometres by 40% during the strictest measures—could lead to significant cost savings and more efficient use of resources. Such adaptive planning is not only economically relevant but also environmentally beneficial, as it reduces unnecessary energy consumption and associated CO₂ emissions during periods of reduced demand. At the same time, capacity reductions must be carefully balanced to maintain sufficient space for safe distancing during health crises, which highlights the need for flexible train composition and dynamic crew management. The recommended capacity reduction of 10–40% is directly based on the demand drops predicted by the NARX model under different levels of restrictive measures. Specifically, the model scenarios that included the most severe nationwide restrictions (e.g., full school closures, mobility bans) showed passenger reductions approaching 40%, while moderate interventions such as partial workplace limitations or mask mandates resulted in approximately 10–20% demand loss. These forecasts provided the quantitative basis for translating predicted ridership decreases into practical capacity adjustment recommendations, rather than relying solely on descriptive historical trends.

By employing a neural network–based simulation framework, the study contributes to the field of data-driven modelling and simulation of transportation systems. This aligns with the scope of the Special Issue, demonstrating how machine learning methods can capture complex dynamics under crisis conditions and provide actionable insights for policymaking.

Finally, the applicability of the model extends beyond pandemic situations. The NARX-based framework can be adapted to forecast mobility under other crisis conditions, such as extreme weather events, energy shortages, or regional epidemics. In such contexts, the ability to quickly simulate different restriction scenarios and their impact on passenger demand can significantly support decision-making and improve the resilience of transport systems.

Future research should focus on integrating additional data sources, such as real-time ticketing systems or automated passenger counts, to refine predictive accuracy. Extending this approach to other modes of transport would provide a more comprehensive understanding of multimodal mobility dynamics. Combining neural network models with

agent-based simulations could further capture behavioural heterogeneity at the individual level, thereby enhancing the robustness and policy relevance of crisis mobility models.

7. Conclusions

This study analysed the demand for rail passenger transport in the Slovak Republic during the COVID-19 pandemic, with a particular focus on the Bratislava region. The results confirmed that the most decisive factors influencing mobility were measures in schooling, access to workplaces, restrictions on movement, and retail activities. Among these, school closures had the strongest impact, causing an average reduction of up to 75% in passenger numbers. The application of the NARX neural network model demonstrated a strong correlation ($R = 0.88$) between the severity of restrictions and the decline in passenger demand, thereby providing a reliable framework for forecasting mobility during crisis situations. Unlike most previous studies that analysed pandemic-related impacts on rail passenger demand at a national or highly aggregated level, this work applies a nonlinear NARX neural network to the Slovak rail system and explicitly integrates government restrictions as exogenous variables. This combination provides a regionally applicable yet nationally interpretable modelling framework that, to our knowledge, has not been presented in the existing literature.

The findings highlight important implications for railway policy and operations. Adaptive capacity management, including reductions of 10–40% depending on the severity of restrictions, could improve the efficiency of resource allocation and reduce social costs, while maintaining safe travel conditions. Although the analysis focused on the Bratislava region, similar mobility patterns were observed across other Slovak regions, with differences mainly reflecting regional variations in infection intensity and the COVID automat. The proposed capacity adjustments (10–40%) reflect the model's scenario outputs: the NARX predictions under varying policy intensities were translated into actionable ranges for railway planners. This strengthens the practical link between the modelling results and operational recommendations.

Nevertheless, the study has certain limitations. First, the model was built on binomial variables representing the tightening or relaxation of measures, which cannot fully capture the heterogeneity and complexity of real-world restrictions. Second, the analysis relied on aggregated weekly data, which may conceal short-term fluctuations in mobility. Third, the regional focus on Bratislava, while justified by its role as the country's most important commuting hub, constrains the generalizability of the findings to regions with different mobility structures.

Future research should address these limitations by integrating higher-resolution data (daily or real-time passenger counts), incorporating more nuanced representations of measures, and extending the analysis to additional regions and transport modes. Such extensions would provide a more comprehensive understanding of mobility dynamics under crisis conditions and further strengthen the capacity of transport authorities to design resilient and adaptive transport policies.

Despite the robust performance of the proposed NARX model, several limitations should be acknowledged. The use of binary indicators for tightening or relaxing measures cannot fully represent the complexity and heterogeneity of real government interventions. Weekly and regionally aggregated data may conceal short-term fluctuations and differences in passenger segments, such as commuters versus occasional travellers. Moreover, the model was trained on data from 2019–2021 and primarily validated on the Bratislava region, which may limit its direct transferability to areas with different mobility structures.

Future research should focus on incorporating higher-resolution and real-time passenger data, such as automated counts and smart-ticketing systems, to refine predictive

accuracy. Extending the model to multimodal mobility (bus, metro, long-distance coach) could provide a more holistic view of crisis-induced travel changes. Combining neural networks with agent-based or discrete choice models may also better capture behavioural heterogeneity. Finally, evaluating the model's performance on new disruptions beyond pandemics (for example, energy price shocks or extreme weather events) could further strengthen its resilience and practical applicability. Future work should replace the current binary representation of policy measures with a more granular or weighted coding scheme that reflects the true intensity of interventions. This refinement could help the NARX model better differentiate between mild and severe restrictions and thus improve forecasting accuracy for various crisis scenarios. Future research should incorporate explicit benchmarking of the NARX neural network against simpler baseline models, such as ARIMA or regression with exogenous inputs, to confirm its advantage and mitigate concerns about overfitting. Such comparative analysis would provide stronger evidence for selecting neural network methods in crisis-related mobility modelling. Future work should include scenario-based testing and sensitivity analysis to complement fit statistics. By systematically altering the intensity and sequence of governmental restrictions or other exogenous variables, researchers could evaluate the robustness of the NARX predictions and provide decision-makers with a clearer understanding of confidence intervals and risk margins.

The presented model is limited to using a simplified coding of restriction severity and by reliance on historical data from a single crisis period. Benchmarking against simpler time-series approaches (for example, ARIMA) and sensitivity or scenario analysis would strengthen robustness and generalizability. Future research should refine the coding of interventions, test the model on multimodal transport systems, and explore dynamic capacity allocation strategies for different crisis scenarios.

Based on the NARX demand forecasts under different levels of restrictive measures, railway operators can prepare tiered capacity strategies for crisis situations. When education closures or severe workplace restrictions are introduced, a capacity reduction of 30–40% is justified, while moderate measures may require only a 10–20% adjustment. Essential commuter services, especially on main interregional corridors and to/from Bratislava, should be prioritized to maintain connectivity for key workers. Operators could pre-design flexible emergency timetables that reduce train frequency but preserve core routes and use demand-monitoring dashboards to dynamically add or remove services as restrictions change. Coordination with regional authorities is crucial to ensure that school and work policies inform capacity decisions in real time.

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

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This publication was created thanks to support under the Operational Program Integrated Infrastructure for the Identification and possibilities of implementation of new technological measures in transport to achieve safe mobility during a pandemic caused by COVID-19 project (ITMS code: 313011AUX5), co-financed by the European Regional Development Fund. The paper was also supported by the KEGA Agency by Project 010ŽU-4/2023, “Innovative approaches in teaching

in the field of transport focused on railway traffic management, with the support of risk and crisis management”, led by the Faculty of Operations and Economics of Transport and Communication, University of Žilina.

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

References

1. Gkiotsalitis, K.; Cats, O. Public transport planning adaptation under the COVID-19 pandemic crisis: Literature review of research needs and directions. *Transp. Rev.* **2020**, *41*, 374–392. [CrossRef]
2. Barbieri, D.M.; Lou, B.; Passavanti, M.; Hui, C.; Hoff, I.; Lessa, D.A.; Sikka, G.; Chang, K.; Gupta, A.; Fang, K.; et al. Impact of COVID-19 pandemic on mobility in ten countries and associated perceived risk for all transport modes. *PLoS ONE* **2021**, *16*, e0245886. [CrossRef] [PubMed]
3. Basu, R.; Ferreira, J. Sustainable mobility in auto-dominated Metro Boston: Challenges and opportunities post-COVID-19. *Transp. Policy* **2021**, *103*, 197–210. [CrossRef]
4. Basunia, A.; Muttaqi, A.; Bhuiyan, M.R.H.; Badhon, F.A. Transportation mobility during COVID-19: A systematic review and bibliometric analysis. *Digit. Transp. Saf.* **2024**, *3*, 169–183. [CrossRef]
5. Borkowski, P.; Jazdzewska-Gutta, M.; Szmelter-Jarosz, A. Lockdowned: Everyday mobility changes in response to COVID-19. *J. Transp. Geogr.* **2021**, *90*, 102906. [CrossRef]
6. Černá, L.; Pribula, D.; Bulková, Z.; Abramović, B. Draft of public rail passenger transport during the COVID-19 pandemic. *LOGI—Sci. J. Transp. Logist.* **2023**, *14*, 77–88. [CrossRef]
7. Hadjidemetriou, G.M.; Sasidharan, M.; Kouyialis, G.; Parlikad, A.K. The impact of government measures and human mobility trend on COVID-19 related deaths in the UK. *Transp. Res. Interdiscip. Perspect.* **2020**, *6*, 100167. [CrossRef]
8. Dudziak, A.; Caban, J. The urban transport strategy on the example of the city bike system in the city of Lublin in relation to the COVID-19 pandemic. *LOGI—Sci. J. Transp. Logist.* **2022**, *13*, 1–12. [CrossRef]
9. Harantová, V.; Ilavská, I.; Böhm, P.; Böhmová, G. The impact of the pandemic on the daily mobility of residents. *Transp. Tech. Technol.* **2023**, *19*, 13–18. [CrossRef]
10. Musselwhite, C.; Avineri, E.; Susilo, Y. Restrictions on mobility due to the coronavirus COVID-19: Threats and opportunities for transport and health. *J. Transp. Health* **2021**, *20*, 101042. [CrossRef]
11. De Sousa, L.E.; Neto, P.H.D.O.; Da Silva Filho, D.A. Kinetic Monte Carlo model for the COVID-19 epidemic: Impact of mobility restriction on a COVID-19 outbreak. *Phys. Rev. E* **2020**, *102*, 032133. [CrossRef]
12. Štefancová, V.; Čulík, K.; Abramović, B.; Pálková, A. Restriction of mobility due to follow-up measures caused by COVID-19. In *TRANSBALTICA XIII: Transportation Science and Technology; Lecture Notes in Intelligent Transportation and Infrastructure*; Prentkovskis, O., Yatskiv, I., Skačkauskas, P., Maruschak, P., Karpenko, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2023; pp. 627–633. [CrossRef]
13. Azevedo, J.P.; Cojocar, A.; Talledo, V.M.; Narayan, A. COVID-19 School Closures, Learning Losses and Intergenerational Mobility. Policy Research Working Papers 2023, 10381. World Bank. Available online: <https://openknowledge.worldbank.org/handle/10986/39607> (accessed on 20 August 2025).
14. Bulková, Z.; Dedík, M.; Štefancová, V.; Gašparík, J. Proposal of the systematic measures to support rail passenger transport during the pandemic period. In Proceedings of the 26th International Scientific Conference Transport Means 2022, Kaunas, Lithuania, 5–7 October 2022; pp. 378–383.
15. Štefancová, V.; Harantová, V.; Mazanec, J.; Mašek, J.; Brůhová Foltýnová, H. Analysis of passenger behaviour during the COVID-19 pandemic situation. *LOGI—Sci. J. Transp. Logist.* **2023**, *14*, 203–214. [CrossRef]
16. Dedík, M.; Bulková, Z.; Gašparík, J.; Šperka, A. How to ensure safe mobility during the pandemic? *Transp. Res. Procedia* **2023**, *74*, 64–71. [CrossRef]
17. Bulková, Z.; Čamaj, J.; Šperka, A.; Jursová, S. Simulation of travel document purchase measures aimed at preventing the spread of COVID-19—Case study from Slovakia. *Transp. Res. Interdiscip. Perspect.* **2023**, *20*, 100853. [CrossRef]
18. Singh, J.; Homem de Almeida Correia, G.; van Wee, B.; Barbour, N. Change in departure time for a train trip to avoid crowding during the COVID-19 pandemic: A latent class study in the Netherlands. *Transp. Res. Part A Policy Pract.* **2023**, *170*, 103628. [CrossRef]
19. Deplomo, B.N.I.; Villaberde, J.F.; Paglinawan, A.C. Train passengers prediction for railway management system using regression analysis. In Proceedings of the 7th International Conference on Computer and Communications, Chengdu, China, 10–13 December 2021; pp. 1767–1772. [CrossRef]
20. Lei, B.; Liu, X.; Cao, Z.; Hao, Y.; Zhang, Y.; Chen, X. Modeling and forecasting of COVID-19 spread in urban rail transit system. *J. Traffic Transp. Eng.* **2020**, *20*, 139–149. [CrossRef]

21. Elias, W.; Zatmeh-Kanj, S. Extent to which COVID-19 will affect future use of the train in Israel. *Transp. Policy* **2021**, *110*, 215–224. [CrossRef] [PubMed]
22. Ferreira, S.; Amorim, M.; Lobo, A.; Kern, M.; Fanderl, N.; Couto, A. Travel mode preferences among German commuters over the course of COVID-19 pandemic. *Transp. Policy* **2022**, *126*, 55–64. [CrossRef] [PubMed]
23. Zhang, J.; Zhang, R.; Ding, H.; Li, S.; Ma, S.; Zhai, B.; Kashima, S.; Hayashi, Y. Effects of transport-related COVID-19 policy measures: A case study of six developed countries. *Transp. Policy* **2021**, *110*, 37–57. [CrossRef]
24. Guo, Y.; Hao, Y.; Zhang, G.; Ma, D.T. Exploring the impacts of travel-implied policy factors on COVID-19 spread within communities based on multi-source data interpretations. *Health Place* **2021**, *69*, 102538. [CrossRef] [PubMed]
25. Bian, Z.; Zuo, F.; Gao, J.; Chen, Y.; Pavuluri Venkata, S.S.C.; Duran Bernardes, S.; Ozbay, K.; Ban, X.J.; Wang, J. Time lag effects of COVID-19 policies on transportation systems: A comparative study of New York City and Seattle. *Transp. Res. Part A Policy Pract.* **2021**, *145*, 269–283. [CrossRef]
26. Tan, L.; Ma, C. Choice behavior of commuters' rail transit mode during the COVID-19 pandemic based on logistic model. *J. Traffic Transp. Eng.* **2020**, *8*, 186–195. [CrossRef]
27. Molin, E.; Kroesen, M. Train travel in corona time: Safety perceptions of and support for policy measures. *Transp. Res. Part A Policy Pract.* **2022**, *158*, 196–209. [CrossRef]
28. Ding, H.; Zhang, J. Dynamic associations between temporal behavior changes caused by the COVID-19 pandemic and subjective assessments of policymaking: A case study in Japan. *Transp. Policy* **2021**, *110*, 58–70. [CrossRef] [PubMed]
29. Cartenì, A.; Di Francesco, L.; Martino, M. The role of transport accessibility within the spread of the coronavirus pandemic in Italy. *Saf. Sci.* **2021**, *133*, 104999. [CrossRef]
30. Coppola, P.; De Fabiis, F. Impacts of interpersonal distancing on-board trains during the COVID-19 emergency. *Eur. Transp. Res. Rev.* **2021**, *13*, 13. [CrossRef] [PubMed]
31. Ding, P.; Feng, S.; Jiang, J. The impact of urban rail transit epidemic prevention measures on passengers' safety perception. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4161. [CrossRef]
32. Haque, M.T.; Hamid, F. An optimization model to assign seats in long distance trains to minimize SARS-CoV-2 diffusion. *Transp. Res. Part A Policy Pract.* **2022**, *162*, 104–120. [CrossRef] [PubMed]
33. Vichiensan, V.; Hayashi, Y.; Kamnerdsap, S. COVID-19 countermeasures and passengers' confidence of urban rail travel in Bangkok. *Sustainability* **2021**, *13*, 9377. [CrossRef]
34. Navarrete-Hernandez, P.; Rennert, L.; Balducci, A. An evaluation of the impact of COVID-19 safety measures in public transit spaces on riders' worry of virus contraction. *Transp. Policy* **2023**, *131*, 1–12. [CrossRef]
35. Thaithatkul, P.; Sanghatawatana, P.; Anuchitchanchai, O.; Laosinwattana, W.; Liang, J.; Chalermpong, S. Travel behavior change of public transport users during the COVID-19 pandemic: Evidence from Bangkok. *Asian Transp. Stud.* **2023**, *9*, 100102. [CrossRef]
36. Haque, M.T.; Hamid, F. Social distancing and revenue management—A post-pandemic adaptation for railways. *Omega* **2023**, *114*, 102737. [CrossRef]
37. Abdullah, M.; Ali, N.; Hussain, S.A.; Aslam, A.B.; Javid, M.A. Measuring changes in travel behavior pattern due to COVID-19 in a developing country: A case study of Pakistan. *Transp. Policy* **2021**, *108*, 21–33. [CrossRef]
38. De Vos, J. The effect of COVID-19 and subsequent social distancing on travel behavior. *Transp. Res. Interdiscip. Perspect.* **2020**, *5*, 100121. [CrossRef]
39. Gašparík, J.; Bulková, Z.; Dedík, M. Prediction of transport performance development due to the impact of COVID-19 measures in the context of sustainable mobility in railway passenger transport in the Slovak Republic. *Sustainability* **2024**, *16*, 5283. [CrossRef]
40. Bulková, Z.; Dedík, M.; Lovíšek, M.; Janošková, K.; Vaculík, J. Increasing the quality and efficiency of the transportation process in railway passenger transport in the case of another pandemic period. *Transp. Tech. Technol.* **2023**, *19*, 7–12. [CrossRef]
41. Mašek, J.; Pálková, A.; Blaho, P.; Halajová, Š.; Jursová, S.; Šipuš, D. Proposal for using IT solutions in public passenger transport in Slovak Republic to reduce the spread of COVID-19. *LOGI—Sci. J. Transp. Logist.* **2023**, *14*, 181–191. [CrossRef]
42. Public Health Authority of the Slovak Republic. Development of the Pandemic Situation at the District Level in the Slovak Republic. 2021. Available online: <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/> (accessed on 20 August 2025).
43. Gavin, P.H. The Levenberg-Marquardt Algorithm for Nonlinear Least Squares Curve-Fitting Problems. Duke University, 2024. Available online: <https://people.duke.edu/~hpgavin/lm.pdf> (accessed on 20 August 2025).
44. Majerčák, J.; Vakulenko, S.P. The impact of COVID-19 pandemic on population mobility in the Czech Republic and Slovakia. *LOGI—Sci. J. Transp. Logist.* **2023**, *14*, 158–168. [CrossRef]
45. Fazio, M.; Pluchino, A.; Inturri, G.; Le Pira, M.; Giuffrida, N.; Ignaccolo, M. Exploring the impact of mobility restrictions on the COVID-19 spreading through an agent-based approach. *J. Transp. Health* **2022**, *25*, 101373. [CrossRef] [PubMed]
46. Tirachini, A.; Cats, O. COVID-19 and public transportation: Current assessment, prospects, and research needs. *J. Public Transp.* **2020**, *22*, 1–21. [CrossRef] [PubMed]
47. Parino, F.; Zino, L.; Porfiri, M.; Rizzo, A. Modelling and predicting the effect of social distancing and travel restrictions on COVID-19 spreading. *J. R. Soc. Interface* **2021**, *18*, 20200875. [CrossRef]

48. Noland, R.B. Mobility and the effective reproduction rate of COVID-19. *J. Transp. Health* **2021**, *20*, 101016. [CrossRef]
49. Department for Transport. *Transport Use by Mode: Great Britain, Since 16 March 2020*; Department for Transport: London, UK, 2021. Available online: <https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-covid-19-pandemic> (accessed on 20 August 2025).
50. Anke, J.; Francke, A.; Schaefer, L.M.; Petzoldt, T. Impact of SARS-CoV-2 on the mobility behaviour in Germany. *Eur. Transp. Res. Rev.* **2021**, *13*, 10. [CrossRef] [PubMed]
51. Bin, E.; Andruetto, C.; Susilo, Y.; Pernestål, A. The trade-off behaviours between virtual and physical activities during the first wave of the COVID-19 pandemic period. *Eur. Transp. Res. Rev.* **2021**, *13*, 14. [CrossRef]
52. Jenelius, E.; Cebecauer, M. Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales and passenger counts. *Transp. Res. Interdiscip. Perspect.* **2020**, *8*, 100242. [CrossRef] [PubMed]
53. Molloy, J.; Schatzmann, T.; Schoeman, B.M.; Tchervenkova, C.M.; Hintermann, B.M.; 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]
54. De Haas, M.; Faber, R.; Hamersma, M. How COVID-19 and the Dutch ‘intelligent lockdown’ changed activities, work, and travel behaviour: Evidence from longitudinal data in the Netherlands. *Transp. Res. Interdiscip. Perspect.* **2021**, *6*, 100150. [CrossRef]
55. Irawan, M.Z.; Belgiawan, P.F.; Joewono, T.B.; Bastarianto, F.F.; Rizki, M.; Ilahi, A. Exploring activity-travel behavior changes during the beginning of COVID-19 pandemic in Indonesia. *Transportation* **2022**, *49*, 529–553. [CrossRef]
56. Arellana, J.; Márquez, L.; Cantillo, V. COVID-19 outbreak in Colombia: An analysis of its impacts on transport systems. *J. Adv. Transp.* **2020**, *2020*, 8867316. [CrossRef]
57. Gibbs, H.; Liu, Y.; Pearson, C.A.B.; Jarvis, C.I.; Grundy, C.; Quilty, B.J.; Diamond, C.; LSHTM CMMID COVID-19 Working Group; Eggo, R.M. Changing travel patterns in China during the early stages of the COVID-19 pandemic. *Nat. Commun.* **2020**, *11*, 5012. [CrossRef]
58. Huang, J.; Wang, H.; Fan, M.; Zhuo, A.; Sun, Y.; Li, Y. Understanding the impact of the COVID-19 pandemic on transportation-related behaviors with human mobility data. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA, USA, 6–10 July 2020; pp. 3443–3450. [CrossRef]
59. Hensher, D.A.; Wei, E.; Beck, M.J.; Balbontin, C. The impact of COVID-19 on cost outlays for car and public transport commuting: The case of the Greater Sydney Metropolitan Area after three months of restrictions. *Transp. Policy* **2021**, *101*, 71–80. [CrossRef]
60. Parady, G.; Taniguchi, A.; Takami, K. Travel behavior changes during the COVID-19 pandemic in Japan: Analysing the effects of risk perception and social influence on going-out self-restriction. *Transp. Res. Interdiscip. Perspect.* **2020**, *7*, 100181. [CrossRef]
61. 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*, e417–e424. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Integrated Information System for Parking Facilities Operations and Management

Vasile Dragu, Eugenia Alina Roman, Mircea Augustin Roșca, Floriana Cristina Oprea, Andrei-Bogdan Mironescu and Oana Maria Dinu *

Faculty of Transport, National University of Science and Technology Politehnica Bucharest,
060042 Bucharest, Romania; vasile.dragu@upb.ro (V.D.); eugenia_alina.dragu@upb.ro (E.A.R.);
mircea.rosca@upb.ro (M.A.R.); cristina.oprea@upb.ro (F.C.O.); andrei.mironescu@tpbi.ro (A.-B.M.)

* Correspondence: oana.dinu@upb.ro; Tel.: +40-745504092

Abstract: Parking management and operation represent a major challenge for both users and administrators, who seek to ensure efficient utilization, accommodate as many demands as possible, and reduce maintenance costs. This paper presents a theoretical model for an integrated IT system designed for parking management and administration. The modeling process involved designing a parking facility using the AutoCAD Vehicle Tracking v25.00.2775 software package, in accordance with current design standards. To simulate system operation, a dedicated Python v2025.12.0 program was developed to assign parking spaces to arriving vehicles based on specific allocation criteria. Three allocation strategies were applied: random allocation, allocation aimed at minimizing the driving distance within the parking lot, and allocation aimed at reducing the walking distance from the assigned space to the destination. The simulation results show that, in the absence of allocation criteria, parking spaces are utilized in a quasi-uniform manner. The calculated values of variance and standard deviation are significantly lower in this case, increasing as allocation restrictions are introduced, but then returning to reduced values as the occupancy rate grows, since under intensive use the potential for controlled allocation decreases. The relationship between the number of allocations of each parking space and the applied allocation strategies was examined using Pearson and Spearman correlation coefficients. The results reveal a direct linear dependence under moderate demand and an inverse dependence under high demand—patterns consistent with situations observed in practice. The proposed software application provides a practical tool for effective parking management, contributing to the rational use of parking spaces, reduced travel distances within the facility, lower fuel consumption, and consequently, reduced pollution.

Keywords: parking systems; management; operation; optimization; performance metrics

1. Introduction

Parking, particularly for passenger cars, is an essential component of the transportation system. It plays a significant role in traffic planning, with the potential to either encourage or discourage transport demand. Parking involves considerable social costs, and its capacity planning, location, and management represent a complex issue for both transport planners and public authorities [1]. In the absence of appropriately dimensioned and efficiently managed parking facilities, vehicles tend to park along the roadway, resulting in the occupation of the curbside lane and a consequent reduction in the effective number and width of traffic lanes. This situation has major implications for traffic flow and network performance. In many cases, parking infrastructure is either insufficient or mismanaged,

compelling drivers to travel additional distances in search of an available parking slot [2,3]. Parking accessibility and the satisfaction of the driver are major concerns when it comes to the development of smart cities [4]. This has led to an increase in traffic congestion, a decrease in zone accessibility and parking spaces, and the need for specific measures to estimate accessibility [5].

When properly implemented, parking management can significantly reduce the number of parking slots required in a given area while providing a range of economic, social, and environmental benefits [6]. Parking management can be categorized into on-site and area-wide management. This study focuses on on-site management, which addresses how a single parking facility can be administered in order to maximize benefits and ensure the rational use of parking slots. Area-wide management refers to the coordination of multiple parking facilities within a specific zone and the allocation of vehicles based on various criteria—such as the minimum travel distance to the assigned facility, search costs, fuel consumption, or the number of vehicle occupants [7–11]. As a result, traffic intensity in the area is reduced, contributing to the mitigation of congestion—an issue that increasingly absorbs substantial community resources, both economic and environmental [12]. It is estimated that approximately 30% of traffic congestion in central urban areas is caused by vehicles searching for a parking slot, with the average search time being around 7 to 8 min [13,14].

In 2024, TomTom, a company specializing in traffic coordination systems and traffic analysis [15], conducted a study on congestion levels in 500 cities across 62 countries worldwide, including 254 cities from Europe. According to this analysis, Bucharest ranked first in Europe in terms of congestion levels, and fifth globally, with a congestion level of 48% (compared to 55% in 2023, 52% in 2020, and 41% in 2015). Bucharest was followed by Plovdiv (Bulgaria), Łódź (Poland), Dublin (Ireland)—each also reporting a 48% congestion level—and Hull (United Kingdom) with 46%. In this ranking, the top three cities share the same congestion level, with their positions differentiated based on additional traffic quality criteria. It is notable that between 2015 and 2023, the congestion level in Bucharest increased steadily, while in 2024, a decrease was recorded. However, despite this reduction, Bucharest remains first in the European ranking, whereas in previous years it occupied fifth place. This suggests that other cities have managed to reduce traffic congestion more effectively than the public administration in Bucharest.

With the widespread implementation of the smart city model across urban areas, a range of intelligent digital solutions has been developed. Among these, smart parking systems stand out by integrating advanced technologies with human-centered innovation, aiming to optimize resource utilization and reduce the consumption of non-renewable energy sources [16–18]. As a result, urban areas can benefit from faster, more convenient, and higher-density vehicle parking, ensuring that each parking slot is utilized efficiently. Ultimately, smart parking systems are expected to contribute to more effective urban management, leading to reduced traffic congestion and lower parking operation and management costs [19,20].

Smart parking systems contribute to positive changes for various urban stakeholders by reducing the incidence of traffic accidents, particularly those caused by driver inattention, as drivers are often distracted while searching for available parking spaces or rushing to occupy vacant ones [21,22]. The use of parking applications on mobile devices enables the localization and advanced reservation of parking spaces, eliminating the need for drivers to rush in search of an available spot. In addition to assisting drivers, smart parking systems are also expected to support parking operators and facility owners in maximizing the utilization of available space and resources in a way that increases revenue while enhancing the overall parking experience for their customers [23]. The use of smart parking

systems enables administrators to implement dynamic pricing strategies based on real-time occupancy levels. Consequently, these systems provide both economic and operational benefits for parking facility owners, local authorities, and end users [24].

Overall, the literature underscores the need for integrative approaches that balance algorithmic efficiency with infrastructural constraints and user diversity. Moreover, the system's performance under peak visitor demand or in socioeconomically mixed districts remains unexplored.

Based on these considerations, the research team developed a fully original integrated IT system for simulating the allocation of parking spaces. The allocation is dynamic in both time and space, and is performed according to several parameters, such as the following: assigning a space so that the driving distance within the parking lot is minimized; assigning a space so that both the driving distance within the lot and the walking distance from the parking space to the final destination are minimized; or assigning spaces randomly so that demand across parking areas is evenly distributed. The simulation model was implemented in a Python v2025.12.0 programming environment for a large-scale parking lot and a fleet of 500 vehicles. This number can be easily adjusted depending on parking demand intensity. The model is innovative both in terms of its programming environment and its allocation criteria. It was validated through a case study of a large-scale parking lot (1000 m × 510 m) but can be applied to parking facilities of any size and with any space arrangement, in compliance with current design standards. Simulation parameters can be easily modified within the program, which makes this model universally applicable to any parking configuration—unlike other models found in the literature. This flexibility allows parking administrators to adapt their management policies to meet both system demands and user requirements. For each allocation criterion used in the simulation, key operational indicators of the parking facility are evaluated, and conclusions are formulated to ensure that the benefits are maximized for both the administrator and users.

These insights may support improved parking utilization by enabling the implementation of differentiated pricing strategies based on the degree to which the defined parameters are satisfied. In doing so, this study addresses an existing gap in the technical literature on parking systems, which generally focuses on the interaction between traffic flow and parking, or on the relationship between parking demand and the urban areas served by road infrastructure.

This document is structured as follows: in Section 2, the paper delves into the existing research and literature that provides contextual and general evidence on the main issues in urban parking, slot allocation in private parking areas, priority routines and fairness policies, allocation, and reservation systems. Section 3 outlines the design methodology, the spatial configuration, and the analytical tools used to implement a rule-based, criteria-driven distribution of parking spaces, and Section 4 presents the simulation of the parking facility's operation, with its logical framework and functional description detailed in the previous section. Further, Section 5 succeeds with an analysis of the distribution of parking slot usage, synthesizes the results, and engages in assessing the implications of the observed data. Finally, Section 6 draws conclusions on the automation and digitalization of parking systems and the indirect contribution to improved traffic flows, thus mitigating the negative externalities of road congestion.

2. Literature Review

2.1. Main Issues in Urban Parking Research

Urban areas frequently face imbalances between parking demand and supply, which manifest either as oversupply—leading to inefficient land use—or as shortages, resulting in traffic congestion caused by “cruising” for parking. Additionally, a spatial and functional

mismatch between the type and location of available parking and actual demand further complicates the issue. Addressing this requires adaptive policies, including dynamic pricing mechanisms and zoning regulations [25].

The study of [26] addressed parking congestion through data-driven strategies, including dynamic pricing, infrastructure expansion, and policy interventions. The analysis was limited as it mainly relied on synthetic projections due to uncollected employment/land-use data, and modern solutions of real-time monitoring (e.g., IoT sensors, mobile apps) were not integrated.

A significant body of literature has investigated how drivers select parking based on a combination of factors, including cost, walking distance, expected search time, safety, and trip purpose [27]. These studies often rely on stated preference surveys and discrete choice modeling, though discrepancies between stated and revealed preferences are commonly reported.

User heterogeneity is a central theme, with behavioral differences observed across user groups (e.g., residents vs. visitors, commuters vs. shoppers) [28]. Cao et al. introduced a behavior model based on cumulative prospect theory to capture drivers' risk attitudes and decision-making biases under uncertainty, particularly focusing on how unreliable cruising times influence choices between on-street and off-street parking [29]. Nevertheless, this approach did not incorporate real-time IoT data (e.g., smart parking sensors) for adaptive decision support.

Rodríguez et al. extended this field by integrating alternative availability and user heterogeneity into a mixed logit framework. Their methodology combined stated preference surveys with discrete choice models—including error component models—to analyze parking behavior. The study accounted for systematic and random taste variation, as well as the correlation between alternatives, offering actionable insights for cities on how pricing policies and arrival times influence parking demand [27].

Smart parking research increasingly emphasizes the use of sensors and mobile applications to reduce search time by guiding drivers to available parking slots [30]. Additional solutions include park-and-ride strategies, which integrate parking with public transit to reduce inner-city traffic volumes [27]. These initiatives are aligned with broader efforts to optimize urban mobility using data-driven parking policies, as highlighted in [25,31]. However, implementation challenges remain—particularly regarding data privacy, infrastructure investment, and user adoption barriers.

Contemporary parking management approaches increasingly shift from supply-side solutions—such as expanding parking stock—to demand-side strategies that focus on optimizing existing capacity. These include policies aimed at improving turnover, efficiency, and support for sustainable urban planning [32].

Shared parking arrangements allow multiple users with different peak usage times to access the same spaces, reducing the overall number of required parking spots and improving land-use efficiency [25].

Time restrictions are another commonly used tool to increase parking space utilization in high-demand zones. Several studies have examined the effectiveness of such policies: Weinberger discussed their influence on driver behavior and parking space utilization [33]; Marsden provided a comprehensive review of parking regulations [34]; Jakobsson explored how time restrictions affect both parking behavior and mode choice [35]; Arnott modeled the impact of parking time limits on traffic congestion and efficiency [36]; and Litman offered practical guidelines for implementation [37].

2.2. Main Issues in Slot Allocation in Private Parking Areas

One of the most persistent challenges in the management of private parking areas lies in how individual slots are allocated. The traditional approach—assigning fixed slots to specific users—remains widespread due to its administrative simplicity and perceived fairness. However, this model often results in inefficiencies, particularly in mixed-use facilities or residential buildings where vehicle ownership varies across households, or where peak usage periods are not aligned. Underutilization of assigned slots during working hours or extended absences is a frequent occurrence [13]. Although such studies highlight the mismatch between allocation and actual demand, there is limited exploration of how user behaviors (e.g., sharing agreements, informal reallocations) influence or mitigate these inefficiencies.

In response, dynamic allocation schemes have emerged as promising alternatives. These models rely on real-time data to assign slots based on current availability, the estimated parking time, or user category. Solutions like the Dy PARK system incorporate dynamic pricing and rematching algorithms to optimize usage under fluctuating demand conditions [38]. However, while simulation outcomes suggest improved slot utilization and user throughput, such results are context-dependent and are rarely validated in operational environments. The Dy PARK model, for instance, assumes broad technological penetration and high user compliance—factors that remain uncertain in many residential settings. This illustrates a broader gap in the literature: the transition from simulation to scalable, user-accepted deployment remains insufficiently addressed.

Several contributions have attempted to formalize parking slot assignment using optimization methods. Genetic algorithms have been employed to reduce search time and on-site traffic congestion in complex facilities like shopping centers [39]. Similarly, Elfaki et al. proposed a constraint optimization model that accounts for parking spot availability, proximity, and priority categories [40]. While these models show computational effectiveness, they frequently assume homogeneity in vehicle types, driver behavior, and infrastructure—assumptions that limit their adaptability to real-world constraints. Moreover, most algorithmic approaches do not include mechanisms for handling social or legal constraints, such as the rights of parking slot holders, parking disputes, or allocation fairness.

Beyond allocation logic, space optimization has received attention as a complementary research stream. Mladenović et al. introduced a scalable dynamic framework that adapts to usage patterns across heterogeneous facilities, and Nakazato et al. explored dynamic pricing to balance demand across time slots [41,42]. These works provide robust mathematical models and emphasize operational scalability. However, a notable limitation is their lack of attention to user-centered outcomes: few studies have analyzed how these models impact user satisfaction, acceptance, or perceptions of fairness—particularly in environments where space is both limited and socially sensitive, such as residential complexes.

Reservation-based systems are often proposed as partial solutions, enabling more predictable parking behavior and reduced internal traffic. Yet even here, critical issues remain. While such systems are effective in structured environments like office buildings or automated garages, their adaptability to legacy infrastructures and informal user cultures is uncertain [43]. Additionally, few studies have addressed the long-term maintenance, system fatigue, or policy implications of relying on automated scheduling, particularly in settings where not all users have equal digital access.

Although some studies have attempted to link technical allocation systems with user preferences—such as through real-time slot suggestions or price-based incentives—there is minimal integration between algorithmic design and behavioral science. This disconnect leaves open critical questions about accessibility, trust, and the actual willingness of users

to adapt to dynamic or automated allocation models. As the focus shifts from system performance to user access and fairness, another critical question arises: how should parking be distributed when demand exceeds supply, particularly in residential or mixed-use environments?

2.3. Priority Routine and Fairness Policies

In residential and mixed-use parking environments, slot allocation must address not only operational efficiency but also equity of access, particularly when the parking demand exceeds capacity. Unlike public lots, private and semi-private parking structures often cater to defined user groups (e.g., residents, staff, and visitors), making the perceived fairness of allocation routines a central issue. From an engineering management perspective, fairness policies must be designed to balance utilization efficiency, user satisfaction, and system enforceability under spatial and behavioral constraints.

One of the most intuitive allocation strategies is First-Come-First-Served (FCFS). While widely implemented due to its simplicity and transparency, FCFS does not account for user heterogeneity. In their empirical study on reserved residential parking systems, Yang et al. found that residents tend to perceive FCFS as inequitable when it consistently favors early arrivals—often visitors or commuters—at the expense of long-term tenants. This suggests a key limitation in relying on time-of-arrival alone as an allocation metric in environments with habitual or entitled users. The literature lacks longitudinal studies analyzing how repeated FCFS exposure influences resident satisfaction or informal workarounds (e.g., unofficial slot exchanges) [44].

To address equity concerns, priority-based allocation systems have emerged, often giving preferential access to elderly users, people with disabilities, electric vehicles (EVs), or residents with seniority. Chen et al. developed a comparative model to analyze how such a system balances efficiency and fairness. Their results show that while priority rules improve access for vulnerable or high-need users, they can reduce the total system throughput under constrained capacity. The trade-off curve between allocative fairness and operational efficiency remains a key gap, as few studies have addressed real-time system adaptability in response to demand shifts or behavioral compliance [45].

Another widely used approach is the implementation of permit-based systems, where access is pre-assigned to specific users through long-term agreements. Xu & Sun proposed a dynamic permit reservation model that adjusts slot assignments based on usage frequency while penalizing monopolization of premium locations. Their formulation explicitly incorporates a fairness function into the cost optimization process, a critical contribution to bridging algorithmic efficiency and ethical considerations. However, their study did not explore how such systems scale in socially diverse or informally governed parking environments, where enforcement and digital access may be inconsistent [46].

An alternative to gaining traction is a cyclic slot allocation strategy, where access to premium parking spots is distributed in turn among eligible users over a fixed time interval. This approach has been explored primarily in the context of electric vehicle (EV) charging infrastructure. Panayiotou et al. demonstrated that fair time-based cycling, combined with reservation windows, enhances temporal equity and reduces peak-period traffic congestion. Although the study focused on EV facilities, the underlying principles are broadly applicable to conventional parking environments. However, implementing such cyclic strategies requires high levels of user compliance and strong enforcement mechanisms—factors that remain insufficiently examined in residential contexts where parking personalization is often the default [47].

Recent contributions have also highlighted the role of digital systems in enabling adaptive and differentiated access control. Alahmadi et al. introduced a privacy-enhanced

parking management framework that integrates user profiles into slot reservation logic, assigning different rules based on attributes such as driver age, vehicle type, or mobility level. Their system demonstrated scalability and security through a blockchain-based infrastructure. Yet, the study did not address sociotechnical adoption barriers such as digital illiteracy, low smartphone penetration among older residents, or resistance to algorithmic decisions—all of which can affect real-world effectiveness [48].

The complexity of fairness increases further in mixed-use urban environments, where residents and non-residents often compete for the same parking infrastructure. İnaç presented a case study from Istanbul where proximity-based digital permits were introduced to prioritize residents over visitors. The deployment significantly reduced unauthorized occupation and improved satisfaction scores among permit holders. Importantly, the study emphasized the role of transparency and communication, suggesting that procedural fairness—how rules are applied and explained—can be as important as outcome fairness. However, the system’s performance under peak visitor demand or in socioeconomically mixed districts remains unexplored [49].

Considering electrification trends, the intersection of fairness policies and EV infrastructure presents an emerging frontier. Ebrahimi et al. investigated how EV parking and charging lots can support grid operations under uncertainty. Although their focus was on distribution systems, the implications for user-level fairness are critical: how should limited charging-enabled spots be shared among growing EV populations with different usage patterns and priorities? Their model integrated stochastic demand but did not address policy-level questions such as whether EV drivers with longer parking times should have the same access priority as short-term users [50].

2.4. Smart Allocation and Reservation Systems

Smart allocation and reservation systems represent a pivotal research direction in the context of intelligent transport and parking infrastructure. The core objective of these systems is to dynamically assign parking slots to users in real time, based on availability, cost, proximity, and—in more advanced designs—anticipated demand and user preferences.

A foundational contribution by Drašković and Vukićević introduced a branch-and-bound optimization model, supported by wireless sensors and Bluetooth communication, to guide vehicles to the nearest vacant slot in interconnected parking blocks. While the framework demonstrates algorithmic feasibility, its testing was limited to controlled environments with ideal infrastructure assumptions. Moreover, it does not integrate predictive features such as demand anticipation or user-specific preferences, which are crucial for adaptive urban deployments [51].

More recent developments have addressed some of these limitations. Wang et al. proposed a heuristic model based on Lagrangian relaxation, which optimizes the allocation of shared private parking spaces while accounting for EV charging requirements and platform-level rules. However, despite its scalability in large scenarios, the model assumes full information and does not capture demand volatility or user behavior [52].

Likewise, Burns et al. contributed a dynamic reservation framework focused on delivery vehicles, formulating the problem as a sliding time window mixed-integer program. While this approach significantly improves utilization under structured demand, its applicability in general-purpose parking environments remains limited by its assumptions of reliable arrival forecasts and delivery-focused constraints [53].

On the practical implementation side, Renuka and Dhanalakshmi developed a mobile-based smart parking system using slot allocation, IR sensors, and RFID-based billing. Though effective in small-scale settings, it relies on rule-based scheduling (e.g., FCFS) without optimization or feedback control [54]. Aswath et al. expanded on hardware-

oriented approaches by integrating YOLOv8-based vision, Raspberry Pi processing, and web-based reservation interfaces, yet still without a decision layer capable of dynamic prioritization or equity-based allocation [55].

Finally, Zhang et al. synthesized the research landscape across four dimensions—inventory control, allocation decision-making, pricing, and performance evaluation—and highlighted important missing elements such as fairness, demand elasticity, and multimodal integration [56]. In parallel, Ismael and Holguín-Veras tackled the problem of parking allocation under vehicle heterogeneity, formulating a mixed-integer optimization model that explicitly distinguishes between vehicle classes such as cars, vans, and trucks. Their model emphasizes capacity constraints and compatibility rules in shared facilities, showing notable gains in total system efficiency and fairness. However, the study operated under centralized information assumptions and did not incorporate behavioral or stochastic uncertainty in arrival patterns [57].

Despite these contributions, current systems still lack comprehensive integration of adaptive, user-centered, and context-aware logic. There remains a need for hybrid systems that not only allocate resources efficiently but also adapt to stochastic conditions, incorporate behavioral modeling, and ensure equitable access in diverse urban contexts.

2.5. Physical Design Constraints and Integration into Smart Parking Systems

The physical design of parking facilities fundamentally determines their usability, efficiency, and integration with evolving urban mobility systems. Core design parameters such as parking slot dimensions, aisle widths, and layout geometry are tightly linked to easy movement, spatial yield, and user experience. Yet, as the complexity of urban mobility systems increases—with the introduction of automated vehicles (AVs), electric vehicle (EV) infrastructure, and sensor-based guidance systems—traditional geometric principles alone are no longer sufficient.

At the base of parking layout design lies the notion of the design vehicle, as articulated by Mikušová et al., which provides a structured approach to defining turning radius, parking slot lengths, and clearance zones. This model ensures compatibility with a range of vehicle types, including commercial and articulated vehicles, offering a robust baseline for geometric planning. However, it inherently assumes homogeneity in traffic composition and static infrastructure, overlooking the increasingly diverse and dynamic nature of road vehicle composition [58].

A complementary approach is offered by mathematical models such as that of Maina et al., who used mixed-integer linear programming (MILP) to identify optimal parking angles [59]. Their results, showing 75° as the most space-efficient configuration, align with empirical optimization efforts like those by Wang et al., who, through field-based layout trials in a surface parking lot in Panzhuhua, demonstrated how aisle width and angle impact capacity and flow [60]. Still, both studies operated under idealized, static conditions and lacked adaptability to shifting demand, temporary closures, or real-time reconfiguration.

The geometric demands of automated vehicle systems challenge these static assumptions even further. Kong et al. demonstrates that AVs, by eliminating human error and reducing the space required for vehicle movements, allow for much denser configurations [61]. Chen and Sun extended this by proposing hybrid layouts accommodating both AVs and conventional vehicles, integrating structural constraints such as EV charging points and obstacle buffers. However, these studies were predominantly simulation-based and overlooked behavioral variability, pedestrian movement patterns, and clarity of traffic signs—factors that strongly influence usability in real deployments [62].

3. Materials and Methods

Based on the insights outlined in the literature, the present study transitions from conceptual considerations to the practical development of a structured parking allocation model. This section outlines the design methodology, the spatial configuration of the facility, and the analytical tools used to implement a rule-based, criteria-driven distribution of parking spaces.

The design of the information system for parking management and administration begins with the physical layout planning of the parking facility, which was carried out in accordance with [63], specifying the structural dimensions of parking slots as well as the adopted orientation angles of the parking layout—90, 45, 60, and 75 degrees. The design process was conducted using the AutoCAD Vehicle Tracking v25.00.2775 software, selecting a 90-degree parking orientation relative to the circulation direction in the parking area, as this configuration represents the most applied solution in practice.

AutoCAD Vehicle Tracking v25.00.2775 is a software tool used for analyzing vehicle movement trajectories. It enables designers to simulate vehicle movements within parking facilities, intersections, or confined spaces, assessing accessibility and maneuverability. Among its extended functionalities, the following can be highlighted:

- Simulation of vehicles with varying dimensions, including passenger cars, trucks, and buses.
- Vehicle path analysis, by identifying critical areas where large vehicles cannot fit within the geometric constraints.
- Integration with Building Information Modeling (BIM), allowing 3D visualization of parking layouts in the context of surrounding constructions.
- Generation of reports and the ability to export data to other software, such as automatic generation of traffic flow data and spatial occupancy analysis.

Figure 1 presents the designed layout of the parking facility, which consists of 8 rows (A, B, C, . . . , H), each containing 12 parking slots, resulting in a total of 96 parking slots distributed over an area measuring 1000 m in length and 510 m in width. For the proposed layout, it was assumed that access to the parking facility is made from the northern side, while the exit is through the southern side. It should be noted that the access points, the number of parking slots, as well as their inclination angle, can be configured under any of the aforementioned orientation hypotheses using the AutoCAD Vehicle Tracking v25.00.2775 software.

After designing the parking layout in accordance with the shape and dimensions of the area designated for the parking facility, a program was developed in Python to simulate vehicle behavior within a public parking lot. The main features of the simulation are as follows:

- Dynamic allocation of parking slots based on predefined criteria, as detailed below:
 - (a) random allocation of a free parking slot within the designated parking area,
 - (b) allocation of parking slots such that the driving distance through the parking area to the assigned space is minimized,
 - (c) the same allocation criterion as in point (b), with the additional condition of minimizing the walking distance from the assigned parking slot to the destination. In the simulation, each vehicle is assigned a free space initially marked in green, and once occupied, the indicator changes to red.
- Parking time. This is a random time interval, predefined as a parameter at the beginning of the simulation. The set value used in the simulation is $t \geq 200$ s, after which the vehicle exits the parking area, and the corresponding space becomes available for allocation to another vehicle.

- Real-time visual update. The graphical interface displays, in real time, the occupied or available status of each parking slot through an animation that can be played at a configurable speed.
- Evaluation of operational metrics. The main performance metrics related to parking usage are calculated based on the initial simulation parameters. These include the number of times each parking space is used during the simulation, the average occupancy time per space, the average driving distance within the parking area, and the average walking distance from the parking space to the intended destination.

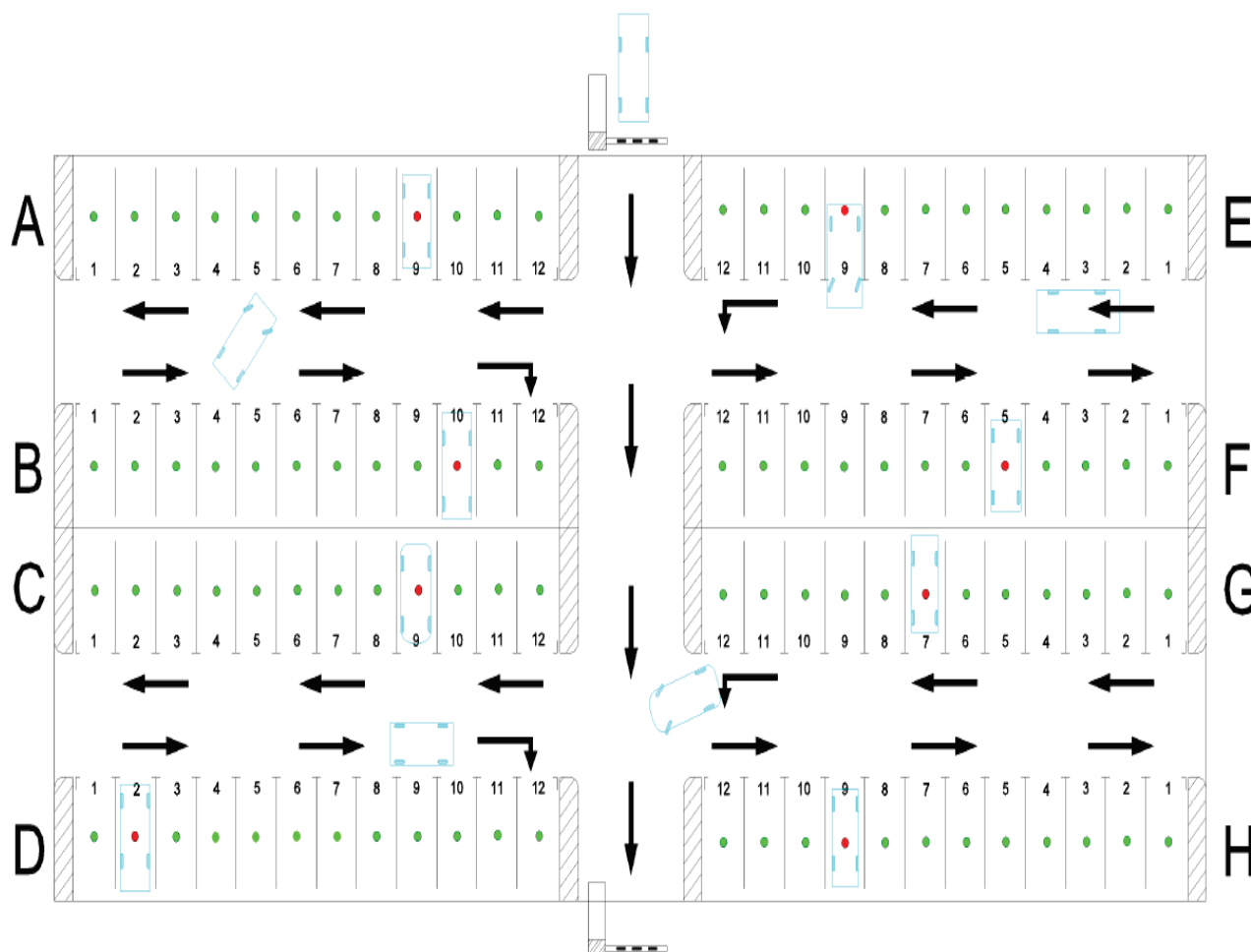


Figure 1. A layout representation of the parking facility using the AutoCAD Vehicle Tracking v25.00.2775 software.

Objectives (a), (b), and (c) of the decision-making process are selected by the parking administrator within the management process, depending on the occupancy rate of the parking facility—determined by the simulation program—and on the declared preferences of the users. For a given time interval (e.g., during a single day), the three management strategies may be applied alternately to increase both the efficiency of use and customer satisfaction. Simulation and modeling of operations can provide threshold values for switching allocation criteria for parking spaces and, implicitly, the tariffs applied. Differentiated pricing can be used depending on the degree to which customer requirements are met, while maintaining budgetary balance. Applying criterion (b) results in a shorter driving route within the parking facility, which benefits the user by reducing fuel consumption, time, and costs; this may justify a higher parking fee, while still maintaining budgetary equilibrium for the user. A similar effect occurs when criterion (c) is applied, in which case, beyond the reduction in the in-parking driving distance, the walking distance

to the destination is also minimized. This, in turn, increases user satisfaction, as all vehicle occupants benefit primarily from time savings. However, applying criteria (b) and (c) leads to more intensive use of certain parking areas and a decline in the use of others. Consequently, this results in uneven wear of the parking infrastructure and may create discomfort in usage. To balance the utilization of different zones, parking management will rely on allocation criterion (a), which naturally entails lower parking costs.

The transition from the simulation model to the real management model will integrate information on user preferences collected in advance, while users will be informed regarding the extent to which they can benefit from the allocation criteria as well as the associated parking costs, thus allowing them to decide whether to use the parking facility or not.

The parking system operates as a queuing system with multiple service stations, where the service stations correspond to the individual parking slots. Arrivals into the system are represented by vehicles seeking to occupy a parking slot, while departures are represented by vehicles leaving the parking facility. The arrival rate is measured in vehicles attempting to enter the parking system within a given time interval ($\lambda = \text{veh/h}$) and is considered a random variable. The service process corresponds to the random parking time, which generates the departure flow, also measured in terms of the number of vehicles exiting the parking facility over a specified period ($\mu = \text{veh/h}$). As long as $\lambda < \mu$, the system remains stable; however, its performance is governed by the value of the utilization ratio $\rho = \frac{\lambda}{\mu}$. The closer this ratio approaches 1, the greater the waiting times generated within the system. In other words, this ratio represents the number of services completed during the time it takes for one unit of demand to arrive for service. If the simulation starts with the arrival of the first vehicle, which finds the parking facility entirely empty, a key question arises: after how much time will the parking facility reach full occupancy? If we denote the total number of available parking slots by n_p , then the total occupancy time (T_p) can be calculated using Equation (1):

$$T_p = \frac{n_p}{\lambda - \mu} [\text{h}] \quad (1)$$

Up to this moment in time (T_p), the value of λ may exceed that of μ . After this point, for the parking facility to continue accepting incoming vehicles, the arrival rate must become equal to or less than the departure rate, i.e., $\lambda \leq \mu$.

The simulation program developed in the Python environment consists of the phases presented in Figure 2.

Figure 3 presents a flowchart of the simulation process, illustrating the functional blocks of the software and their logical sequencing. In this figure, n represents the number of vehicles requesting parking, d_1 represents the distance from the barrier to the parking spot, d_2 represents the distance from the parking slot to the main entrance, and t denotes the parking time.

To ensure that the simulation closely replicates human behavior, three scenarios were proposed:

- (1) A baseline scenario, with no restrictions on the allocation of parking spaces.
- (2) An allocation strategy aimed at minimizing the total travel distance covered by all vehicles within the parking area while searching for a parking space.
- (3) An allocation strategy combining the constraints from scenario 2, while also minimizing the walking distance from the parked vehicle to the destination.

PHASE 1 – Initialization

```

initpygame
hardware & graphics constants ← dimensions, colors, scale
load barrier & arrow images
load and scale vehicle images
create spot lists A–H (left/right)
populate usage_counts (statistics)
num_cars ← Tinter dialog “How many vehicles?”

```

PHASE 2 – Auxiliary Definitions

```

FUNC generate_license_plate() → string
FUNC get_parking_code(spot) → e.g., “A5”
FUNC get_directions_to_spot() → arrows & side direction
CLASS Car
Constructor: selects image, positions vehicle at the entrance gate
– randomly selects a free spot, marks it as occupied
– computes entry and exit paths & directions
– sets parking time (200–250 s)
move() method: FSM with two main branches
▶ Entry/parking (phases 0–4):
0 vertical ↓, 1 rotate 90°, 2 horizontal →/←,
3 rotate to 0°, 4 vertical toward slot → parked = True
▶ Exit (exit_phase 0–4): inverse movements
After parking: logs vehicle in parked_cars and increments usage_counts
draw() method: renders rotated image + trail
Move() method logic:
If parked and time expired → initiate exit
Else, if entering:
execute entry phases 0–4
at phase 4:
parked ← True
store vehicle data in parked_cars
update usage_counts
Else, if exiting:
execute exit phases 0–4
Update trail graphics
Draw(screen) method:
draw rotated car image + trail

```

Scenario 2 – Tabu Search Spot Selection

```

available_spots ← all spots – occupied
cost(spot) ← distance from entrance to spot center
sort available_spots by increasing cost
for each spot in sorted list:
if spot NOT in tabu_list:
target_spot ← spot
break
else (all are tabu):
target_spot ← first spot in sorted list
mark target_spot as occupied
add target_spot to tabu_list
compute entry and exit paths

```

Scenario 3 – Tabu Search with Walking Distance

```

available_spots ← all spots – occupied

```

```

cost(spot) ← distance from entrance to spot center + distance from spot center to store
entrance
sort available_spots by increasing cost
for each spot in sorted list:
if spot NOT in tabu_list:
target_spot ← spot
break
else (all are tabu):
target_spot ← first spot in sorted list
mark target_spot as occupied
add target_spot to tabu_list
compute entry and exit paths

```

PHASE 3 – Main Loop

```

cars_created ← 0; excel_saved ← False
WHILE running
– draw background and parking frame
– draw labels A–H and slot numbers
– if cars_created = num_cars and 4 s have passed:
append Car(...) to cars; cars_created++
– FOR each car in cars:
car.move(); car.draw(); draw trail
– compute and display statistics (free/occupied, etc.)
– if all cars have left AND excel_saved = False:
create two DataFrames (parked_cars, usage_counts)
format and export indicatori_parcare.xlsx
excel_saved ← True
– handle pygame events (QUIT → running ← False)
– pygame.display.update(); clock.tick(30)
END WHILE
pygame.quit()
If window closed prematurely and excel_saved == False:
print warning in console
ELSE:
“Simulation successfully completed”

```

KEY DATA STRUCTURES

```

parked_cars = [ (plate, code, dist_m, duration_min), ... ]
usage_counts = { 'A': {1: n, 2: n, ...}, ... 'H': {...} }
occupied_spots = [ spot_tuples currently in use ]

```

KEY CONCEPTS

Tabu Search avoids assigning any of the last 10 used spots
 Uses a deque(maxlen=10) as the tabu list
 If all available spots are currently in the tabu list, the closest one is still selected.

Figure 2. Phases of the simulation program developed in the Python v2025.12.0.

In the simulation, the points of interest were located between rows C and D, and between rows G and H of the parking facility.

Scenario 1 assigns available parking slots randomly. Users tend to exhibit preferences when choosing a parking slot—such as minimizing walking distance to their point of interest, selecting less crowded or less exposed areas to reduce the risk of vehicle damage, or other personal considerations (e.g., better lighting, wider sidewalks, etc.). This scenario serves as the baseline scenario, providing a reference for comparison with the other allocation strategies to highlight the advantages and disadvantages associated with user-driven choices.

Scenario 3 builds upon the allocation criteria of Scenario 2 by also minimizing the walking distance from the parking slot to the user’s destination. In this case, in addition to reducing internal driving distance, the assignment process considers the convenience of pedestrian access. The simulation assumes that the points of interest are located between rows C and D and between rows G and H of the parking facility. This strategy aims to improve user experience while maintaining efficient vehicle circulation and minimizing discomfort during navigation within confined spaces.

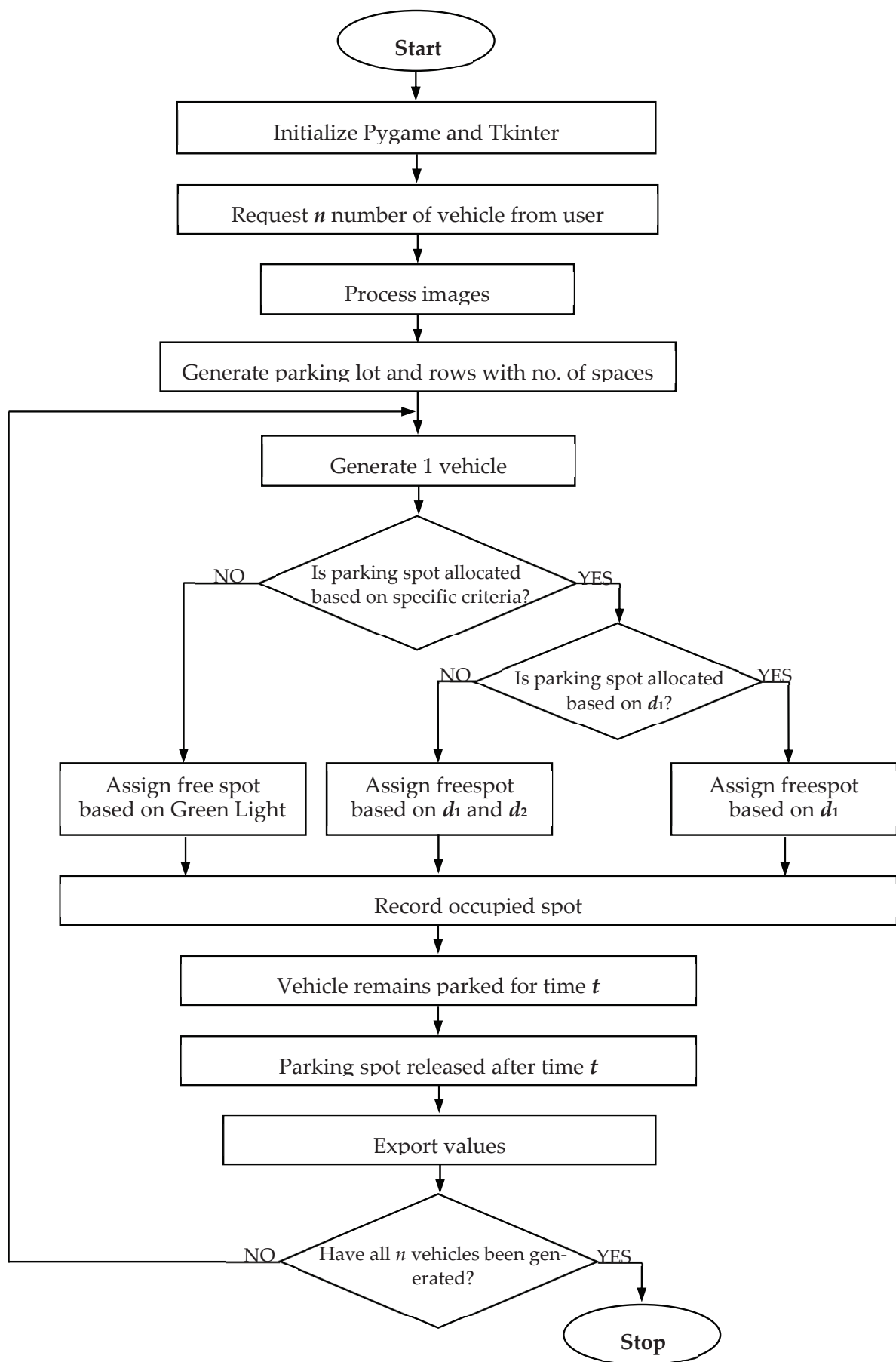


Figure 3. A flowchart of the simulation implemented in Python.

The objective function of this scenario is the minimum total distance traveled by all vehicles within the parking area while searching for an available parking space. This refers specifically to the distance from the entrance of the facility to the parking slot assigned by the program. The objective function is defined by Equation (2).

$$\min \sum_{i,j} d_{ij}^v \quad (2)$$

where v represents the vehicle assigned to the parking spot identified by the coordinates of the corresponding cell i and j :

i —represents the row identifier of the parking cell, with values ranging from A to H;

j —represents the column index of the parking cell, with values ranging from 1 to 12.

Scenario 3 assigns a parking slot while maintaining the constraints from Scenario 2, with the addition of a new requirement aimed at minimizing the walking distance from the assigned parking spot to the point of interest. Thus, to the constraint defined by Equation (1), an additional constraint was introduced—minimizing the total walking distance for all parked vehicles, as calculated by Equation (3):

$$\min(\sum_{i,j} d_{ij}^v + \sum_{i,j} l_{ij}^v) \quad (3)$$

where l_{ij}^v represents the walking distance from the assigned parking slot to the designated point of interest. It should be noted that, in discussions concerning disutility, this walking distance affects each occupant of the parked vehicle. As a simplifying assumption of the model, it was considered that each vehicle is occupied solely by the driver, who is also the one making the parking slot selection decision.

The simulation was carried out for a user-defined number of vehicles, generating information regarding the number of vehicles assigned to each parking slot, the utilization rate of each slot, the average occupancy duration per space, as well as the average occupancy time per row and for the entire parking facility. To ensure a meaningful analysis of parking lot operations, only scenarios where vacant spaces are available were considered. Once the parking facility reaches full occupancy, no additional vehicles are generated until at least one spot becomes available again. This constraint was introduced to assess the dynamics of space occupation and release, recognizing that in non-automated parking systems, all arriving vehicles are granted access regardless of the occupancy level. In such situations, when parking demand is high, newly entering vehicles tend to congest the drive aisles, thereby increasing the exit time for vehicles attempting to leave the facility.

To determine whether there is a dependency between the utilization rate of parking slots and specific parking-related indicators, namely, the distance to the parking slot, or the combined distance composed of the driving segment plus the walking distance to the destination, both Pearson and Spearman correlation coefficients were applied. These correlation coefficients serve as important tools for describing the linear relationship between the selected variables; however, Spearman's correlation is used to quantify dependence and model the variation of nonlinear value pairs [64].

The Pearson correlation coefficient quantifies the strength of the linear relationship between two continuous variables, x and y , by producing values between -1 and $+1$. A value of $+1$ indicates a perfect positive linear dependency, where an increase in one variable is consistently accompanied by an increase in the other. A value of -1 reflects a perfect negative linear dependency, implying that one variable increases as the other decreases proportionally. A value of zero suggests the absence of any linear dependency, indicating that the two variables are not linearly related.

The Pearson correlation coefficient (r_P) which quantifies the degree of linear dependency between two continuous variables, is computed according to Equation (4):

$$r_P = \frac{n \sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)}{\sqrt{[n \sum_i x_i^2 - (\sum_i x_i)^2][n \sum_i y_i^2 - (\sum_i y_i)^2]}}, \quad (4)$$

where x_i and y_i represent the paired values of the two variables being compared, indexed by $i = 1, 2, \dots, n$, with n denoting the total number of observed data pairs for the variables x and y .

The Pearson correlation coefficient specifically assesses both the strength and the direction of the linear dependency between two continuous variables. When the dependency structure between variables deviates from linearity, the Pearson correlation fails to accurately quantify the strength of the association. To address this limitation, alternative methods can be applied—such as Spearman’s rank correlation coefficient, which is particularly suitable for ordinal data and for capturing more complex, non-linear associations. Spearman’s coefficient is used when the goal is to assess the degree of association between two sets of values, even when precise measurements are not available. In certain scenarios, we deal with characteristics that cannot be directly quantified but can be ranked based on a specific criterion. In these cases, the ordinal position of the data is considered rather than its numerical value. We assume a set of n elements ranked according to qualitatively distinct characteristics. For a population of n statistical units $A_1, A_2, \dots, A_n$, each unit is described by two attributes, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. We denote by d_i the difference between the ranks of the two variables for the i -th observation:

$$d_i = x_i - y_i \quad (i = 1, 2, \dots, n), \quad (5)$$

where the vector with components d_i represents a measure of the degree of proximity in the correspondence between the variables x and y . If $d = 0$, it indicates that the variables under consideration have identical ordering for the two characteristics; therefore, a perfect correspondence exists. The greater the values of d_i , the higher the level of discordance between the two variables. To avoid compensation effects between positive and negative differences, the squared values of d_i are used in the calculation. Accordingly, Spearman’s rank correlation coefficient (r_S) is defined by the following expression:

$$r_S = \frac{6 \sum_i d_i^2}{n(n^2 - 1)} \quad (6)$$

As with Pearson’s correlation coefficient, a value of $r_S = 1$ indicates perfect concordance, whereas a value of $r_S = -1$ reflects maximum discordance between the two ranked variables.

Both correlation coefficients were employed to assess the strength of the relationship between parking slot utilization and three key distance metrics: the minimum driving distance covered by vehicles within the parking area to reach the allocated spot; the walking distance from the parking slot to the final destination; and the total distance, defined as the sum of the driving distance and the subsequent walking distance to the destination. The corresponding results are presented in the case study section and are further interpreted in the discussion chapter.

4. Case Study

The case study was conducted on the parking facility illustrated in Figure 1, which comprises 96 parking slots arranged in four rows symmetrically aligned with a central

vertical axis. This axis also serves as the main access and exit route. The four rows are divided by the central axis into eight half-rows, labeled from A to H. Each half-row contains 12 consecutively numbered parking slots. Vehicle movement within the parking facility, to reach the slot automatically allocated by the system, is constrained to two directions: north–south (NS) and east–west (EW). Vehicles also follow these two directions when exiting the facility, which allows for the calculation of the total travel distance within the parking layout. The geometric configuration of the parking area, including slot orientation angles compliant with current standards and physical dimensions, was modeled using the AutoCAD Vehicle Tracking software v25.00.2775. This tool enables the design of any layout adapted to the available surface area for the parking arrangement.

The simulation of the parking facility's operation was implemented using Python, with its logical framework and functional description detailed in the previous section.

The simulation program was executed under three analytical scenarios: (a) random allocation of an available parking slot within the designated facility; (b) allocation of parking slots aimed at minimizing the distance traveled within the parking area to the assigned slot; and (c) the same allocation criterion as in point (b), with the additional objective of minimizing the walking distance from the allocated parking slot to the final destination.

The simulation was conducted for a total of 500 vehicles, with their arrival times at the parking entrance barrier randomly generated by the program. At the start of the simulation, the parking lot is assumed to be empty, entering a loading and transition phase until the moment the first vehicle exits the facility. From that point onward, both entry and exit events occur simultaneously, each governed by a specific intensity parameter defined within the program. The minimum parking time was set to 200 s, while a time shorter than the typical real-world duration was selected to ensure manageable simulation runtimes. This value remains consistent with the time intervals between successive vehicle arrivals requesting a parking slot. Given that the simulation includes both entries and exits, translating the adopted test values to real-world conditions can be achieved by a uniform time-axis scaling of all durations involved. Initially, the simulation was executed for parking times ranging between 200 and 250 s.

A representative segment of the simulation output, including data for the first 11 and the last 11 vehicles generated under scenario (a), is detailed in Table 1. Similar tables were produced for the remaining two scenarios; however, to conserve space, only the aggregated results are shown in Tables 2–4. The license plate number of each vehicle was automatically generated by the simulation program and does not correspond in any way to real-life vehicle registrations. In a real-world implementation, the program could identify each vehicle at the entrance barrier and process this information in accordance with applicable legal regulations regarding the handling of personal data.

Table 1 presents a detailed overview of the 500 vehicles simulated, including the assigned parking slot, the distance traveled from the entrance gate to the designated space, and the parking time.

The distribution of vehicles across individual parking slots, as determined by the simulation under scenario (a), is summarized in Table 2.

As shown in Table 2, the most frequently used parking spot was B7, which accommodated eight vehicles throughout the simulation. Additionally, there were seven parking spots that were each used by seven vehicles: A3, D4, F4, A6, C8, D8, and D11. At the other end of the spectrum, the least used parking spots were A8 and F11, each of which accommodated only two vehicles. Four other spots—E4, B6, C9, and G12—were used by just three vehicles. The sum of all values in Table 2 confirms that the total number of vehicles simulated and successfully parked was 500.

Table 1. Representative simulation output for scenario (a)—parking time: 200–250 s.

No.	License Plate Number	Assigned Parking Slot	Distance Traveled (m)	Parking Time (min)
1	B 206 KTV	B4	28.54	3.63
2	B 610 OJE	B7	23.42	4.1
3	B 262 OWN	A1	30.69	3.87
4	B 366 DZN	C8	26.38	3.63
5	B 289 QZD	F7	23.34	3.75
6	B 603 XCA	C6	29.08	4.17
7	B 34 EJO	B2	32.33	3.93
8	B 707 DCA	E11	9.55	3.48
9	B 537 LTU	D1	44.41	3.93
10	B 250 SEN	C11	23.4	3.62
11	B 952 SUO	A2	28.49	3.52
.				
.				
.				
490	B 585 MUE	G2	35.51	3.7
491	B 491 JYM	D5	38.97	4.07
492	B 810 NCB	D4	40.24	3.57
493	B 946 ONB	G9	25.23	3.73
494	B 277 MJI	A10	11.43	3.87
495	B 603 PPW	G7	27.68	3.98
496	B 410 OJH	E8	15.53	3.72
497	B 490 IQV	B12	17.23	4.1
498	B 718 UNA	A3	26.26	3.83
499	B 570 CDH	H10	34.09	3.6
500	B 733 FZC	B7	23.34	3.93
		Sum	13949.18	1874.82
		Average	27.898	3.75

Table 2. Allocation of vehicles across individual parking slots according to scenario (a).

Row Parking Slot	A	B	C	D	E	F	G	H
1	6	5	5	5	6	6	5	6
2	4	5	4	6	6	5	6	4
3	7	6	6	5	5	5	5	6
4	6	6	5	7	3	7	4	4
5	5	6	5	6	6	5	6	5
6	7	3	6	5	5	6	6	6
7	5	8	4	5	6	4	6	4
8	2	4	7	7	6	5	6	5
9	4	5	3	6	6	4	5	5
10	6	6	5	4	4	4	6	5
11	5	5	4	7	6	2	6	6
12	5	6	5	5	4	6	3	4
Total	62	65	59	68	63	59	64	60

For scenario (b), the vehicle allocation to parking slots—based on the same parking time—was conducted according to the logic described and is presented in Table 3.

From Table 3, it can be observed that the most frequently used parking slot was utilized by nine parked vehicles—this occurred 28 times—and eight parked vehicles in 22 instances. At the opposite end of the spectrum, some parking slots were never used, which occurred in 39 cases. A highly uneven usage pattern is evident, driven by the allocation algorithm’s emphasis on minimizing the in-parking driving distance. This led to a strong

preference for parking spaces located near the entrance of the car park. For scenario (c), the distribution of vehicles across parking slots, under the same parking time conditions, was carried out according to Table 4.

Table 3. Distribution of vehicles per parking slot under scenario (b).

Row Parking Slot	A	B	C	D	E	F	G	H
1	6	0	0	0	5	0	0	0
2	8	0	0	0	8	0	0	0
3	9	7	0	0	8	7	0	0
4	8	7	3	0	8	7	2	0
5	9	8	7	0	9	8	7	0
6	9	8	7	0	9	8	7	0
7	9	8	8	0	9	8	8	0
8	9	9	8	0	9	8	8	0
9	9	9	8	0	9	9	8	0
10	9	9	8	0	9	9	8	0
11	9	9	8	0	9	9	8	0
12	9	9	9	0	9	9	9	0
Total	103	83	66	0	101	82	65	0

Table 4. Distribution of vehicles across each parking slot according to scenario (c).

Row Parking Slot	A	B	C	D	E	F	G	H
1	0	3	8	5	0	1	8	0
2	0	7	9	7	0	7	8	0
3	0	9	9	7	0	8	9	0
4	0	9	9	6	0	8	9	0
5	7	9	9	0	0	8	9	0
6	7	9	9	0	7	9	8	0
7	8	9	9	0	7	9	8	0
8	9	9	9	0	8	8	7	0
9	9	9	8	0	8	8	7	0
10	9	9	7	0	8	8	0	0
11	9	8	7	0	9	7	0	0
12	9	8	0	0	9	7	0	0
Total	67	98	93	25	56	88	73	0

The data presented in Table 4 indicate a highly uneven distribution of space usage across the car park. A total of 27 parking spaces were each used by nine vehicles, while 18 others accommodated eight vehicles. At the opposite end of the spectrum, 32 parking slots were never utilized throughout the simulation. This pronounced disparity in allocation is largely attributable to the dual objective of minimizing both the in-parking driving distance and the walking distance to the designated point of interest. Consequently, the allocation algorithm favored parking slots located near both the entrance and the predefined destinations.

The simulation was extended to cover multiple vehicle parking time intervals across the three adopted study scenarios. It was assumed that the randomly generated vehicle parking time falls within the following ranges: $T_1 = [200, 250]$ s, $T_2 = [250, 300]$ s, $T_3 = [300, 350]$ s, and $T_4 = [350, 380]$ s. The results concerning the usage of individual parking slots and parking rows, the minimum and maximum space utilization rates, as well as the peak instantaneous occupancy of the car park, are summarized in Table 5.

Table 5. Simulation results for the three analytical scenarios and various vehicle parking times within the car park.

Parking Time	Scenario	Row								Parking Space Utilization		Max. Instant. Car Park Use	
		A	B	C	D	E	F	G	H	Min	Max	Val.	%
T1	a	62	65	59	68	63	59	64	60	2	8	59	61.45
	b	103	83	66	0	101	82	65	0	0	9	57	59.38
	c	67	98	93	25	56	88	73	0	0	9	57	59.38
T2	a	62	63	63	62	62	67	62	59	4	7	71	73.96
	b	89	79	69	17	86	77	68	15	0	8	69	71.88
	c	63	86	85	37	57	82	73	17	0	8	71	73.96
T3	a	61	64	60	61	63	64	63	64	4	6	84	87.50
	b	75	71	69	38	76	71	68	32	0	7	81	84.38
	c	65	72	75	48	63	72	70	65	0	7	82	85.46
T4	a	65	62	61	61	65	62	62	62	5	6	93	96.88
	b	66	64	63	62	62	61	61	61	1	6	92	95.83
	c	65	68	67	59	63	63	62	53	1	6	91	94.80

As shown in Table 5, for scenario (a), which does not impose any allocation constraints, the use of parking slots becomes progressively more evenly distributed as the vehicle parking time increases. This uniform exploitation presents potential advantages for car park operators, primarily by promoting even wear of the pavement surface, which may lead to less frequent and more cost-effective maintenance interventions over the long term. In contrast, once specific allocation criteria are introduced (as in scenarios b and c), the spatial distribution of parking activity becomes increasingly uneven. This is a direct result of the prioritization logic embedded in the allocation algorithm, which favors certain parking slots over others based on defined optimization objectives. Nonetheless, such non-uniform patterns of use may still prove beneficial operationally: since the wearing and degradation of the driving surface would occur unevenly across the facility, maintenance activities could be localized to the most affected areas. This targeted approach would reduce the need for full facility closures and allow unaffected areas to remain in operation. Furthermore, the effectiveness of these allocation strategies is particularly evident under moderate demand conditions—interpreted here as shorter parking times. Notably, as the average parking time increases, the discrepancy between minimum and maximum usage levels across parking slots tends to diminish, suggesting that longer parking times contribute to a more balanced distribution of use over time.

Figure 4 illustrates the simulation results obtained under the scenario of random vehicle parking, with parking times in the range $T1 = [200, 250]$ s. This scenario was selected because it features the shortest parking times, resulting in lower peak demand at any given moment. Consequently, the system can accommodate all 500 vehicles without exceeding capacity. The car park operates continuously, without interruptions.

To provide a clear visualization of the simulation results, a three-dimensional representation was chosen. On the horizontal axis, the parking rows (labeled A to H, as shown in Figure 1) are plotted for the 500 vehicles that received a parking space during the simulation interval. The minimum, maximum, and utilization percentage values are also shown. The vertical axis represents the magnitude of these values according to the scale on the left side of the figure. The axis perpendicular to this plane represents the three simulation scenarios defined in Section 3, Materials and Methods.

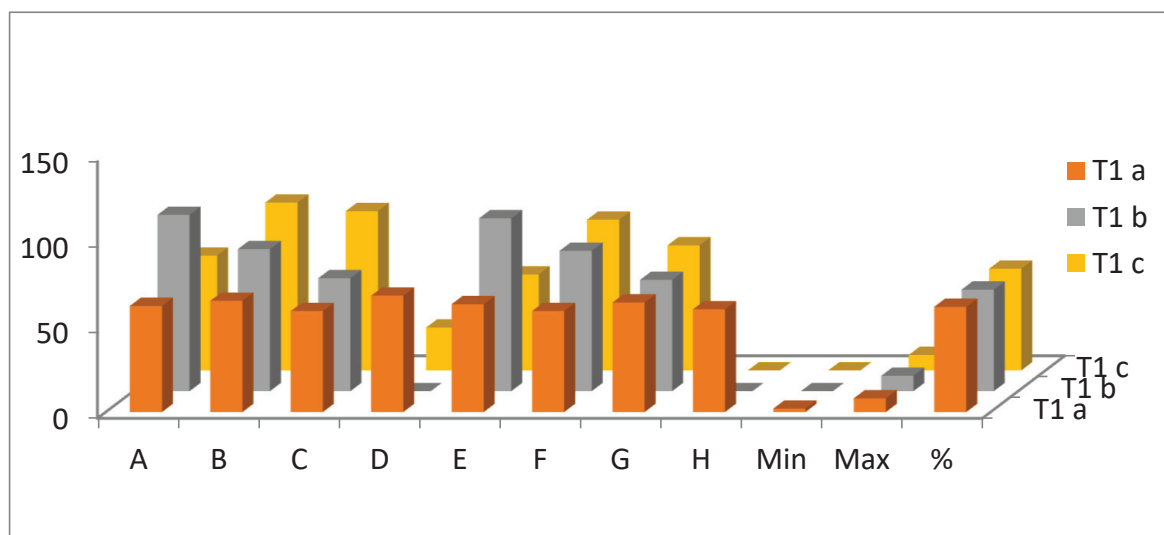


Figure 4. Simulation results for the three scenarios, $T1 = [200, 250]$ s.

Through this three-dimensional representation, all three scenarios for the selected parking duration can be visualized simultaneously.

The degree of numerical dispersion in parking slot allocation around the mean value, based on the data presented in Table 5, is assessed through the calculation of variance and standard deviation for the three analytical scenarios. The resulting values are presented in Table 6.

Table 6. The standard deviation and variance of the values resulting from parking slot allocation.

Parking Time	Scenario	Row								Average	Variance	Standard Deviation
		A	B	C	D	E	F	G	H			
T1	a	62	65	59	68	63	59	64	60	62.64	8.90	2.98
	b	103	83	66	0	101	82	65	0	86.01	219.23	14.81
	c	67	98	93	25	56	88	73	0	79.15	351.04	18.74
T2	a	62	63	63	62	62	67	62	59	62.57	4.34	2.08
	b	89	79	69	17	86	77	68	15	74.77	292.80	17.11
	c	63	86	85	37	57	82	73	17	71.10	307.53	17.54
T3	a	61	64	60	61	63	64	63	64	62.54	2.22	1.49
	b	75	71	69	38	76	71	68	32	66.67	169.35	13.01
	c	65	72	75	48	63	72	70	65	71.23	72.96	8.54
T4	a	65	62	61	61	65	62	62	62	62.54	2.30	1.52
	b	66	64	63	62	62	61	61	61	62.54	2.82	1.68
	c	65	68	67	59	63	63	62	53	62.82	18.64	4.32

An analysis of the values presented in Table 6 reveals that both the variance and the standard deviation are significantly lower in the scenario without any parking constraints. These values increase as allocation restrictions are introduced. However, they tend to decrease again as the overall occupancy level of the car park rises. This is because, under high congestion, the capacity for controlled allocation—based on predefined criteria—diminishes, leading to an almost uniform distribution of parked vehicles.

To determine whether there is any correlation between the number of allocations of a parking slot and the criteria used for space assignment, two correlation coefficients were computed, as described in Chapter 2—Materials and Methods: Pearson’s correlation coefficient and Spearman’s rank correlation coefficient. The results of the analysis are presented in Table 7.

Table 7. Pearson and Spearman correlation coefficients for allocation criteria (b) and (c).

Car Park Occupancy [%]	Correlation (b)			Correlation (c)				Total
	Pearson Driving Distance	Spearman Driving Distance	Driving Distance	Pearson Walking Distance	Total	Driving Distance	Spearman Walking Distance	
60	0.795	0.791	0.772	0.802	0.941	0.841	0.887	0.695
72	0.712	0.513	0.746	0.721	0.946	0.663	0.753	0.439
86	0.424	0.186	0.526	0.557	0.833	0.465	0.451	0.125
95	−0.169	0.051	−0.331	−0.024	−0.738	0.005	0.387	−0.344

An analysis of the values presented in Table 7 reveals a strong correlation between parking slot occupancy and the allocation criteria under conditions of lower parking demand. As demand increases, the strength of this correlation diminishes. By the time occupancy reaches 95%, certain values even exhibit a negative correlation. This indicates that, as the utilization of a parking slot increases, the allocation criterion is no longer followed; on the contrary, the relationship between the two variables becomes inversely proportional.

5. Discussion

The simulation outcomes provide valuable insights into the functioning of a parking facility under both unconstrained conditions and cases involving allocation criteria. An analysis of the distribution of parking slot usage indicates that, when no allocation rules are applied, parking slots tend to be used in a quasi-uniform manner, with each slot being occupied multiple times. This pattern becomes more pronounced as the intensity of car park usage increases, which was modelled by increasing the duration of vehicle parking. The introduction of allocation criteria—based either on minimizing the in-parking driving distance or the walking distance to the final destination—predictably results in a non-uniform usage pattern, prioritizing spaces that minimize the respective objectives. In certain cases, particularly under Cases *T1*, *T2*, and *T3* (representing shorter parking times), some parking slots remain completely unused, regardless of overall demand levels. In Case *T4*, which involves the longest parking times, the distribution of allocated spaces begins to rebalance. This shift occurs because high occupancy levels limit the program’s capacity to enforce the predefined allocation criteria.

For the parking time span *T1* (200–250 s), it is observed that the maximum instantaneous utilization rate of the entire parking lot ranges between 61.45% and 59.38%, with the higher value corresponding to the case without restrictions on space allocation. This outcome reflects practical reality, as users are generally assigned spaces without imposed restrictions. When allocation restrictions are introduced, utilization becomes more dispersed (min. 0 and max. 9), and the maximum instantaneous utilization rate decreases to 59.38%. As the parking time span increases and reaches *T4* (350–380 s), the maximum instantaneous utilization rate ranges between 96.88% and 54.80%. The higher value again corresponds to the unrestricted allocation scenario, while the lower value is associated with the simultaneous application of two allocation criteria—namely, the minimum driving distance to the assigned parking space and the minimum walking distance to the point of interest. These findings may support parking managers in making decisions aimed at balancing and optimizing parking utilization through the implementation of differentiated pricing schemes based on user-selected allocation criteria.

The analysis of standard deviation and variance supports this trend: these values are lowest in the unconstrained case, increase under constrained allocation, and then decrease again as congestion rises. This reflects the practical reality that, as congestion increases,

the ability to allocate spaces based on strict criteria diminishes, leading to a more uniform distribution of usage across the car park.

The lowest values of the standard deviation are obtained for time spans $T3$ and $T4$, in the absence of allocation restrictions, with values of 1.49 and 1.52, respectively. These results indicate a strong clustering of values around the mean. When correlated with the maximum instantaneous utilization rate (which is very high for $T3$ and $T4$), it can be concluded that as the parking duration increases, the possibilities for allocation according to the predefined criteria are reduced, and allocation becomes quasi-uniform. Thus, an optimal parking interval can be determined, for which the allocation criteria can still be effectively applied. In the case study, this value ranges between 250 and 300 s. The optimal value can be easily obtained by running the simulation program with alternative input parameters specified by the user.

The maximum instantaneous car park use can reach levels as high as 97%, which is considered extremely high for the queueing system. This phenomenon was also identified through the computation of the Pearson and Spearman correlation coefficients. These indicate a direct linear dependency under relatively moderate levels of demand but reveal an inverse relationship when the parking facility becomes highly saturated—an observation that aligns with patterns commonly encountered in real-world scenarios.

The Pearson and Spearman correlation analyses performed for the number of allocations of a parking space with respect to allocation criteria (b) and (c) indicate that, at a 60% occupancy level, the highest Pearson coefficient is 0.941. This corresponds to the allocation criterion combining the minimum driving distance to the assigned parking space with the minimum walking distance to the point of interest, suggesting a direct linear relationship between the number of allocations and the chosen allocation criterion. The highest Spearman coefficient is 0.887, corresponding to the correlation between the number of allocations and allocation criterion (c), defined as the walking distance from the parking space to the point of interest. The difference between the two maximum values is only 0.054, reflecting an exceedingly small discrepancy and thus a strong concordance between the variations of the two analyzed variables. At the opposite extreme, for high occupancy levels (95%), the correlation between the number of allocations and the allocation criteria reaches a maximum value of -0.738 , which indicates an inverse correlation between the analyzed measures, i.e., non-compliance with the allocation criteria.

This application provides a robust tool for the implementation of effective car park management and has the potential to significantly reduce the number of parking slots required within a specific area, while offering a variety of economic, social, and environmental benefits, as also demonstrated in the work of [6].

The application can offer financial benefits to car park operators by enabling the implementation of differentiated pricing strategies, based on the ability to allocate preferential parking slots relative to key points of interest served by the facility, or according to the current level of demand. Consequently, such parking management and administration systems provide both financial and logistical advantages, as also highlighted in the work of [24].

The data obtained may assist parking managers in improving financial and operational parameters to maximize profit and enhance customer satisfaction. Financial benefits can be achieved through reductions in maintenance and operational costs, allocation of spaces according to user-selected criteria, implementation of differentiated tariffs based on satisfaction levels, and energy savings derived from reduced pollution within the parking facility. Shorter driving distances inside the parking lot not only decrease pollution but also reduce fuel consumption and physical effort for users, as assigned spaces are closer to their destinations.

The proposed simulation model is of general applicability and can be adapted to any parking facility, regardless of its size, orientation of parking spaces, or distribution of destinations and points of interest. All input variables can be modified quickly and easily within the model, yielding corresponding results and efficiency indicators of parking utilization.

The simulation model was also tested for smaller parking facilities, but the results did not produce significant benefits or noteworthy advantages; thus, they were not included in this work. Since the allocation criteria are defined in terms of travel distances (either the distance driven through the parking lot or the sum of driving and walking distances to the destination), the operational indicators become most relevant in parking facilities where such distances are considerable. This consideration was the rationale for selecting, in the case study, a parking lot measuring 1000 m in length and 510 m in width. These dimensions were derived from a design created using AutoCAD Vehicle Tracking v25.00.2775 for 96 available spaces. A symmetric parking layout was chosen for testing to identify different patterns of functional similarity. The main limitation of this simulation model lies in the fact that the results obtained for parking management are significant primarily in large-scale parking facilities with substantial internal distances to be traveled.

This application can be improved by incorporating functionalities for visualizing vehicle movement within the parking area, including the detection of potential conflict points. Further enhancements may also involve integration with external traffic monitoring systems, mobile applications, or IoT platforms that capture users' declared preferences. In an initial stage, such data could be collected at the parking entrance, where users specify their preferred parking location or final destination.

6. Conclusions

This study presented an information system for the management and operation of a parking facility, aimed at maximizing space utilization, reducing search times, enabling preferential allocation, and promoting the rational use of the available area. The software application, developed entirely by the research team, is original and was designed with the objectives of ease of generalization and universal applicability, regardless of the surface area or geometric layout of the parking lot. For modeling and simulation purposes, a parking facility with 96 spaces was considered; however, the number of spaces and the lot size can be easily scaled due to the use of AutoCAD Vehicle Tracking v25.00.2775. The operation of the facility was simulated using a custom-developed Python program, which assigns parking spaces to vehicles based on predefined allocation criteria that aim to optimize utilization while accommodating user preferences. Operational indicators were also evaluated to support the development of marketing policies by parking administrators. The results indicate that the proposed model enhances parking operations and can be adjusted to meet different management objectives—whether to maximize space efficiency or financial revenue, through differentiated pricing based on ease of access or user-specific constraints. These constraints include walking distance from the allocated parking space to the destination and user-defined preferences. The calculated performance indicators include average parking time per space and for the entire facility, individual space occupancy rates within a given time interval, and the average vehicle travel distance required to reach the assigned parking slot. The model was tested under a simulated demand of 500 randomly generated vehicles with randomly varied parking durations constrained within defined time intervals.

The simulation model employs as input data the arrival times of vehicles, generated randomly, as well as the durations of parking occupancy, also randomly generated within

various time intervals. For the study and analysis, four intervals of parking duration were considered, denoted T_1 , T_2 , T_3 , and T_4 , with values ranging between 200 s and 380 s.

In practical application, the model will rely on real input data: vehicle state information and user-declared preferences collected via mobile IoT systems or in-vehicle embedded systems, which will communicate with fixed parking management systems. Feedback on user behavior will be gathered through automated vehicle transmission systems, whereby the parking management system will receive information on the occupied space, parking duration, the distance traveled within the facility, and the walking distance to the destination.

The system is compatible with real-time data input, either from entry barriers or from external traffic monitoring systems. Parking durations may also be captured directly from sensors installed at each space, detecting occupancy and vacancy.

We conclude that the automation and digitalization of parking systems can contribute indirectly to improved traffic flow by reducing the number of vehicles driving in search of parking, thus mitigating the negative externalities of road congestion.

Author Contributions: Conceptualization, V.D. and A.-B.M.; methodology, V.D.; software, A.-B.M. and M.A.R.; validation, O.M.D., E.A.R. and F.C.O.; formal analysis, V.D., O.M.D. and E.A.R.; investigation, F.C.O.; resources, O.M.D.; data curation, E.A.R.; writing—original draft preparation, V.D. and A.-B.M.; writing—review and editing, V.D., E.A.R. and O.M.D.; visualization, M.A.R.; supervision, V.D.; project administration, V.D.; funding acquisition, O.M.D. All authors have read and agreed to the published version of the manuscript.

Funding: The research work contained in this paper was supported within the framework of the grant no. 42/09.10.2023 from GNAC ARUT 2023.

Data Availability Statement: The data presented in this study is available on request from the corresponding author.

Acknowledgments: This work was supported by a grant from the National Program for Research of the National Association of Technical Universities—GNAC ARUT 2023.

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

References

1. Shaaban, K.; Tounsi, H. Parking detection method using quadtree decomposition analysis. *J. Traffic Transp. Eng. (Engl. Ed.)* **2022**, *9*, 645–653. [CrossRef]
2. Zhu, Y.; Ye, X.; Chen, J.; Yan, X.; Wang, T. Impact of cruising for parking on travel time of traffic flow. *Sustainability* **2020**, *12*, 3079. [CrossRef]
3. Samaranayake, P.; Gunawardana, U.; Stokoe, M. Kerbside parking assessment using a simulation modelling approach for infrastructure planning—A metropolitan city case study. *Sustainability* **2023**, *15*, 3301. [CrossRef]
4. Arellano-Verdejo, J.; Alonso-Pecina, F.; Alba, E.; Arenas, A.G. Optimal allocation of public parking spots in a smart city: Problem characterisation and first algorithms. *J. Exp. Theor. Artif. Intell.* **2019**, *31*, 575–597. [CrossRef]
5. Harpalete, O.L. Methods to Estimate Urban Accessibility in a Medium Size City. Case Study: Iasi, Romania. *U.P.B. Sci. Bull.* **2023**, *85*, 120–134. Available online: https://www.scientificbulletin.upb.ro/rev_docs_arhiva/full1ba_354364.pdf (accessed on 22 May 2025).
6. Karri, S.; Dhabu, M.M. Multistage Game Model Based Dynamic Pricing for Car Parking Slot to Control Congestion. *Sustainability* **2022**, *14*, 11808. [CrossRef]
7. Shoup, D.C. Cruising for parking. *Transp. Policy* **2006**, *13*, 479–486. [CrossRef]
8. Azari, K.A.; Arintono, S.; Hamid, H.; Rahmat, R.A.O. Modelling demand under parking and cordon pricing policy. *Transp. Policy* **2013**, *25*, 1–9. [CrossRef]
9. Mondal, M.A.; Rehena, Z.; Janssen, M. Smart parking management system with dynamic pricing. *J. Ambient. Intell. Smart. Environ.* **2021**, *13*, 473–494. [CrossRef]
10. Thompson, R.G.; Richardson, A.J. A parking search model. *Transp. Res. Part A Policy Pract.* **1998**, *32*, 159–170. [CrossRef]
11. Arnott, R. On the optimal target curbside parking occupancy rate. *Econ. Transp.* **2014**, *3*, 133–144. [CrossRef]

12. Qian, Z.; Xiao, F.; Zhang, H.M. Managing morning commute traffic with parking. *Transp. Res. Part B Methodol.* **2012**, *46*, 894–916. [CrossRef]
13. Geng, Y.; Cassandra, C.G. New “smart parking” system based on resource allocation and reservations. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1129–1139. [CrossRef]
14. Weinberger, R.R.; Adam Millard-Ball, A.; Hampshire, R.C. Parking search caused congestion: Where’s all the fuss? *Transp. Res. Part C Emerg. Technol.* **2020**, *120*, 102781. [CrossRef]
15. TomTom Traffic Index. Available online: <https://www.tomtom.com/traffic-index/ranking/> (accessed on 10 May 2025).
16. Naphade, M.; Banavar, G.; Harrison, C.; Paraszczak, J.; Morris, R. Smarter cities and their innovation challenges. *Computer* **2011**, *44*, 32–39. [CrossRef]
17. Biyik, C.; Allam, Z.; Pieri, G.; Moroni, D.; O’Fraifer, M.; O’Connell, E.; Olariu, S.; Khalid, M. Smart parking systems: Reviewing the literature, architecture and ways forward. *Smart Cities* **2021**, *4*, 623–642. [CrossRef]
18. Atiqur, R. Smart car parking system model for urban areas. *Comput. Sci. Inf. Technol.* **2021**, *2*, 95–102. [CrossRef]
19. Babic, M.; Vekj, A.; Stanojevic, M.; Ostojic, G.; Borocki, J.; Stankovski, S. Modern parking solutions for smart cities. In Proceedings of the 30th DAAAM International Symposium on Intelligent Manufacturing and Automation, Zadar, Croatia, 23–26 October 2019.
20. Parmar, J.; Das, P.; Dave, S.M. Study on demand and characteristics of parking system in urban areas: A review. *J. Traffic Transp. Eng.* **2020**, *7*, 111–124. [CrossRef]
21. Borgonovo, E.; Gatti, S.; Peccati, L. What drives value creation in investment projects? An application of sensitivity analysis to project finance transactions. *Eur. J. Oper. Res.* **2010**, *205*, 227–236. [CrossRef]
22. Channamallu, S.S.; Kermanshachi, S.; Rosenberger, J.M.; Pamidimukkala, A. A review of smart parking systems. *Transp. Res. Proc.* **2023**, *73*, 289–296. [CrossRef]
23. Kuran, M.S.; Viana, A.C.; Iannone, L.; Kofman, D.; Mermoud, G.; Vasseur, J.P. A smart parking lot management system for scheduling the recharging of electric vehicles. *IEEE Trans. Smart Grid* **2015**, *6*, 2942–2953. [CrossRef]
24. Sajeev, A.; Vidwans, S.; Mallick, C.; Jog, Y. Understanding smart and automated parking technology. *Int. J. Serv. Sci. Technol.* **2015**, *8*, 251–262.
25. Shoup, D. *The High Cost of Free Parking, 2011 Updated Edition*; reprinted by Routledge: London, UK, 2017.
26. Kattimani, R. Appraisal of Multilevel Car Parking Facility at KG Road–CBD Area, Bengaluru City. *Int. J. Res. Eng. Technol.* **2014**, *3*, 404–410. [CrossRef]
27. Rodríguez, A.; dell’Olio, L.; Moura, J.L.; Alonso, B.; Cordera, R. Modelling Parking Choice Behaviour Considering Alternative Availability and Systematic and Random Variations in User Tastes. *Sustainability* **2023**, *15*, 8618. [CrossRef]
28. Hess, S.; Polak, J.W. Mixed Logit Modelling of Airport Choice in Multi-Airport Regions. *J. Air Transp. Manag.* **2005**, *11*, 59–68. [CrossRef]
29. Cao, Y.; Ren, Y.; Jia, H.; Sun, M.; Dali, Z. Modeling Parking Choice Behavior Using Cumulative Prospect Theory. *Sustainability* **2024**, *16*, 1596. [CrossRef]
30. Fabusuyi, T.; Hampshire, R.C.; Hill, V. Evaluation of a Smart Parking System. *Transp. Res. Rec.* **2013**, *2359*, 10–16. [CrossRef]
31. Kelly, J.A.; Clinch, J.P. Influence of Varied Parking Tariffs on Parking Occupancy Levels by Trip Purpose. *Transp. Policy* **2006**, *13*, 487–495. [CrossRef]
32. Stover, V.G.; Koepke, F.J. *Parking Design, Transportation and Land Development*, 2nd ed.; Institute of Transportation Engineers: Washington, DC, USA, 2002.
33. Weinberger, R.R. Death by a Thousand Curb Cuts: Evidence on the Effect of Minimum Parking Requirements on the Choice to Drive. *Transp. Policy* **2012**, *20*, 93–102. [CrossRef]
34. Marsden, G.R. The Evidence Base for Parking Policies—A Review. *Transp. Policy* **2006**, *13*, 447–457. [CrossRef]
35. Jakobsson, C.; Fujii, S.; Gärling, T. Determinants of private car users’ acceptance of road pricing. *Transp. Policy* **2000**, *7*, 153–158. [CrossRef]
36. Arnott, R.; Inci, E. An Integrated Model of Downtown Parking and Traffic Congestion. *J. Urban Econ.* **2006**, *60*, 418–442. [CrossRef]
37. Litman, T. *Parking Management Best Practices, Updated ed.*; Victoria Transport Policy Institute: Victoria, BC, Canada, 2024. [CrossRef]
38. Singh, G.; Jyoti, D.; Kumar, R. DyPARK: A Dynamic Pricing and Allocation Scheme for Smart On-Street Parking System. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4217–4234. [CrossRef]
39. Thomas, D.; Kovoov, B.C. A Genetic Algorithm Approach to Autonomous Smart Vehicle Parking System. *Procedia Comput. Sci.* **2018**, *125*, 68–76. [CrossRef]
40. Elfaki, A.O.; Messoudi, W.; Bushnag, A.; Abuzneid, S.; Alhmiedat, T. Constraint Optimization Model for Dynamic Parking Space Allocation. *Sensors* **2024**, *24*, 3988. [CrossRef]
41. Mladenović, M.; Delot, T.; Laporte, G.; Wilbaut, C. A Scalable Dynamic Parking Allocation Framework. *Comput. Oper. Res.* **2021**, *125*, 105080. [CrossRef]
42. Nakazato, T.; Fujimaki, Y.; Namerikawa, T. Parking Lot Allocation Using Rematching and Dynamic Parking Fee Design. *IEEE Trans. Control Network Syst.* **2022**, *9*, 1692–1703. [CrossRef]

43. Sahu, D.; Sinha, P.; Prakash, S.; Yang, T.; Rathore, R.S.; Wang, L. A Multi-Objective Optimization Framework for Smart Parking Using Digital Twin, Pareto Front, MDP and PSO for Smart Cities. *Sci. Rep.* **2025**, *15*, 7783. [CrossRef]
44. Yang, T.; Zhao, L.; Huang, M. Empirical Research on Residents' Acceptance of Reserved Parking. In Proceedings of the 2024 International Conference on Artificial Intelligence and Data Mining (ICAIDM 2024), Hangzhou, China, 20–22 April 2024; Atlantis Press: Dordrecht, The Netherlands. [CrossRef]
45. Chen, R.; Gao, G.; Kang, L.-J.; Zhang, L.-Y. Efficiency and Equity Analysis on Parking Reservation. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *187*, 103606. [CrossRef]
46. Xu, D.; Sun, H. Research on Parking Service Optimization Based on Permit Reservation and Allocation. *J. Adv. Transp.* **2022**, *2022*, 8364988. [CrossRef]
47. Panayiotou, T.; Mavrovouniotis, M.; Ellinas, G. On the Fair-Efficient Charging Scheduling of Electric Vehicles in Parking Structures. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; pp. 1627–1634. [CrossRef]
48. Alahmadi, S.; Hakeem, A.; Attiah, A.; Alghamdi, B.; Mohaisen, L. Towards Scalable and Privacy Enhanced On-Street Parking Management: A Roadmap for Future Inquiry. *Electronics* **2023**, *12*, 4160. [CrossRef]
49. İnaç, H. Efficient On-Street Parking Management: Priority Parking Permits for Residents Integrated with Digital Systems—Case of Istanbul. *J. Intell. Transp. Syst. Appl.* **2025**, *8*, 256. [CrossRef]
50. Ebrahimi, M.; Ebrahimi, M.; Shafie-khah, M.; Laaksonen, H. The Role of EV Parking Lots for Supporting the Distribution System Operation Considering EV Uncertainties. *Transp. Res. Procedia* **2023**, *70*, 263–270. [CrossRef]
51. Drašković, D.; Vukićević, S. A Model of Software System for Parking Using Search Algorithms. In Proceedings of the 35th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2012), Opatija, Croatia, 21–25 May 2012.
52. Wang, X.; Guo, Z.; Hou, J.; Li, T. Optimal Allocation of Private Parking Spaces in Shared Parking Systems Considering EV Charging: A Lagrangian Relaxation Based Decomposition Approach. *Comput. Ind. Eng.* **2025**, *205*, 111170. [CrossRef]
53. Burns, A.; Forsythe, C.R.; Michalek, J.J.; Whitefoot, K. Estimating the Potential for Dynamic Parking Reservation Systems to Increase Delivery Vehicle Accommodation. *Transp. Res. Part A Policy Pract.* **2025**, *193*, 104380. [CrossRef]
54. Renuka, R.; Dhanalakshmi, S. Android Based Smart Parking System Using Slot Allocation and Reservations. *ARPJ. Eng. Appl. Sci.* **2015**, *10*, 3116–3120. Available online: https://arpnjournals.com/jeas/research_papers/rp_2015/jeas_0415_1892.pdf (accessed on 5 June 2025).
55. Aswath, S.; Kishore, N.; Thinakaran, S.K.; Vishwaa, M.; Hariharan, A.; Vasavi, C.S.; Karthikeyan, B. Smart Car Parking System with Online Reservation. *Procedia Comput. Sci.* **2025**, *258*, 2777–2786. [CrossRef]
56. Zhang, X.; Pitera, K.; Wang, Y. Parking Reservation Techniques: A Review of Research Topics, Considerations, and Optimization Methods. *J. Traffic Transp. Eng. (Engl. Ed.)* **2023**, *10*, 1099–1117. [CrossRef]
57. Ismael, A.; Holguín-Veras, J. Optimal Parking Allocation for Heterogeneous Vehicle Types. *Transp. Res. Part A Policy Pract.* **2025**, *192*, 104357. [CrossRef]
58. Mikušová, M.; Abdunazarov, J.; Żukowska, J.; Usmankulov, A. Designing of Parking Spaces Taking into Account the Parameters of Design Vehicles in Russia. *Computation* **2020**, *8*, 71. [CrossRef]
59. Maina, R.M.; Kiprop, A.K.; Muriithi, C.M. A mathematical model for optimization of vehicle parking space. *OPSEARCH* **2025**, *Online First*, 1–18. [CrossRef]
60. Wang, Q.; Zhang, Y.; Zhang, Y. Optimization Design of Parking Lot Layout—Taking Panzhuhua Wanda Outdoor Parking Lot as an Example. In Proceedings of the International Conference on Smart Transportation and City Engineering (STCE 2023); SPIE: Bellingham, WA, USA, 2024; Volume 13018, p. 130184C. [CrossRef]
61. Kong, Y.; Le Vine, S.; Liu, X. Capacity Impacts and Optimal Geometry of Automated Cars' Surface Parking Facilities. *J. Adv. Transp.* **2018**, *2018*, 6908717. [CrossRef]
62. Chen, S.; Sun, Q. An Optimization Design of Hybrid Parking Lots in an Automated Environment. *Sustainability* **2023**, *15*, 15475. [CrossRef]
63. *Design Standards for Car Parks—Code Reference NP 24–2022*; Ministry of Development, Public Works and Administration of Romania; Part I, No. 110 bis/2023; Official Gazette of Romania: Bucharest, Romania, 2023.
64. Târcolea, C.; Paris, A.S.; Sylvan, D. Modeling Dependencies in Bivariate Distributions. *U.P.B. Sci. Bull. Series A* **2020**, *82*, 150–156. Available online: https://www.scientificbulletin.upb.ro/rev_docs_arhiva/full902_166350.pdf (accessed on 2 July 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

A CNN-LSTM-GRU Hybrid Model for Spatiotemporal Highway Traffic Flow Prediction

Jinsong Zhang ¹, Junyi Sha ¹, Chunyu Zhang ¹ and Yijin Zhang ^{2,*}

¹ School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, China; jinsong_zhang@dlmu.edu.cn (J.Z.); junyi_sha@163.com (J.S.); qingqian9962022@163.com (C.Z.)

² School of Economics and Management, Dalian Minzu University, Dalian 116650, China

* Correspondence: zhangyijin@dlmu.edu.cn

Abstract: The rapid growth in the number of motor vehicles has exacerbated traffic congestion. The occurrence of congestion not only poses significant challenges for traffic management authorities but also severely impacts residents' travel and daily routines. Against this backdrop, predicting traffic flow can provide crucial insights for anticipating changing traffic patterns. Therefore, this paper proposes a novel hybrid deep learning architecture (CNN-LSTM-GRU) for highway traffic flow prediction that integrates spatiotemporal and meteorological dimensions. Our approach constructs a multidimensional feature matrix encompassing temporal sequences, spatial correlations, and weather conditions. Convolutional Neural Networks (CNN) are employed to capture spatial patterns, while Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks jointly model temporal dependencies. Through systematic hyperparameter tuning and step-length optimization, we validate the model using real-world traffic data from a provincial highway network. The experimental evaluation analyzes the following two critical dimensions: (1) holiday vs. non-holiday traffic patterns, and (2) the impact of weather data integration. Comparative analysis reveals that our hybrid model demonstrates superior prediction accuracy over standalone LSTM, GRU, and their CNN-based counterparts (CNN-LSTM, CNN-GRU).

Keywords: highways; traffic flow prediction; CNN-LSTM-GRU

1. Introduction

The development of highway infrastructure constitutes a pivotal component in national modernization efforts and serves as the backbone of contemporary transportation networks. Amidst the exponential growth in motor vehicle ownership, highway congestion has emerged as a pressing challenge, with traffic volumes having reached critical thresholds during holiday periods. This phenomenon not only undermines public mobility efficiency but also constrains service quality enhancement, thereby impeding regional economic integration and societal advancement.

Traffic flow dynamics exhibit complex spatiotemporal patterns influenced by multifactorial factors. Crucially, holiday periods and meteorological conditions have been identified as primary determinants of traffic volatility, resulting in nonlinear relationships that traditional prediction models struggle to capture. In response to these challenges, this paper proposes a novel framework for highway traffic prediction that integrates key contextual factors, such as holiday-induced travel patterns and weather conditions, into multidimensional analytics.

In recent years, with the rapid development of deep learning algorithms, an increasing number of models have been employed for traffic flow prediction research. To enhance

model accuracy, this paper proposes a hybrid CNN-LSTM-GRU model. This model extracts spatial features from data using Convolutional Neural Networks (CNN) and subsequently captures temporal features through Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. The advantages of this approach are twofold. (1) Spatial Feature Enhancement: By constructing two-dimensional feature matrices of highway network topologies, the CNN effectively captures spatial correlations between adjacent road segments, addressing the limitation of traditional time-series models that often overlook spatial dependencies. (2) Temporal Feature Synergy: The dual-path architecture combining LSTM and GRU preserves LSTM's strength in long-term memory retention while leveraging the GRU's computational efficiency, creating a complementary mechanism for temporal sequence analysis. This integrated framework enables more accurate traffic flow prediction by simultaneously modeling spatial dependencies and temporal dynamics, offering a robust solution for modern highway management challenges.

2. Literature Review

Currently, research on short-term traffic flow prediction is diverse, and with the continuous advancement of technology, prediction levels and accuracy are also improving. This section summarizes traffic flow prediction methods based on a review of the existing literature.

Firstly, the most common methods are those based on mathematical statistics, which include both linear and nonlinear models. Linear models include Kalman filtering and historical average models, while nonlinear methods include non-parametric regression models and chaos theory. Guo et al. [1] proposed using Kalman filtering to achieve real-time data processing capabilities, employing process variance to handle real-world data. Wang and Papageorgiou [2] used a macroscopic traffic flow model along with the measurement model, which is designed by use of the extended Kalman-filtering method. Some other researches relied on time series analysis, such as autoregressive integrated moving average models (ARIMA) and their variants (SARIMA), to capture periodic patterns through linear assumptions. The ARIMA method was first proposed in 1979, by Ahmed et al. [3] and applied to traffic flow prediction research. Smith et al. [4] conducted an investigation into the theoretical underpinnings of nonparametric regression, with the dual objectives of establishing its methodological framework and evaluating whether heuristically optimized forecast generation techniques in nonparametric regression could attain traffic flow prediction accuracy comparable to seasonal ARIMA models in single-interval forecasting scenarios. Kumar et al. [5] developed a predictive framework utilizing the Seasonal Autoregressive Integrated Moving Average (SARIMA) model for short-term traffic flow forecasting, achieving reliable performance even when constrained by limited input data. However, this type of method has weak adaptability to nonlinear and unexpected events, and it ignores spatial correlation. For the nonlinearity of complicated urban traffic flow, Chen and Xiao [6] proposed a switching autoregressive integrated moving average (ARIMA) model and employed it to explore how traffic flow varies with time.

There are also many studies that use machine learning algorithms to conduct traffic flow prediction research. Algorithms such as Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbor (k-NN) enhance predictive ability through nonlinear mapping. Zhang et al. [7] developed a high-precision multi-step traffic flow prediction model using Support Vector Machine (SVM) methodology. The framework incorporates actual traffic volume data as input vectors, with a systematic comparison of four distinct input vector configurations conducted to evaluate their comparative predictive performance. Hou et al. [8] developed four traffic flow forecasting models—Random Forest, regression tree, multilayer feed-forward neural network, and nonparametric regression—specifically

for planned work zone events. Concurrently, Zhang et al. [9] proposed a short-term traffic flow prediction method using k-Nearest Neighbor (KNN)-based nonparametric regression, systematically analyzing how key parameter configurations influence model performance. Building on these efforts, Lu et al. [10] introduced a hybrid traffic flow prediction framework combining signal decomposition and machine learning. The improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) was applied to decompose traffic sequences into multiple intrinsic mode functions (IMFs), which were then processed by machine learning algorithms for enhanced predictive accuracy. Nevertheless, traditional machine learning approaches often necessitate extensive manual feature engineering and demonstrate limited capacity in processing high-dimensional spatiotemporal datasets.

In recent years, deep learning has garnered significant attention and has been widely applied to the research of traffic flow prediction. The application of foundational models, such as CNN, enables the effective utilization of spatial convolution to capture topological relationships in road networks. For instance, Zhang et al. [11] proposed ST-ResNet, which models the spatiotemporal correlation of urban traffic flow through residual CNN. LSTM, with its gating mechanism addressing long-range dependencies, has become a benchmark for processing temporal data. Ma et al. [12] were the first to apply LSTM to traffic prediction, demonstrating its superiority over traditional methods. Additionally, numerous studies have integrated various models for traffic flow prediction. The spatiotemporal fusion architecture, constructed by combining CNN and LSTM [13], extracted the temporal dependencies of spatial features and models. Mackenzie et al. [14] used HTM (Hierarchical Temporal Memory) to integrate and evaluate data from the adaptive traffic system on the main roads of Sydney and Adelaide. The results showed that HTM achieved comparable prediction outcomes to the LSTM model. Tian et al. [15] proposed using multi-scale temporal smoothing to infer missing data and conducted comparisons on the PEMS dataset and their own dataset. Experiments demonstrated that this method achieved high accuracy. Feng et al. [16] proposed the AMSVM-STC model based on support vector machines to accurately predict traffic congestion. They used an adaptive particle swarm optimization algorithm to find optimal parameters. The validation of datasets showed that, even during periods of abnormal traffic flow fluctuations, the model could provide accurate and timely predictions, outperforming the results of other mentioned models. This has further evolved into a combination of 3D convolution (C3D) and Transformer [17], enhancing multi-scale feature extraction. For the non-Euclidean structure of road networks, Graph Neural Networks (GNNs), such as GCN and GAT, capture inter-node relationships through graph convolution. The DCRNN model proposed by Li et al. [18], which combines graph diffusion convolution with sequence modeling, has become a classic framework.

In the research on predicting traffic flow on highways, due to the periodicity of highway traffic flow, a certain section of road has a relatively stable flow and change trajectory at the same time period in different historical times. Therefore, historical time is considered an important entry point for predicting traffic flow. Fang et al. [19] added an attention mechanism to the LSTM model, effectively assigning different weights to different inputs of the model and focusing on filtering important information. After, four datasets were used as examples to prove that this model has good accuracy. Shuai et al. [20] decomposed the traffic flow components through SSA, predicted the traffic flow through LSTM and SVR, and finally combined their respective prediction results to obtain the final prediction results. Through the verification of the traffic flow data of the Guizhou expressway, it was concluded that the combined model based on component decomposition was superior to the single model. Bing et al. [21] proposed a multi-step prediction method consisting of VMD and LSTM. The VMD algorithm decomposes traffic flow into IMF

components, and each LSTM predicts one component before integrating the predictions. Xu et al. [22] proposed using the whale optimization algorithm to optimize the structure of the BiLSTM-Attention network, and using this algorithm to find the best parameters to input into the network structure, they formed the WOA_BiLSTM-Attention model.

At the same time, during the process of reading the literature, it was found through induction and summary that traffic flow prediction considering time and space has also formed quite rich results. Due to the high degree of closure of highways, the traffic flow of a certain section will be affected by the upper and lower nodes; thus, it is necessary to consider the spatial impact of highways. Zheng et al. [23] considered spatiotemporal correlation in the prediction of urban road network traffic flow, selected similar and target road segments as independent and dependent variables, and finally used them as inputs for CNN-LSTM traffic flow prediction. The final results showed that, compared with other models, this model has higher accuracy. Zheng et al. [24] proposed using an attention mechanism on the basis of Conv LSTM to solve the problem of poor spatial feature capture, and using multi-layer architecture for feature extraction to improve the capturing power of Bi LSTM on temporal features, they thereby improved the overall prediction accuracy of the model. Some combination models have also been proposed for short-term traffic flow prediction research, i.e., Bi-LSTM-CNN [25] and GTO-CNN-LSTM [26].

Traffic flow prediction is influenced by a multitude of factors. Beyond the inherent temporal and spatial characteristics of traffic flow itself, it is also affected by fundamental elements such as driver behavior and road conditions. These conventional factors can generally be categorized as routine influences, as their impact on traffic flow tends to remain relatively stable. This study, however, places particular emphasis on analyzing the effects of dynamic factors—including weather conditions, holidays, and other time-varying elements—on traffic flow. For instance, adverse weather such as rain or snow can render road surfaces slippery to varying degrees, compromising driver control, reducing vehicle speeds, and ultimately leading to congestion. Similarly, holidays often induce significant fluctuations in highway traffic flow, with notable surges observed before and after holiday periods. Hence, this research aims to integrate these dynamic factors (weather, holidays, etc.) with the intrinsic temporal and spatial attributes of traffic flow to develop a more comprehensive predictive framework.

In terms of prediction methods, the combination model can better utilize the advantages of each model and fully explore data features. Therefore, in this study, a combination model based on LSTM-GRU is constructed. CNN is used to better extract spatial features of the data and to complete the prediction.

3. Data Sources and Characteristics Analysis

3.1. Data Sources and Preprocessing

The dataset used in this study consists of traffic flow data and weather data collected from toll stations on a provincial highway from 19 July 2016 to 24 October 2016. The traffic flow data include entry and exit traffic data from three toll stations, labeled as Station 1, Station 2, and Station 3. Each toll station has two directions (0 and 1), where 0 represents the entrance and 1 represents the exit. Station 2 was only allowed to enter the highway due to maintenance, which is a one-way passage. The weather data span from 19 July 2016, 00:00, to 24 October 2016, 24:00, and includes the following seven fields: atmospheric pressure, sea-level pressure, wind direction, wind speed, temperature, humidity, and rainfall.

The raw traffic flow data were collected at a frequency of 20 min. Specifically, the entrance of toll Station 1 contains 2084 records, the exit of toll Station 1 contains 2084 records, the exit of toll Station 2 contains 1725 records, the entrance of toll Station 2 is 0 in this dataset,

the exit of toll Station 3 contains 2086 records, and the entrance of toll Station 3 contains 2085 records.

The weather data consist of 231 records collected at a frequency of every 3 h.

Preprocessing raw data is an essential and critical step in data mining, machine learning, and artificial intelligence. Data collected directly often lack quality assurance, with missing values being a common issue due to machine malfunctions or human errors. Missing data can lead to errors during feature extraction, thereby affecting prediction accuracy. Linear interpolation, the simplest form of algebraic interpolation, is widely used for handling nonlinear functions. Therefore, this study employs linear interpolation to fill in missing values.

The traffic flow data for direction 1-0 from 19 July 2016 to 24 October 2016 contain four missing values. The traffic flow data for direction 2-0 during the same period contain 362 missing values. The traffic flow data for direction 3-0 contain two missing values. The traffic flow data for direction 1-1 contain four missing values, and the traffic flow data for direction 3-1 contain three missing values. Below is a portion of the raw traffic flow data table for direction 1-0, which includes one missing value, as shown in Table 1. The missing value is filled using the mean imputation method. Since the data represent the number of vehicles passing through at 20 min intervals, the repaired data are rounded to 14.5. Due to the large number of missing values in direction 2-0, likely caused by detector malfunctions, linear interpolation is used for filling.

Table 1. The original data of the 1-0 station entrance direction somewhere missing value and its filling value processing.

Tollgate_ID	Time_Window	Direction	Volume
1	[2016-10-01 22:40:00, 2016-10-01 23:00:00]	0	41
1	[2016-10-01 23:00:00, 2016-10-01 23:20:00]	0	35
1	[2016-10-01 23:20:00, 2016-10-01 23:40:00]	0	16
1	[2016-10-01 23:40:00, 2016-10-02 00:00:00]	0	
1	[2016-10-02 00:00:00, 2016-10-02 00:20:00]	0	13
1	[2016-10-02 00:20:00, 2016-10-02 00:40:00]	0	7
1	[2016-10-02 00:40:00, 2016-10-02 01:00:00]	0	3

Since the weather data collection frequency differs from the traffic flow data, linear interpolation is applied to align the weather data frequency to 20 min, resulting in a total of 2088 weather data records. A portion of the filled weather data is shown in Table 2.

Table 2. Weather data linear interpolation filling result table.

Time_Window	Pressure (hPa)	Sea_Pressure (hPa)	Wind_Direction (°)	Wind_Speed (m/s)	Temperature (°C)	Rel_Humidity (%)	Precipitation (mm)
[2016-09-19 00:00:00, 2016-09-19 00:20:00]	1008.2	1013.2	329	2.8	22.2	76	0
[2016-09-19 00:20:00, 2016-09-19 00:40:00]	1008.213	1013.213	289.75	3.0125	22.475	74.875	0
[2016-09-19 00:40:00, 2016-09-19 01:00:00]	1008.225	1013.225	250.5	3.225	22.75	73.75	0
[2016-09-19 01:00:00, 2016-09-19 01:20:00]	1008.238	1013.238	211.25	3.4375	23.025	72.625	0
[2016-09-19 01:20:00, 2016-09-19 01:40:00]	1008.25	1013.25	172	3.65	23.3	71.5	0
[2016-09-19 01:40:00, 2016-09-19 02:00:00]	1008.263	1013.263	132.75	3.8625	23.575	70.375	0

Data visualization is crucial for in-depth analysis and trend prediction. By presenting traffic flow and weather data in graphical form, it helps uncover patterns over time and

space, explore correlations between the two, analyze long-term traffic demand trends, and provide scientific references for future traffic flow predictions and traffic management under extreme weather conditions.

Therefore, to better observe the characteristics of the data, visualizations of the features of traffic flow data and weather data are provided, as shown in Figure 1 below.

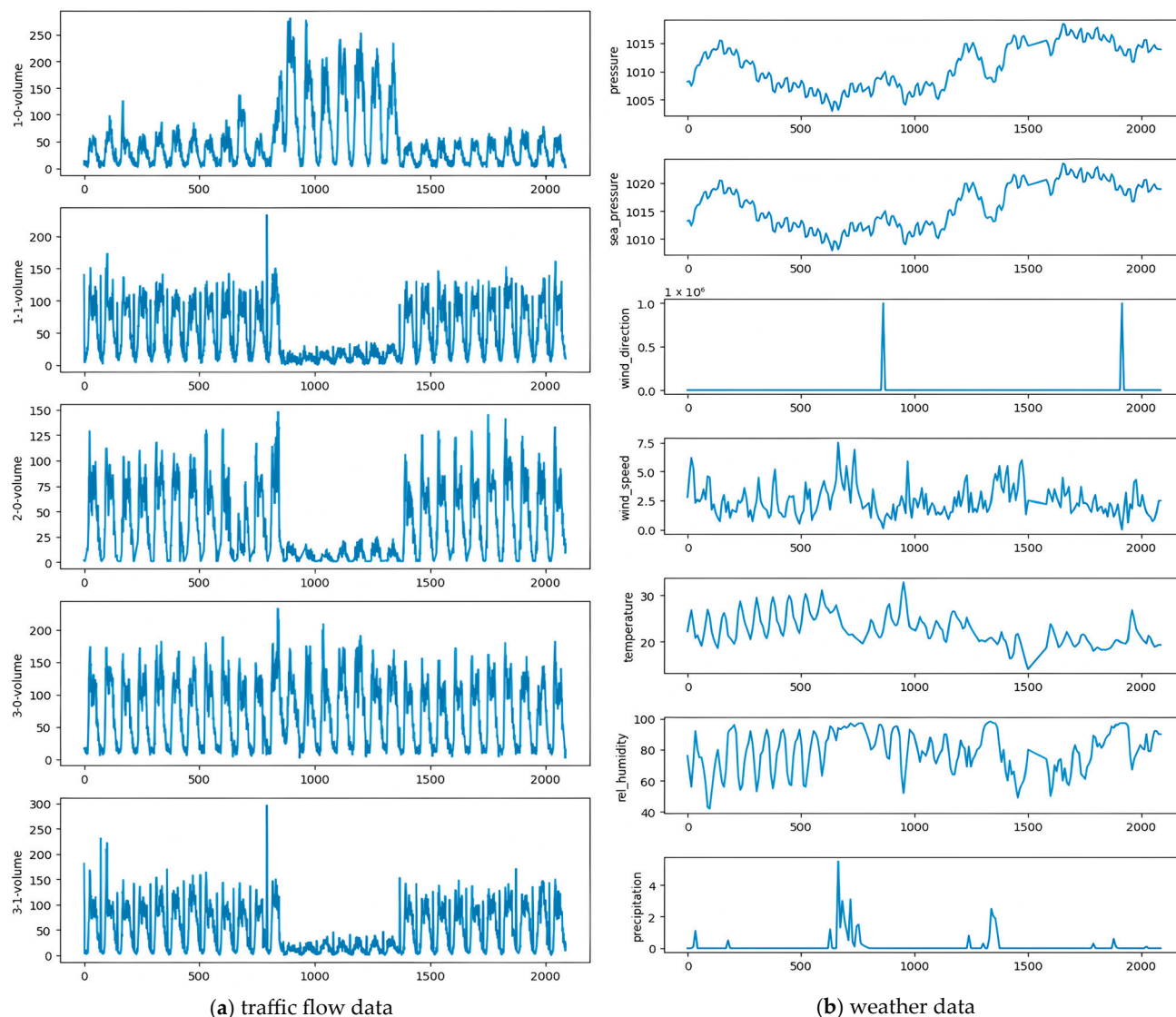


Figure 1. Traffic flow and weather data feature map.

From the figure, it can be observed that the traffic flow trends at each station are similar over certain periods. When studying traffic flow prediction, integrating data from all stations can be considered. Additionally, understanding traffic flow trends can help traffic management departments to allocate resources more effectively and to simplify management models.

Based on the previous analysis, periodic changes in traffic flow, holidays, and weather are key factors influencing traffic flow. Therefore, the weather and traffic flow data are integrated here, with a portion of the data shown in Table 3 below.

Table 3. Traffic flow and weather data integration table.

Time_Window	Pressure (hPa)	Sea_Pressure (hPa)	Wind_Direction (°)	Wind_Speed (m/s)	Temperature (°C)	Rel_Humidity (%)	Precipitation (mm)	Tollgate_ID	Direction	Volume
[2016-09-19 00:00:00, 2016-09-19 00:20:00]	1008.2	1013.2	329	2.8	22.2	76	0	1	0	3
[2016-09-19 00:20:00, 2016-09-19 00:40:00]	1008.213	1013.213	289.75	3.0125	22.475	74.875	0	1	0	6
[2016-09-19 00:40:00, 2016-09-19 01:00:00]	1008.225	1013.225	250.5	3.225	22.75	73.75	0	1	0	9
[2016-09-19 01:00:00, 2016-09-19 01:20:00]	1008.238	1013.238	211.25	3.4375	23.025	72.625	0	1	0	0
[2016-09-19 01:20:00, 2016-09-19 01:40:00]	1008.25	1013.25	172	3.65	23.3	71.5	0	1	0	4
[2016-09-19 01:40:00, 2016-09-19 02:00:00]	1008.263	1013.263	132.75	3.8625	23.575	70.375	0	1	0	10

After integration, this study finds that the data features are not on the same scale, which could affect prediction accuracy. Therefore, before analysis, the obtained data were normalized. This study employs the commonly used min–max method for data normalization, scaling the data to a range between 0 and 1. The min–max method is shown in Formula 1 below:

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x_i represents the original data, x'_i represents the normalized data, x_{min} is the minimum value, and x_{max} is the maximum value.

3.2. Analysis of Highway Traffic Flow Characteristics

Highway traffic flow, as a crucial indicator reflecting the operational status of road transportation systems, is shaped and driven by a series of complex internal conditions and external environments. Therefore, to better predict traffic flow, it is necessary to analyze both internal and external characteristics.

(1) Analysis of Internal Traffic Flow Characteristics

(a). Periodicity

The periodicity analysis of highway traffic flow focuses on temporal characteristics, primarily manifested in daily and weekly cycles, exhibiting both complexity and regularity. As shown in Figure 2, the data represent traffic flow from 00:00 on 7 October 2016 to 00:00 on 8 October 2016 in Figure 2a. It can be observed that the traffic flow exhibits significant fluctuations within a single day, indicating strong complexity. Figure 2b shows the traffic flow data from 19 September 2016 to 25 September 2016, revealing a high degree of similarity in traffic flow trends over a week, with peak commuting times being largely consistent. Figure 2c displays traffic flow data from 21 September 2016 to 11 October 2016, including both workdays and holidays. The data show two distinct patterns, but within individual workdays or holiday periods, the trends remain similar.

By leveraging long-term accumulated highway traffic data, these periodic phenomena can be precisely quantified, and the patterns can be used to predict future traffic conditions. This, in turn, guides the optimal allocation of traffic resources and the design of more accurate and efficient traffic management and control measures. Additionally, the results of the periodicity analysis provide indispensable scientific support for highway system expansion plans and the development of emergency response strategies for unexpected events.

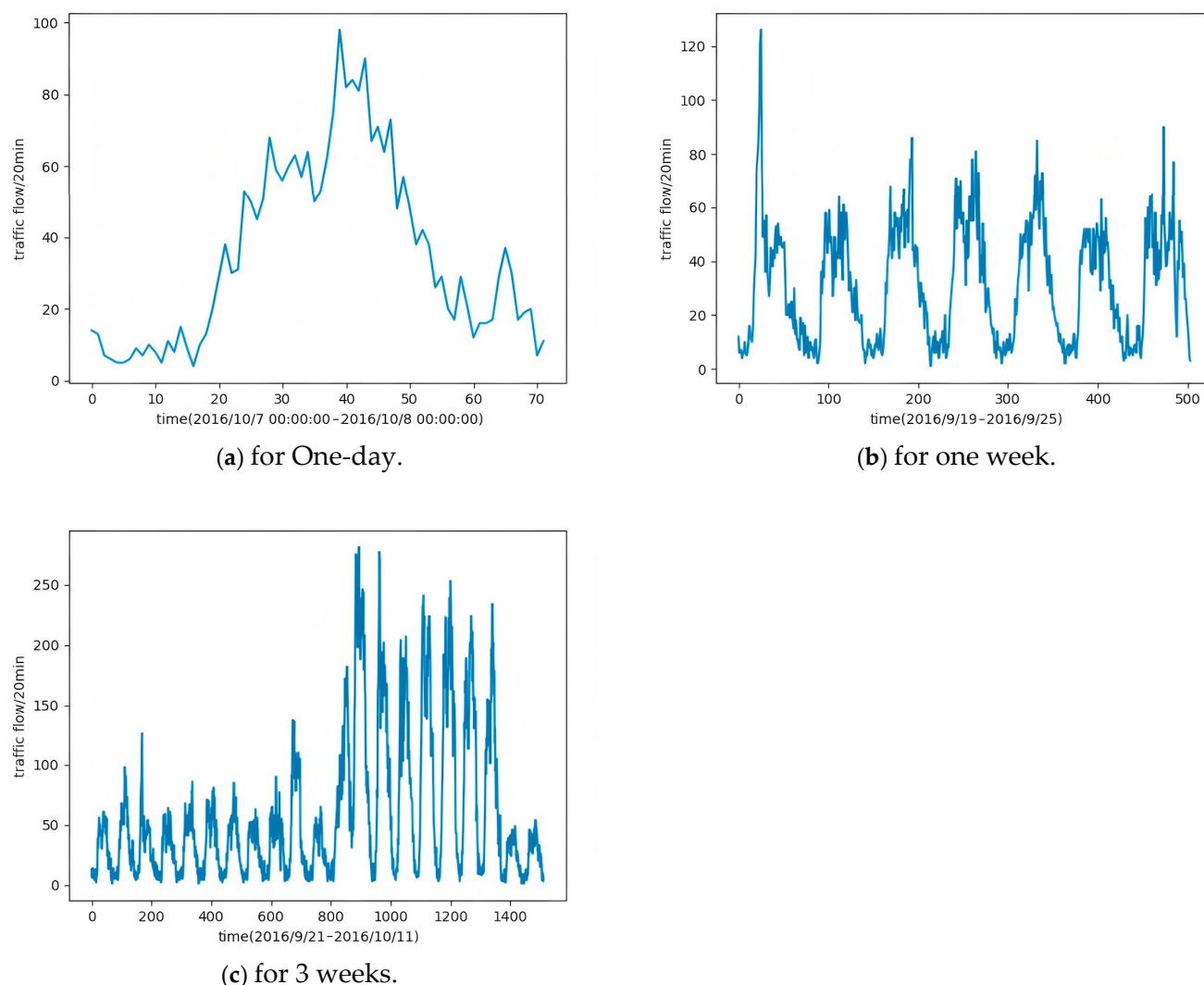


Figure 2. Traffic flow data cycle trend chart: (a) for one day, (b) for one week, (c) for 3 weeks.

(b). Spatial Correlation

In the highway system, the spatial correlation of traffic flow refers to the inherent connections and influencing mechanisms between traffic volume, density, and speed across different road segments. Understanding this characteristic allows for a global analysis of the operational patterns of the traffic system. This not only provides recommendations for optimizing existing traffic management models but also enables better allocation of road resources to alleviate congestion. In the long term, grasping the spatial correlation of highway traffic flow helps in designing more forward-looking and adaptable road network layouts, thereby enhancing the operational efficiency and service quality of the road transportation system.

Figure 3 shows the traffic flow trends from 20 September 2016 to 29 September 2016 for the 0-direction of the three adjacent toll stations (Station 1, Station 2 and Station 3). It can be observed that the traffic flow changes at adjacent stations are highly correlated, with consistent trends.

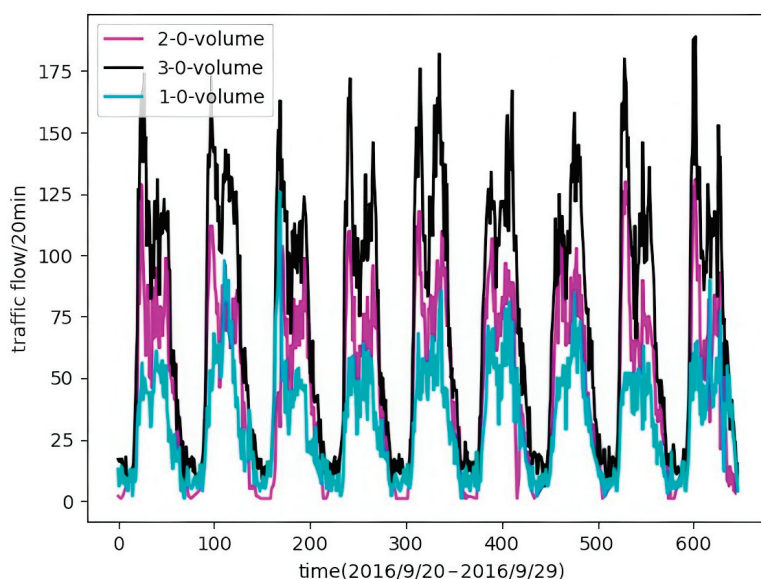


Figure 3. Traffic flow trends at adjacent toll stations.

(2) Analysis of External Traffic Flow Characteristics

Traffic flow has uncertainty, which means that the influence of other factors outside the transportation system can cause dynamic changes in traffic flow. According to daily experience, the occurrence of unconventional congestion is usually due to factors such as commuting time, holidays, and weather. In order to achieve a more accurate prediction of highway traffic flow, this article integrates the impact of key factors on traffic flow and conducts research.

The Pearson correlation analysis method is used to calculate the correlation between dynamic weather factors and traffic flow. The Pearson correlation coefficient is calculated using Formula (2) as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (2)$$

The strength of the correlation results is judged based on the range shown in Table 4 below.

Table 4. Correlation coefficient value and intensity correspondence table.

Correlation Coefficient ($ r $) Range	Strength of Correlation
$0.8 \leq r \leq 1$	Very Strong
$0.6 \leq r \leq 0.8$	Strong
$0.4 \leq r \leq 0.6$	Moderate
$0.2 \leq r \leq 0.4$	Weak
$0 \leq r \leq 0.2$	Very Weak or No Correlation

The calculation results are shown in Figure 4.

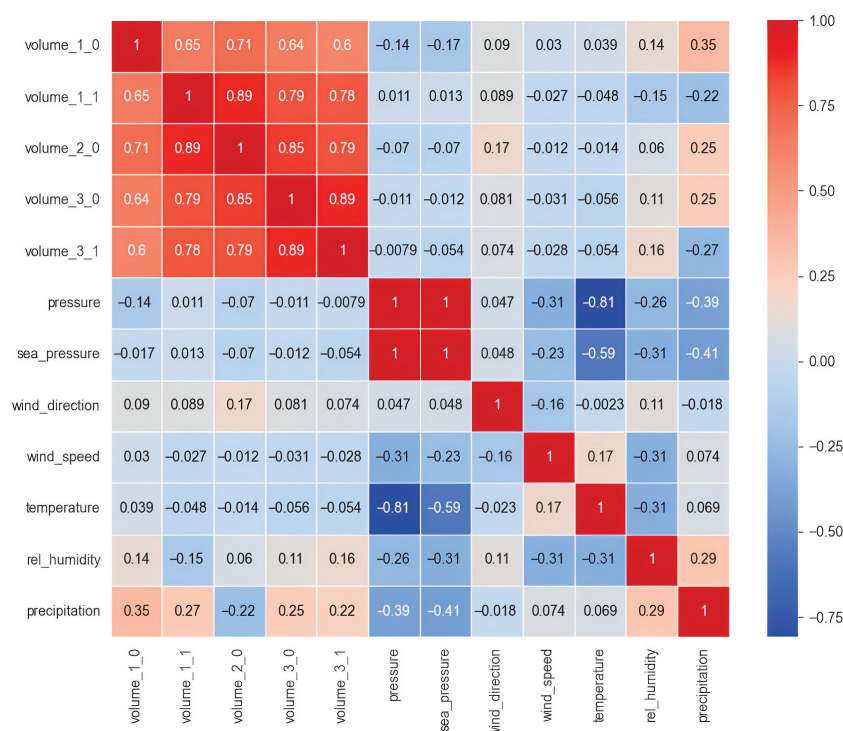


Figure 4. Correlation diagram between each station and weather factors.

As shown in Figure 4, we found that the correlations between Station 1-0 and other stations are 0.65, 0.74, 0.64, and 0.60, all above 0.6, indicating strong correlations between the traffic flows at different stations. Therefore, when predicting traffic flow at a target station, the traffic flow data from other stations should also be considered.

Taking Station 1-0 as an example, the correlations between traffic flow and weather factors such as atmospheric pressure (pressure), sea-level pressure (sea_pressure), wind direction (wind_direction), wind speed (wind_speed), temperature (temperature), and relative humidity (rel_humidity) are -0.14 , -0.017 , 0.09 , 0.03 , 0.039 , and 0.14 , respectively, all below 0.2 . This indicates that these factors have almost no correlation with the traffic flow data at this station. However, the correlation between precipitation (precipitation) and traffic flow data is 0.35 , indicating a strong correlation. This also proves that natural factors like weather do indeed influence traffic flow.

4. Methodology

LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) are models developed to address the vanishing gradient problem encountered by recurrent neural networks (RNNs) when processing continuous and lengthy data. Both models introduce gating mechanisms to more effectively regulate the flow and storage of information, thereby improving the processing of time-series data. While LSTM can finely handle long-term dependencies, its structure is complex. On the other hand, GRU has a simpler structure but sacrifices some prediction accuracy. Therefore, this study adopts an integrated LSTM-GRU approach to leverage the strengths of both models in handling long-term dependencies. However, neither model can process the spatial features in traffic flow data. To address this, the study incorporates CNN (Convolutional Neural Network) to extract spatial features and validates the model using highway traffic flow data and weather data from a specific province.

4.1. Model Architecture Design

The architecture mainly consists of three parts, which are data preprocessing, feature matrix construction, and model building. First, missing values in the raw traffic flow and weather data are filled, and the data are normalized. Then, the processed data are constructed into a matrix containing time, space, and weather features, which serves as the input to the model. In the model phase, the CNN is used to extract spatial features, while a combination of LSTM and GRU layers processes the temporal features. Finally, a fully connected layer (Dense) is used to output the results. The main architecture of the model is shown in Figure 5 below.

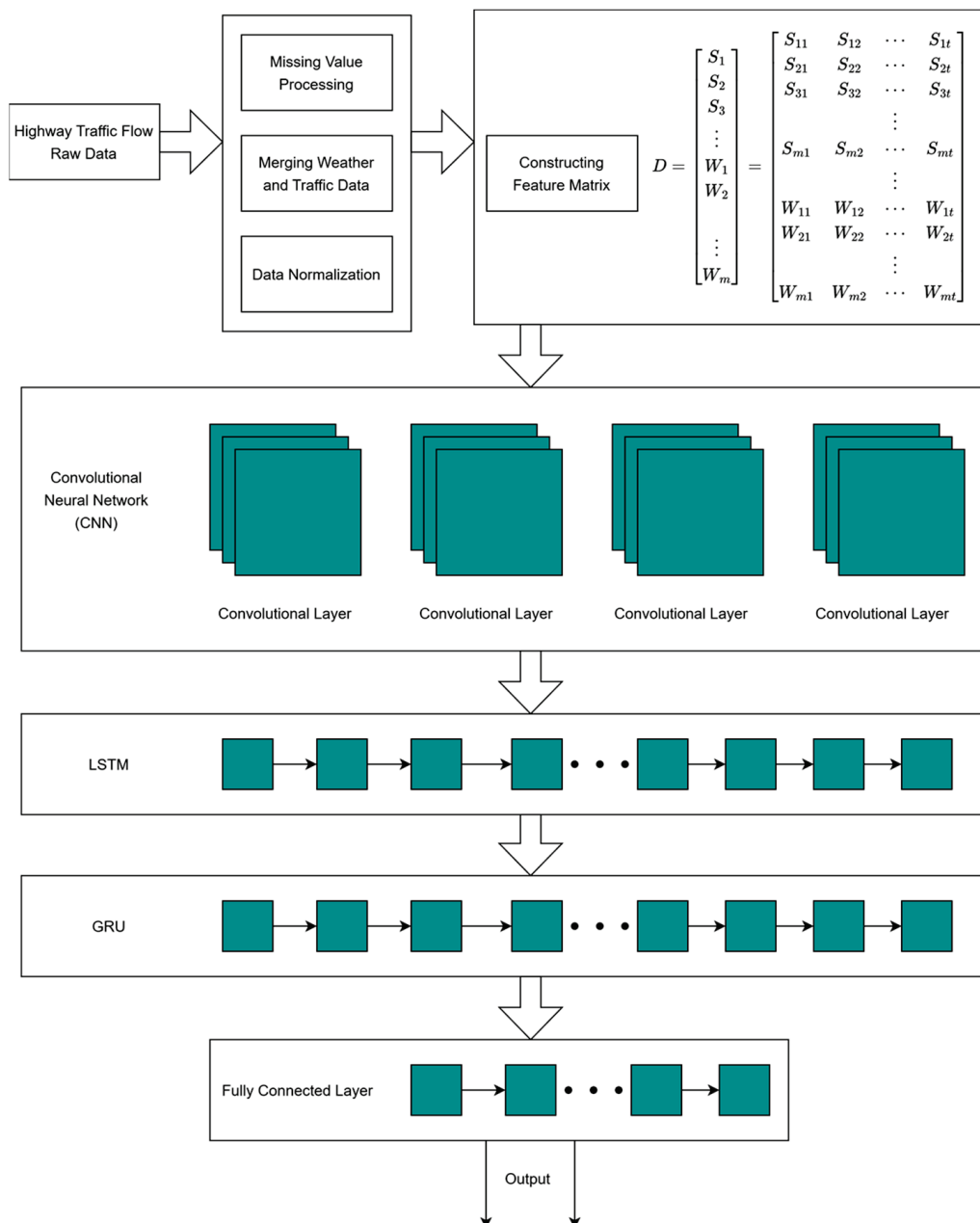


Figure 5. Model architecture diagram.

When extracting temporal features using LSTM-GRU, the Keras Sequential model is first used, with an LSTM layer embedded using the add function. The parameter `return_sequences` is set to `True`. Then, a GRU layer is added, using the output of the LSTM layer as its input. Finally, a fully connected layer (Dense) and an output layer are appended to obtain the final prediction. The specific structure is shown in Figure 6 below.

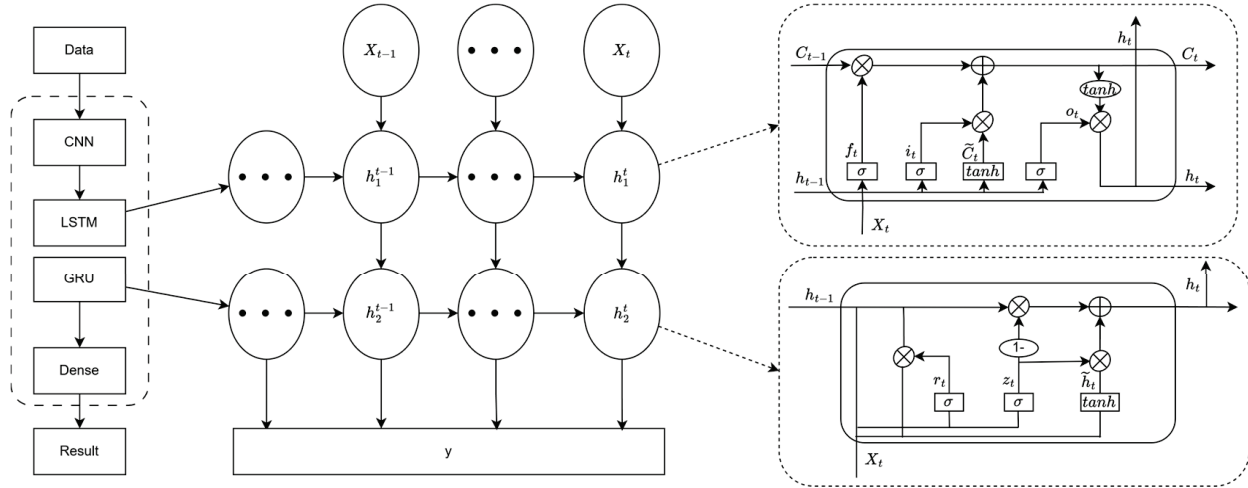


Figure 6. LSTM-GRU internal structure diagram.

- (1) The LSTM layer mainly includes structures such as the forget gate, input gate, and output gate. The calculation process is shown in Formulas (3)–(8) below:

$$f_t = \sigma(f_t) = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (3)$$

$$i_t = \sigma(i_t) = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (4)$$

$$\tilde{C}_t = \tanh(c_t) = \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \quad (5)$$

By integrating the above formulas, the following calculation formula is derived:

$$c_t = c_{t-1} * f_t + \tilde{C}_t * i_t \quad (6)$$

where x_t represents the input variable at the current time step. h_{t-1} represents the hidden state at the previous time step. c_{t-1} represents the cell memory at time $t - 1$. c_t represents the cell memory at time t . $\tilde{C}_t$ represents the candidate cell memory. W represents the weight matrices for the input state and hidden state. b represents the value of bias.

$$m_t = \tanh(c_t) \quad (7)$$

$$h_t = o_t * m_t \quad (8)$$

The hidden state h_{t-1} and input x_t are processed through the output gate's Sigmoid function to calculate o_t . The cell memory c_t is activated using the tanh function, ultimately yielding the updated hidden state h_t .

- (2) The GRU layer mainly includes the reset gate and update gate. The calculation process is shown in Formulas (9)–(13) below:

$$r_t = \delta(W_r \cdot [h_{t-1}, x_t]) \quad (9)$$

$$z_t = \delta(W_z \cdot [h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (11)$$

$$\tilde{h}_{t-1} = r_t * h_{t-1} \quad (12)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (13)$$

where r_t represents the portion of the hidden state from the previous time step that needs to be reset. δ is a Sigmoid activation function ranging from 0 to 1, quantifying the degree of selective forgetting. h_{t-1} represents the hidden state from the previous time step. x_t is the current input. W_r is the relevant weight matrix.

z_t represents the degree of retention for the current data, ranging from 0 to 1. A value closer to 0 indicates more forgetting, while a value closer to 1 indicates more retention.

During the calculation of $\tilde{h}_t$, the reset data are first combined with h_{t-1} to obtain $\tilde{h}_{t-1}$. Then, $\tilde{h}_{t-1}$ is combined with x_t and scaled using the tanh function.

Finally, the updated hidden state h_t is obtained.

4.2. Construction of the Two-Dimensional Matrix

As mentioned earlier, highway traffic flow is influenced by both internal and external factors. Specifically, since traffic flow is time-series data, historical data are required for traffic flow prediction to forecast future traffic based on historical trends. Additionally, traffic flow at adjacent stations and upstream/downstream stations can also impact the target station's traffic flow. External dynamic factors can cause fluctuations in traffic as well. Therefore, considering these aspects, a matrix containing time, space, and weather features is constructed.

The specific calculation method is shown in Formula (14) below:

$$D = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ W_1 \\ W_2 \\ \vdots \\ W_m \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1t} \\ S_{21} & S_{22} & \dots & S_{2t} \\ S_{31} & S_{32} & \dots & S_{3t} \\ & & \vdots & \\ S_{m1} & S_{m2} & \dots & S_{mt} \\ & & \vdots & \\ W_{11} & W_{12} & \dots & W_{1t} \\ W_{21} & W_{22} & \dots & W_{2t} \\ & & \vdots & \\ W_{m1} & W_{m2} & \dots & W_{mt} \end{bmatrix} \quad (14)$$

where $S_1, S_2, S_3 \dots$ represent the traffic flow data, $W_1, \dots, W_m$ represent the weather data for the same time period. $S_1, S_2, S_3 \dots$ contain two types of data; one is the historical traffic flow data of the station itself, and the other is the traffic flow data from adjacent stations. The horizontal data in the matrix are sorted in chronological order, representing t historical data points.

4.3. Model Parameter Settings

The configuration of the model's network structure significantly impacts its predictive performance. For the highway toll station traffic flow data in this study, parameters such as the number of filters, convolution kernels, pooling layers, LSTM layers and units, GRU layers and units, learning rate, and optimizer are determined to suit the model.

First, the learning rate and step size are determined. The fixed parameter method is used to find the optimal learning rate. Table 5 shows the model's MAE and RMSE values under different learning rates.

Table 5. Model error value under different learning rates.

Learning Rates	0.001	0.002	0.003	0.004	0.005	0.006
MAE	5.016	5.992	6.043	4.940	4.965	6.278
RMSE	6.798	8.086	8.022	7.297	6.542	7.984

From the table, it can be observed that, when the learning rate is 0.005, the model's error values are relatively the smallest. Therefore, the learning rate is set to 0.005.

Next, with the learning rate fixed at 0.005, Table 6 shows the MAE and RMSE values of the model under different iteration counts.

Table 6. Model error value under different iteration times.

Epoch	1	25	50	100	150	200	250	300
MAE	23.947	9.389	7.441	5.270	5.873	5.370	4.795	5.767
RMSE	31.151	13.062	12.005	7.170	9.014	7.185	6.801	9.890

From the table, it can be concluded that, when the iteration count is 250, the model's error values are relatively the smallest. Thus, the iteration count is set to 250.

The number of filters is set to 3, with values of 224, 104, and 72. The number of convolution kernels is set to 3, with values of 6, 5, and 6. The pooling layers are set to 2. The LSTM layers and units are set to 1 and 128, respectively. The GRU layers and units are set to 1 and 96, respectively. The learning rate is fixed at 0.005, and the Adam optimizer is chosen. Other parameter settings are shown in Table 7 below.

Table 7. Parameter setting table.

Parameter Name	Value
Number of maxpool	2
Number of dropout	3
Number of LSTM layers	1
Number of GRU layers	1
Number of con_filters	224, 104, 72
Number of con_kernels	6, 5, 6
Number of dropouts	0.2, 0.3, 0.4
Unit number of each LSTM layer	128
Unit number of each GRU layer	96
Learning rate (lr)	0.005
Optimizer	Adam

4.4. Sliding Window and Evaluation Metrics

When handling time-series data prediction tasks, the model's prediction accuracy is indeed influenced by the sequence length setting. Given this, this study adopts the sliding window technique to select an appropriate sequence length. The sliding window method is an effective approach for processing array or sequence data structures by setting a fixed window size, which can simplify multiple loops into a single loop in specific scenarios, thereby reducing computational complexity.

The choice of window size directly affects the number of generated samples and the number of time-step features included in the samples. In other words, in a given dataset, a smaller window means capturing shorter time segments, resulting in more independent samples. Conversely, a larger window captures longer time spans, resulting in fewer samples. Therefore, selecting an appropriate window size is crucial for accurately predicting traffic flow.

In terms of evaluation metrics, with reference to the approaches in [27,28], the study employs MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) to evaluate the predictive performance of the regression model. RMSE is sensitive to extreme values in the data and measures the gap between predicted results and actual observations. MAE calculates the average of the absolute errors between predicted values and true values, intuitively reflecting the average level of model prediction bias. The mathematical formulas for these two evaluation metrics are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^*| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2} \quad (16)$$

where y_i and y_i^* represent the true values and predicted values, respectively.

Given the above reasons, it is necessary to calculate the appropriate time step size for the highway dataset of the province. The step sizes were set to 72 (1 day), 144 (2 days), 216 (3 days), 288 (4 days), 360 (5 days), 432 (6 days), 504 (7 days), and 576 (8 days), and the MAE and RMSE values are calculated for each. The specific results are shown in Table 8 below.

Table 8. Values of MAE and RMSE under different window sizes.

Window Size	MAE	RMSE
72 (1 day)	13.310	19.230
144 (2 day)	11.681	15.638
216 (3 day)	10.148	13.487
288 (4 day)	10.108	13.255
360 (5 day)	9.363	12.581
432 (6 day)	6.393	8.949
504 (7 day)	4.761	6.250
576 (8 day)	5.283	7.456

From the table, it can be observed that, when the window size is 504 (7 days), the MAE and RMSE values are the lowest, indicating that the model's prediction results are better under this window size. Based on these results, the window size is ultimately set to 504.

5. Results Comparison and Analysis

The experimental results in this section are analyzed from the following two perspectives: whether it is a holiday and whether weather factors are included. To better demonstrate the predictive performance of the proposed model, this study compares it with four other models, which are LSTM, GRU, CNN-LSTM, and CNN-GRU. The following presents the comparison of prediction results.

5.1. Comparison and Analysis of Workday Traffic Flow Prediction Results

This study takes 11–17 October 2016 as the prediction target for workdays and calculates the prediction errors of the model for each station, direction, and date, as shown in Table 9 below.

Table 9. Prediction error of the CNN-LSTM-GRU model for seven-day working days.

Date	Toll Station and Direction									
	1-0		1-1		2-0		3-0		3-1	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
11	3.767	4.781	7.097	10.198	3.202	4.765	5.873	8.048	8.865	12.889
12	3.284	4.462	6.291	9.283	3.512	4.791	5.780	7.988	7.099	9.786
13	3.894	4.871	6.912	10.305	3.492	5.710	5.953	9.778	8.133	10.867
14	4.515	5.842	7.460	10.392	4.576	7.345	6.473	10.089	7.914	11.410
15	4.346	6.468	5.928	8.135	4.156	6.807	5.655	9.319	8.706	14.579
16	4.492	5.992	7.228	10.142	4.682	6.978	8.656	12.995	8.432	11.772
17	5.955	7.857	10.929	16.878	5.676	7.870	7.892	10.558	10.170	14.542
Average	4.322	5.860	7.406	11.073	4.185	6.427	6.612	9.951	8.474	12.379

To better evaluate the model's prediction accuracy, Table 10 provides a comparison of the average prediction errors of the five models for 11–17 October 2016.

Table 10. Average prediction error of the 5 models for seven-day working days.

Date	Models									
	LSTM		GRU		CNN-LSTM		CNN-GRU		CNN-LSTM-GRU	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1-0	5.125	7.103	5.125	7.103	4.611	6.525	4.476	6.040	4.322	5.860
1-1	9.427	14.457	10.032	16.379	7.784	11.744	8.138	12.626	7.407	11.073
2-0	6.480	9.377	6.551	9.434	5.617	8.175	5.711	8.206	4.185	6.427
3-0	8.643	12.592	8.523	12.426	7.239	10.193	6.966	10.067	6.612	9.951
3-1	9.008	17.222	10.865	16.977	10.511	16.819	10.522	15.931	8.474	12.379

From the table above, it can be observed that, when predicting workday traffic flow, the proposed model achieves the lowest MAE and RMSE values, indicating superior prediction accuracy.

To more clearly demonstrate the prediction accuracy of the model, Figure 7 presents the prediction results of the proposed model (HDL) compared to LSTM, GRU, CNN-LSTM, and CNN-GRU for 11 October 2016, as shown in the following figures.

In the previous analysis, it was noted that weather factors can impact the entire traffic system, causing fluctuations in traffic flow. Therefore, this study incorporates weather factors into the model. Using 11–17 October 2016 as the prediction target, the average prediction errors for each station are calculated and compared with the other four models. The comparison results are shown in Table 11 below.

Table 11. Comparison of forecast errors of five models during working days after adding external factors.

Date	Model									
	LSTM		GRU		CNN-LSTM		CNN-GRU		CNN-LSTM-GRU	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1-0	5.266	7.335	5.017	6.858	4.411	6.017	4.198	5.706	2.905	4.229
1-1	9.573	13.755	8.384	12.552	8.601	13.495	7.128	10.977	6.303	9.829
2-0	6.232	9.009	6.348	9.167	5.768	8.437	5.065	7.409	4.105	6.271
3-0	8.488	12.292	8.506	12.506	8.235	11.859	8.013	11.264	6.333	9.279
3-1	10.918	18.150	10.304	15.793	9.845	15.949	10.403	16.056	5.395	8.567

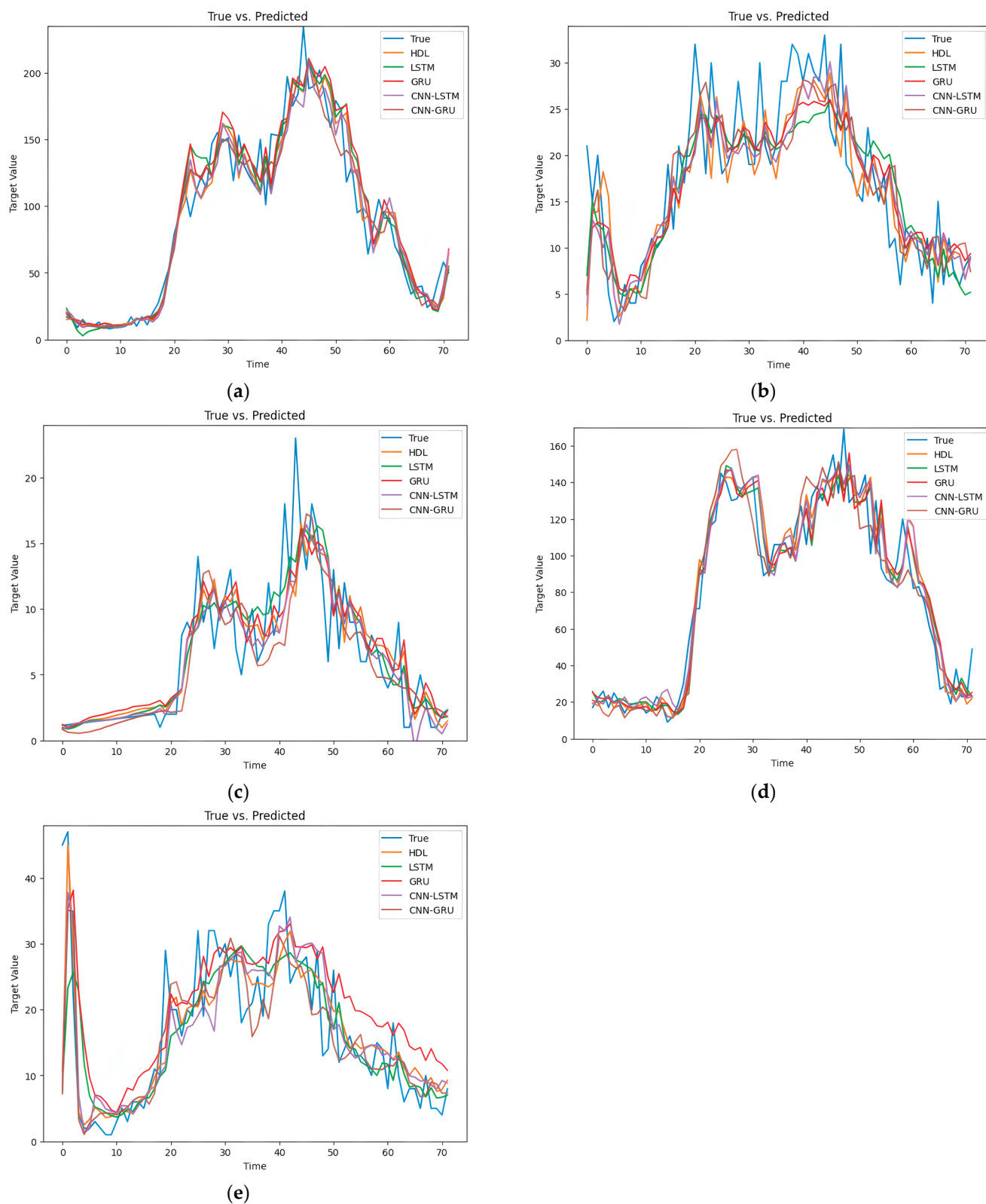


Figure 7. Traffic flow prediction results found by five models: (a) 1-0, (b) 1-1, (c) 2-0, (d) 3-0, and (e) 3-1.

After incorporating external factors, it can be observed that the model's prediction results have significantly improved. For the CNN-LSTM-GRU predictions at each station, the MAE and RMSE decreased by 32.8% and 27.8%, 14.9% and 11.23%, 1.91% and 2.42%,

4.22% and 6.75%, and 36.3% and 30.79%, respectively. This indicates that including weather factors effectively enhances prediction accuracy.

5.2. Comparison and Analysis of Holiday Traffic Flow Prediction Results

In this section, 1–7 October 2016 is taken as the prediction target for holidays. The prediction errors of the model for each station, direction, and date are calculated, as shown in Table 12 below.

Table 12. Prediction error of the CNN-LSTM-GRU model for seven-day holidays.

Date	Toll Station and Direction											
	1-0		1-1		2-0		3-0		3-1			
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	16.946	21.963	2.847	3.545	1.507	2.188	7.090	9.467	2.174	2.770		
2	14.258	20.021	2.690	3.554	1.238	1.883	6.269	10.575	2.556	3.845		
3	12.073	21.151	2.739	3.601	0.916	1.338	5.707	11.003	2.873	4.635		
4	11.047	15.398	2.880	3.689	1.339	1.921	6.745	10.695	3.029	4.571		
5	15.745	21.591	3.268	4.496	1.644	2.322	6.623	10.929	2.909	4.635		
6	11.566	15.373	3.694	4.966	1.667	2.654	6.700	10.328	3.079	4.847		
7	5.955	7.857	4.574	5.612	1.943	2.800	9.624	13.328	4.516	7.284		
Average	13.496	19.333	3.242	4.278	1.464	2.207	6.966	10.947	3.019	4.824		

To better evaluate the model's prediction accuracy, Table 13 provides a comparison of the average prediction errors of the five models for 1–7 October 2016.

Table 13. Average prediction error of the 5 models for seven-day holidays.

Date	Model											
	LSTM		GRU		CNN-LSTM		CNN-GRU		CNN-LSTM-GRU			
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE		
1-0	33.250	49.627	29.083	44.224	14.634	21.506	14.326	20.909	13.496	19.333		
1-1	3.685	4.998	4.269	5.672	3.689	4.886	3.345	4.346	3.242	4.278		
2-0	1.819	2.646	1.828	2.779	2.003	2.982	1.647	2.433	1.465	2.207		
3-0	15.387	23.101	14.568	21.588	12.032	16.479	10.737	15.229	6.966	10.947		
3-1	3.561	4.911	3.901	5.499	3.994	5.691	3.624	5.069	3.019	4.824		

From Table 13, it can be seen that, when predicting holiday traffic flow, the proposed model achieves the lowest MAE and RMSE values, indicating superior prediction accuracy.

In the previous analysis, it was noted that weather factors can impact the entire traffic system, causing fluctuations in traffic flow. Therefore, this section incorporates weather factors into the model. Using 1–7 October 2016 as the prediction target, the average prediction errors for each station are calculated and compared with the other four models. The comparison results are shown in Table 14 below.

After incorporating external factors, the model's prediction results show improvement. For the CNN-LSTM-GRU predictions at each station, the MAE and RMSE decreased by 7.08% and 6.96%, 1.48% and 3.65%, 17.3% and 1.45%, 2.31% and 4.10%, and 1.62% and 9.35%, respectively. This indicates that including weather factors also improves the prediction accuracy for holiday traffic flow. However, compared to workday prediction accuracy, there is a noticeable gap. This is because the data volume for the National Day holiday is smaller than that for workdays, and the traffic flow during the holiday is influenced by more uncontrollable factors and exhibits greater fluctuations, resulting in lower prediction accuracy for holiday traffic flow.

Table 14. Comparison of prediction errors of the 5 models during holidays after adding external factors.

Date	Model									
	LSTM		GRU		CNN-LSTM		CNN-GRU		CNN-LSTM-GRU	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1-0	31.937	48.243	28.348	44.403	14.437	20.825	14.453	21.049	12.958	17.987
1-1	3.105	4.479	4.027	5.429	3.370	4.609	3.143	4.446	3.194	4.434
2-0	1.672	2.282	1.980	2.521	1.766	2.941	1.422	2.274	1.211	2.175
3-0	14.586	23.567	14.518	21.351	12.801	14.943	10.230	15.042	6.805	10.498
3-1	3.055	4.439	3.408	4.995	3.457	5.252	3.223	4.699	2.970	4.373

Based on the above analysis, after conducting two sets of comparisons, the results indicate that incorporating weather factors and holiday factors leads to smaller prediction errors for the integrated model compared to individual models, both on workdays and holidays. The proposed model in this study demonstrates superior prediction accuracy for highway traffic flow over the other four typical deep learning models. The performance of the model has improved to varying degrees.

The recent literature using relevant datasets for traffic flow prediction was collected, and the prediction results are shown in Table 15 below.

Table 15. Prediction results of highway traffic flow in existing literature.

Model	MAE	RMSE
CNN-LSTM-GRU	2.905/(vehicles-ehiclesG-1)	4.229/(vehicles-ehiclesG-1)
CEEMD-CNN-LSTM-Attention	3.488/(vehicles-ehiclesN-1)	4.512/(vehicles-ehiclesN-1)

From the table, it can be concluded that, within the same dataset used in this study, the proposed model outperforms the CEEMD-CNN-LSTM-Attention model.

6. Conclusions

The main work of this paper included two aspects. Firstly, a deep learning model based on CNN-LSTM-GRU was proposed. Although LSTM could effectively handle long-term dependencies in data, its structure was complex; GRU had the advantage of a simpler structure but sacrificed some prediction accuracy. Therefore, this paper adopted an integrated LSTM and GRU framework to combine their respective strengths in managing long-term dependencies. However, both models were unable to capture spatial features in traffic flow data. To address this, our study integrated LSTM-GRU with CNN to extract spatial features and validated the model using highway traffic flow data from a certain province.

Additionally, in the experimental analysis, considering factors affecting traffic flow prediction, this paper incorporated temporal factors and weather data and analyzed the results from the following two perspectives: whether holidays were included and whether weather data had been added. The findings showed that the proposed model exhibited better prediction performance compared to LSTM, GRU, CNN-LSTM, and CNN-GRU, indicating that incorporating temporal and weather elements could further enhance the model's prediction accuracy.

However, there are still some limitations in the research process of this paper, which need to be further optimized in future studies. For example, in terms of dataset, the dataset used in the paper is relatively limited and outdated. In order to verify the applicability of the algorithm, experiments will need to be conducted on more and updated datasets in the future. In addition, in the study, the impact of road conditions and driver's own reasons on traffic flow was not taken into account. In future research, a more comprehensive indicator system of influencing factors can be established.

Author Contributions: Conceptualization: J.Z. and J.S.; methodology: J.S.; validation: C.Z.; formal analysis: J.S. and Y.Z.; investigation: Y.Z.; resources: J.S.; data curation: C.Z.; writing—original draft

preparation: J.S. and Y.Z.; writing—review and editing: J.Z. and C.Z.; visualization: C.Z.; supervision: J.Z. and Y.Z.; project administration: Y.Z.; funding acquisition: J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities, grant number 3132024302, and the Humanities and Social Sciences Foundation of Ministry of Education under Grant 21YJC630066.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

- Guo, J.; Huang, W.; Williams, B.M. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 50–64. [CrossRef]
- Wang, Y.; Papageorgiou, M. Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. *Transp. Res. Part B* **2005**, *39*, 141–167. [CrossRef]
- Ahmed, M.S.; Cook, A.R. Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transp. Res. Rec.* **1979**, *722*, 1–9.
- Smith, B.L.; Williams, B.M.; Oswald, R.K. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transp. Res. Part C* **2002**, *10*, 303–322. [CrossRef]
- Kumar, S.V.; Vanajakshi, L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *Eur. Transp. Res. Rev.* **2015**, *7*, 1–9. [CrossRef]
- Chen, Y.; Xiao, D. Traffic network flow forecasting based on switching model. *Control Decis.* **2009**, *24*, 1177–1180+1186.
- Zhang, M.; Zhen, Y.; Hui, G.; Chen, G. Accurate Multisteps Traffic Flow Prediction Based on SVM. *Math. Probl. Eng.* **2013**, *12*, 91–109.
- Hou, Y.; Edara, P.; Sun, C. Traffic Flow Forecasting for Urban Work Zones. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 1761–1770. [CrossRef]
- Zhang, T.; Chen, X.; Xie, M.; Zhang, Y. K-NN based nonparametric regression method for short-term traffic flow forecasting. *Syst. Eng. Theory Pract.* **2010**, *30*, 376–384.
- Lu, W.; Hu, Y.; Chen, W.; Qin, Y.; Wu, C.; He, X. Traffic flow prediction for highway vehicle detectors through decomposition and machine learning. *Transp. Lett. Int. J. Transp. Res.* **2025**, *17*, 260–280. [CrossRef]
- Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X.; Li, T. Predicting Citywide Crowd Flows Using Deep Spatio-Temporal Residual Networks. *Artif. Intell.* **2017**, *259*, 147–166. [CrossRef]
- Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C* **2015**, *54*, 187–197. [CrossRef]
- Wu, Y.; Tan, H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. *arXiv* **2016**, arXiv:1612.01022.
- Mackenzie, J.; Roddick, J.F.; Zito, R. An evaluation of HTM and LSTM for short-term arterial traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 1847–1857. [CrossRef]
- Tian, Y.; Zhang, K.; Li, J.; Lin, X.; Yang, B. LSTM-based traffic flow prediction with missing data. *Neurocomputing* **2018**, *318*, 297–305. [CrossRef]
- Feng, X.; Ling, X.; Zheng, H.; Chen, Z.; Xu, Y. Adaptive Multi-Kernel SVM With Spatial-Temporal Correlation for Short-Term Traffic Flow Prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2001–2013. [CrossRef]
- Guo, S.; Lin, Y.; Li, S.; Chen, Z.; Wan, H. Deep Spatial-Temporal 3D Convolutional Neural Networks for Traffic Data Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3913–3926. [CrossRef]
- Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
- Fang, W.; Zhuo, W.; Yan, J.; Song, Y.; Jiang, D.; Zhou, T. Attention meets long short-term memory: A deep learning network for traffic flow forecasting. *Phys. A Stat. Mech. Its Appl.* **2022**, *587*, 126485. [CrossRef]
- Shuai, C.; Pan, Z.; Gao, L.; Zuo, H. Short-term traffic flow prediction of expressway: A hybrid method based on singular spectrum analysis decomposition. *Adv. Civ. Eng.* **2021**, *2021*, 4313970. [CrossRef]
- Bing, Q.; Shen, F.; Chen, X.; Zhang, W.; Hu, Y.; Qu, D. A hybrid short-term traffic flow multistep prediction method based on variational mode decomposition and long short-term memory model. *Discret. Dyn. Nat. Soc.* **2021**, *2021*, 4097149. [CrossRef]

22. Xu, X.; Liu, C.; Zhao, Y.; Lv, X. Short-term traffic flow prediction based on whale optimization algorithm optimized BiLSTM_Attention. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6782. [CrossRef]
23. Zheng, Y.; Dong, C.; Dong, D.; Wang, S. Traffic volume prediction: A fusion deep learning model considering spatial-temporal correlation. *Sustainability* **2021**, *13*, 10595. [CrossRef]
24. Zheng, H.; Lin, F.; Feng, X.; Chen, Y. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 6910–6920. [CrossRef]
25. Lian, R.; Wang, X. Short-time traffic flow prediction based on a hybrid model of Bi-LSTM-CNN. *Adv. Transp. Stud.* **2025**, *65*, 87–102.
26. Ding, R.; Xie, H.; Dai, C.; Qiao, G. Research on ship traffic flow prediction based on GTO-CNN-LSTM. In Proceedings of the International Conference on Traffic Engineering and Transportation System, Dalian, China, 22–24 September 2023.
27. Jia, Y.; Wu, J.; Xu, M. Traffic flow prediction with rainfall impact using a deep learning method. *J. Adv. Transp.* **2017**, *1*, 6575947. [CrossRef]
28. Huang, H.; Xie, J.; Li, Z.; Sun, X.; Peng, T. Traffic flow forecasting method based on Gated Spatial-temporal Spatiotemporal Graph Network and TCN. *J. Traffic Sci. Technol.* **2024**, *6*, 126–131.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Evaluating the Impact of Automated Vehicle Penetration on Intersection Traffic Flow: A Microsimulation-Based Approach

Mircea Augustin Rosca, Floriana Cristina Oprea *, Vasile Dragu, Oana Maria Dinu, Ilona Costea and Stefan Burciu

Transport Faculty, National University for Science and Technology Politehnica Bucharest, Spl Independentei, No 313, RO-060042 Bucharest, Romania; mircea.rosca@upb.ro (M.A.R.); vasile.dragu@upb.ro (V.D.); oana.dinu@upb.ro (O.M.D.); ilona.costea@upb.ro (I.C.); stefan.burciu@upb.ro (S.B.)

* Correspondence: cristina.oprea@upb.ro; Tel.: +40-744774354

Abstract: As automation technologies continue to advance within the automotive industry, urban road traffic is gradually shifting from conventional driving toward fully autonomous. This transition is supported by the progressive integration of partially automated functions, such as Adaptive Cruise Control (ACC) and lane-keeping assistance, which are already implemented in commercial vehicles and increasingly affect both individual driving behavior and overall traffic flow dynamics. The main purpose of this research is to evaluate the impact of automated vehicles presence in a complex signalized intersection under mixed traffic conditions, considering different penetration rates and demand levels. A review of previous modeling approaches from the literature was conducted, highlighting critical aspects to be considered in the design and simulation of road traffic. Field traffic data were collected and used as input for a microsimulation model developed in AIMSUN. A base scenario and a 20% growth scenario were analyzed to assess the impact of AV-ACC penetration, varying the AV-ACC's rates in traffic composition. The results indicate that increased AV-ACC penetration rates, especially beyond 50%, contribute significantly to improving traffic stability and efficiency.

Keywords: automated vehicle; microsimulation; mixed traffic flow; adaptive cruise control (ACC)

1. Introduction

Urban mobility has become an increasingly relevant issue as cities gain prominence, now hosting more than 56% of the world's population and accounting for over 80% of global GDP [1]. Traditionally, urban planning principles were designed to improve efficiency and quality of life. The development of Smart Mobility offers researchers and practitioners innovative approaches to understanding and designing cities [2]. Given the rapid growth of urbanization and the new mobility challenges faced by most cities, paper [3] seeks to examine the evolution of the Smart Urban Mobility concept through a bibliometric analysis.

To further justify the relevance of the study, we incorporate recent pan-European quantitative evidence. Road congestion in Europe imposes substantial economic costs estimated at over EUR 110 billion annually, equating to approximately 1% of GDP [4].

On the technological front, 67% of new private vehicles sold in Europe during Q1 2025 were equipped with advanced Level-2 autonomous features (ADAS), underlining the growing relevance of automation in traffic modeling [5].

Smart urban mobility is a fundamental component of the smart city concept, which involves maintaining a high quality of travel parameters within urban areas [6–8]. The gradual automation of driving processes, including the deployment of automated vehicles,

is expected to contribute to sustainable urban mobility, as users begin to shift from active drivers to passive system supervisors [9–13].

The transition toward higher levels of driving automation, supported by advancements in Artificial Intelligence, is occurring incrementally. Four evolutionary traffic phases are commonly identified: (1) traffic composed entirely of regular vehicles (RVs); (2) mixed traffic with both RVs and vehicles with different levels of automation, with RVs still prevailing; (3) mixed traffic where vehicles with different levels of automation become the majority and (4) fully automated traffic. The overarching goal is to develop effective traffic management policies that account for large-scale vehicle interactions, human behavioral variability and the coexistence of RVs with automated vehicles [14].

The smart city concept is closely tied to the adoption of intelligent vehicles [15,16], which play a key role in making urban environments more sustainable, efficient and livable. Increasingly, cities are integrating automated vehicles (AVs) into their transport systems, reshaping urban infrastructure and influencing its environmental sustainability [17]. Paper [18] examines public acceptance of Connected and Automated Vehicles (CAVs), a category of intelligent vehicles that has recently moved from vision to reality. While there is a degree of acceptance for these vehicles, a lack of public awareness leads to some hesitation. As the share of AVs increases, mobility, safety and environmental benefits will eventually improve [19,20].

The Society of Automotive Engineers International (SAE) defines Connected and Automated Vehicles (CAVs) using the J3016 classification framework, which specifies six levels of driving automation from Level 0 to Level 5 [21]. Level 0 (No Automation)—no driving automation; Level 1 (Driver Assistance)—steering or brake/acceleration assistance, with full driver supervision. Level 2 (Partial Automation)—both steering and brake/acceleration assistance, typically combining lane centering with Adaptive Cruise Control (ACC), which maintains speed and following distance, driver must remain attentive. Level 3 (Conditional Automation)—vehicle drives under specific conditions, driver takes over when requested. Level 4 (High Automation)—full driving within defined operational domains (e.g., local driverless taxi) without driver input. Level 5 (Full Automation)—complete driving automation under all conditions, no driver required [22].

Partial automation, like Advanced Driver Assistance Systems (ADAS), enhances driving comfort and safety by reducing human errors. A key example is Adaptive Cruise Control (ACC), which uses sensors to manage speed and distance from the vehicle ahead. An upgraded version, Cooperative Adaptive Cruise Control (CACC), uses vehicle-to-vehicle (V2V) communication for more precise and faster adjustments. Unlike ACC, CACC can communicate with other vehicles via dedicated short-range communication (DSRC), offering a larger detection range. However, in both systems, the driver still controls the steering and monitors the traffic [23].

The potential of automated vehicles (AVs) to improve urban traffic efficiency is a controversial topic, with conflicting findings reported in the literature. A commonly optimistic hypothesis argues that SAE Level 2 AVs, especially those equipped with ACC and CACC, can improve traffic flow stability by minimizing human driving errors and enabling more consistent car-following behavior [23]. On the other hand, critical hypotheses point out that vehicles with different levels of automation may offer limited benefits in complex urban environments. Thus, mixed traffic with penetration rates of automated vehicles with ACC below 50% will initially have a slight negative effect on traffic flow [22,24].

In our paper, the term “AV-ACC” refers to partially automated vehicles corresponding to SAE Level 2 automation. These vehicles are equipped with Adaptive Cruise Control (ACC), which enables automated longitudinal control, i.e., maintaining speed and headway to the preceding vehicle, but do not perform autonomous lateral maneuvers such as lane

changes or route-based decisions. Therefore, while longitudinal behavior (acceleration and deceleration) is automated, lateral control and situational awareness remain fully under human driver supervision. Although these vehicles are formally categorized as SAE Level 2 autonomous vehicles, in simulation environments, their lateral behavior is modeled similarly to that of conventional regular vehicles (RVs).

To evaluate the potential impact of partially automated vehicles on urban traffic, two traffic demand scenarios—a base scenario and a 20% growth scenario—have been considered in these analyses. In both cases, the presence of automated vehicles influenced traffic performance according to their penetration rates in the traffic. Results indicate that increasing the penetration rate of AV-ACC vehicles generally enhances traffic efficiency. However, under an increased traffic flow, mixed traffic for penetration rates below 50% of AV-ACC initially has a modest negative effect on traffic flow, due to the interaction complexities introduced by their coexistence. For AV-ACC penetration rates exceeding 50%, the improvements become evident, suggesting that automation benefits are dependent on reaching sufficiently high penetration levels.

2. Literature Review

The existing literature on automated vehicles covers a wide range of topics, from socio-economic implications to technical control strategies and traffic management approaches. To provide a clear and structured overview, the following review is organized into thematic categories: (1) socio-economic and environmental context of automated vehicles adoption, (2) driver behavior modeling and mixed traffic flow dynamics, (3) control strategies for connected and automated vehicles, (4) traffic management at intersections and network level, and (5) observed impacts and thresholds in mixed traffic scenarios.

(1) There are numerous studies that review the literature on AVs, examining their social, economic and financial advantages as well as the geographic scope of the studies. They emphasize that concerns exist across all continents regarding the need to reduce traffic congestion effects, decrease the number of accidents and improve traffic safety [25–29].

A taxonomy of socio-mobility was proposed in [30], which defines how AVs strengthen social connections since the mobility of AVs is considered a vehicle-related issue rather than a mobility issue. However, current research has several limitations: most studies focus on potential benefits in isolation without systematically comparing socio-economic trade-offs or considering transitional periods where AVs and RVs share infrastructure. There is limited quantitative analysis of environmental impacts and the interaction with urban land use. This gap highlights the need for studies that integrate socio-economic, environmental, and urban planning dimensions simultaneously, allowing a more realistic assessment of AVs adoption effects.

(2) Several studies focus on understanding driver behavior and its impact on mixed traffic conditions. An analysis of driver behavior models was proposed in [31]. These models have been used as input data for self-coaching, accident prevention studies and the development of driver assistance systems.

In [32], the authors introduced a model designed to examine the interaction between AVs and RVs while accounting for the uncertainty of human driving behavior. The experimental results show that AVs exert a notable influence on both the uncertainty and stability of mixed traffic flow. Moreover, higher penetration rates of AVs can mitigate the unpredictability of RV behavior and enhance the overall stability of the traffic system.

According to [33], the advancement of vehicle automation will lead to a future in which human-driven vehicles and autonomous vehicles coexist. As a result, conventional vehicles, carsharing services, autonomous vehicles and shared autonomous vehicles will operate side by side, reshaping travel behavior. Autonomous vehicles—particularly when

integrated with shared mobility—hold the potential to alleviate or even resolve pressing urban issues such as traffic congestion, road accidents and inefficient land use.

Many studies address the coexistence of conventional human-driven vehicles and autonomous vehicles with different proportions of traffic. Rey and Levin proposed a new stochastic traffic control policy that accounts for both autonomous vehicles and conventional ones, assuming that autonomous vehicles travel in dedicated lanes and have separate or reserved signal phases at intersections, similar to cyclists [34]. Paper [35] studied a scenario where both autonomous and conventional vehicles share the same road. The autonomous vehicles are controlled by a centralized system, while human drivers dynamically react to traffic conditions. The autonomous vehicle management system involves a control policy designed to indirectly influence the decisions of conventional vehicle drivers in order to minimize congestion [35].

CAVs are capable of forming platoons, thereby reducing headways and improving road capacity. In mixed traffic, however - where RVs and CAVs operate together—platoon intensity is strongly affected by the stochastic arrangement of the two vehicle types. As a result, road capacity remains highly uncertain even under identical flow conditions. To address this challenge, ref. [36] introduces a mixed traffic assignment model designed for worst-case scenarios. The model estimates the minimum network capacity and its corresponding equilibrium flow. This assignment problem is formulated as a bilevel programming approach: the lower-level component is expressed as a variational inequality problem to compute equilibrium flows under fixed link capacities, while the upper-level component determines the optimal inputs for all link capacities within their ranges to minimize overall network performance. To solve this bilevel worst-case assignment problem, the authors propose a partition-based norm-relaxed feasible direction algorithm.

Despite the richness of these studies, several knowledge gaps remain. There is no standardized calibration for RVs and AVs behavioral models, leading to inconsistent findings across studies. For example, some studies [32,34,35] suggest substantial traffic stability improvements at moderate AVs penetration, whereas others [22,24,37,38] report negligible or negative impacts under conservative AVs settings. Additionally, few studies account for spatial–temporal correlations in network-level traffic flow, or systematically examine threshold effects, where mixed traffic behavior changes abruptly.

Recent works introduce advanced modeling approaches that could address these gaps. Jorge A. Laval [39] presents a fluid-dynamics perspective on traffic flow, while Eun Hak Lee and Euntak Lee [40] propose a congestion boundary approach to detect phase transitions in traffic flow. Zhao Liu et al. [41] uses a spatial–temporal graph convolution network model for traffic prediction, and Oikonomou et al. [42] simulate CAV safety from conflicts to crashes. These studies provide frameworks that can be applied to understand complex mixed traffic interactions more systematically.

Our study builds upon these approaches by evaluating the impact of AV-ACC functionality in a complex signalized intersection under mixed traffic conditions. Using field traffic data as input for a microsimulation model in AIMSUN, we analyzed scenarios with varying demand levels and AV-ACC penetration rates, highlighting how increased penetration—particularly beyond 50%—affects traffic stability and efficiency.

(3) Research has also examined control techniques that can optimize CAVs performance and safety. Paper [43] analyzed control techniques for CAVs, which are essential for vehicle safety, passenger comfort, transport efficiency and energy savings. It provides a comprehensive overview of vehicle control technology, starting from trajectory prediction for a single vehicle (microscopic level) to the control of a group of connected vehicles (macroscopic level).

In road transport, human error accounts for nearly 90% of accidents. Vehicle automation contributes to improved safety by mitigating driver fatigue and errors, while also enhancing overall mobility efficiency. However, the large-scale adoption of CAVs will unfold over several decades, during which they must coexist with RVs. In [44], the mobility and safety performance of CAVs under mixed traffic conditions is assessed using the cumulative-anticipative car-following (CACF) model. The study compares the CACF approach with the widely used Wiedemann 99 model and cooperative adaptive cruise control (CACC), within the VISSIM simulation platform. The simulations cover both single-lane and multi-lane scenarios and include sensitivity analyses for key mobility and safety parameters. Additionally, Paper [45] analyzed the impact of AVs on urban traffic network capacity through an experimental microscopic traffic simulation using the SUMO platform. The findings indicate that, for both grid and real-world networks, traffic capacity exhibits a quasi-linear growth as the penetration rate of AVs increases.

Future challenges include potential interactions between AVs and RVs, namely vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication [46,47]. Few studies systematically link control strategies with emergent network-level traffic patterns, leaving uncertainty on how microscopic vehicle control translates into macroscopic flow stability. Advanced AI-based models offer opportunities to bridge this gap. Zhao Liu et al. [41] propose a spatial-temporal graph convolution network model informed by traffic fundamental diagram data, and Oikonomou et al. [42] simulate macroscopic safety outcomes from conflicts to crashes. Integrating these approaches can enhance the predictive power and reliability of AVs control in mixed traffic scenarios.

(4) The management of vehicle traffic at urban intersections is mainly approached in two ways. One is the management of traffic control at isolated intersections, or local treatment, and the other is the management of traffic over a traffic zone, or global treatment. In the first approach, the traffic signal cycle durations are sized independently for each intersection in order to minimize delay times, while in the second, decision variables for all intersections are set to minimize the users' travel time throughout the network. These methods are addressed in study [48]. In the case of mixed traffic, consisting of both AVs and RVs, isolated intersection control makes sense. Simulating a fleet of AVs under real traffic conditions requires a dynamic approach since vehicle routes may change over time due to traffic conditions.

(5) Finally, several studies focus on the actual effects of AVs penetration rates on traffic performance. There are studies [22,24,37,38,49] that show that mixed traffic involving low-level AVs will initially have a slight negative effect on traffic flow and road capacities. Both ref. [22] and ref. [24] emphasize the fact that AVs can improve traffic capacity and stability, but their impact depends on factors like penetration rate, AVs settings, traffic volume and human behavior. According to those studies, significant benefits are seen when the penetration rate exceeds 40%. In study [37], the authors conclude that AV-ACC systems tend to have, on average, longer time gaps compared to RVs, a finding supported by other European Commission-funded studies such as [38], which also observed comparable or even longer reaction times in AV-ACC than in RVs.

Despite multiple empirical and simulation studies, clear thresholds and conditions for stability improvement remain debated. Contradictions in reported outcomes arise from differences in behavioral assumptions, AVs control logic and network topology.

Our study contributes by systematically analyzing AV-ACC penetration rates in a complex signalized intersection using AIMSUN microsimulation, identifying how varying AV-ACC shares affect traffic performance under both current and 20% growth demand scenarios. This addresses unresolved questions in the literature regarding the threshold effects

of partially automated vehicles and provides actionable insights for traffic management and intersection design.

3. Materials and Methods

The overall methodological framework employed in this study is illustrated in Figure 1. The experimental design was developed in accordance with the study's main objectives, aiming to assess the effects of increasing levels of partially automated vehicles on intersection traffic performance. Simulation scenarios were created by varying key behavioral parameters influencing vehicle interactions and driver decision-making processes, along with adjustments in traffic volumes, signal control plans, and the proportion of AV-ACC. The simulation was carried out using a microsimulation model developed in AIMSUN (version 8.2.1), calibrated with field data collected from inductive loop detectors.

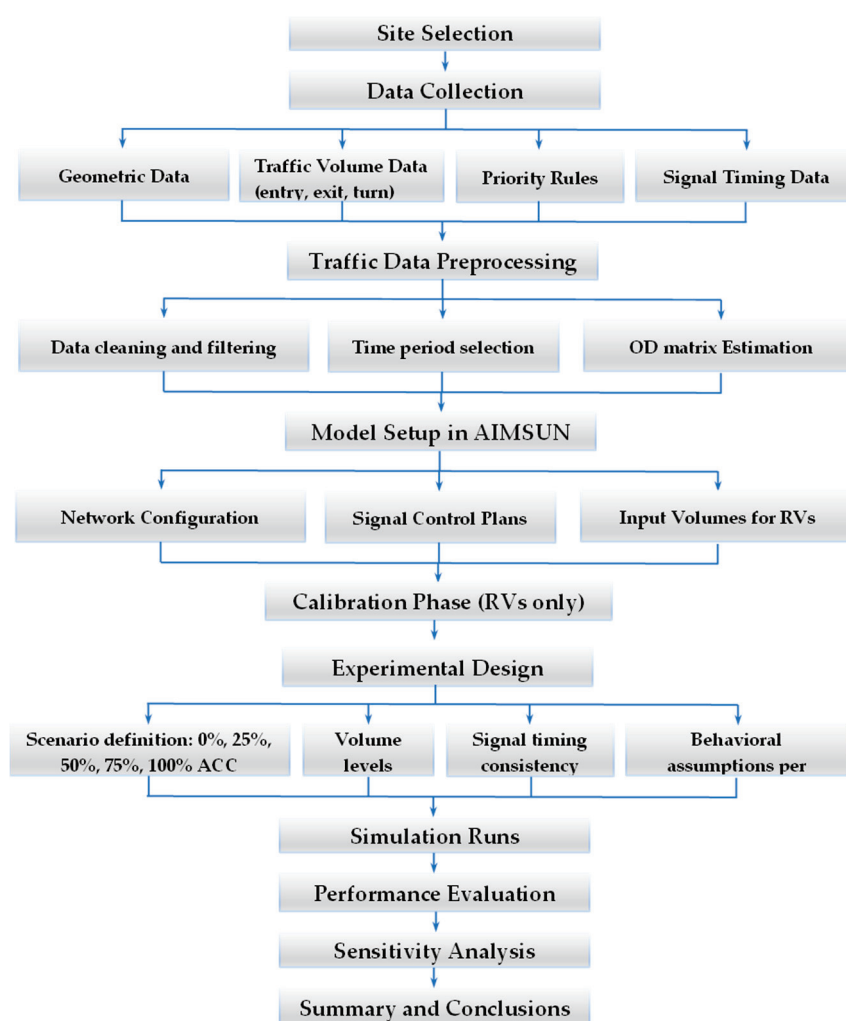


Figure 1. General methodology of the study.

After setting up the simulation environment, all relevant inputs were assigned to each scenario. Each configuration incorporated specific combinations of demand levels, signal timings, and behavior models for both RVs and AV-ACC. Once the scenarios were defined, the simulations were executed, and key performance metrics—including intersection throughput, average vehicle delay and travel speed—were recorded. Finally, the results were compiled, analyzed and interpreted in accordance with the study's main objectives. Each component of the methodology is described in detail in the following sub-sections.

In urban traffic environments, the primary behavioral sub-models implemented in AIMSUN include car-following, lane-changing, gap acceptance during lane changes, gap acceptance for yielding and anticipatory (look-ahead) behavior. An overview of the main parameters defining the behavioral sub-models in AIMSUN is presented in Table 1.

Table 1. Key behavioral parameters in AIMSUN: definitions and functional influence.

Parameter	Definition	Influences
Length	Vehicle length	Effective vehicle length in Car following
Max. Desired Speed	Minimum between the vehicle's assigned maximum speed and the product of its speed acceptance factor and the legal speed limit of the road segment	Car following, lane changing, travel time, queue discharge
Speed Acceptance	Degree of acceptance of the speed limits	Car following, travel time, queue discharge
Max. Acceleration	The highest value that the vehicle can achieve under any circumstances	Car following, lane changing, travel time, queue discharge
Normal Deceleration	The maximum deceleration that the vehicle can use under normal conditions	Car following, lane changing, travel time, queue discharge
Maximum Deceleration	The most severe braking can be applied under special circumstances	Lane changing, travel time, queue discharge
Headway Aggressiveness	Allows vehicles to enter shorter gaps without forcing the rear vehicle to brake, followed by a relaxation process to gradually recover the stability	Car following, lane changing, travel time, queue discharge
Clearance	Distance that vehicle keeps with the preceding one when stopped	Effective vehicle length in car following, capacity, queue length
Maximum give-way time	Time after which the vehicle becomes more aggressive in yield situation	Lane changing, gap acceptance
Sensitivity Factor	How much the vehicle could be sensitive to the deceleration of the leader	Deceleration component of car following
Imprudent lane changing cases	Defines whether a vehicle will still change lane after assessing an unsafe gap	Lane changing, overtaking
Maximum Speed Difference	Highest speed difference allowed between a vehicle and another in the target lane for a safe lane change to occur.	Two-lane car following model
Reaction Time	Delay of a moving vehicle in responding to speed or acceleration changes in the preceding vehicle	All internal models, sections and on-ramp capacities
Reaction Time at Stop	Time it takes for a stationary vehicle to begin moving once the vehicle in front starts accelerating.	All internal models, stop and go capacity
Reaction time at traffic light	Delay between the green phase initiation and the moment when the first vehicle in the queue begins to move.	All internal models
Queue up speed	Maximum speed below which a vehicle is considered to be in a queue	Queue measures
Queue leaving speed	Minimum speed at which a vehicle is considered to have exited a queue after being stopped.	Queue measures

The car-following model is central to representing longitudinal vehicle interactions, as it determines the clearance between a leading and a following vehicle. In AIMSUN, this model is based on the well-established Gipps model [50,51] and has been extended to incorporate localized behavioral parameters. These parameters vary depending on several contextual factors such as driver type (e.g., compliance with speed limits), roadway geometry (e.g., speed limits per segment or turn), and the influence of neighboring vehicles. The model comprises two components: acceleration, which simulates the effort to attain a desired speed under free-flow conditions, and deceleration, which reflects the behavioral adaptation to the leading vehicle's speed and position, thereby regulating longitudinal headway for safety and stability.

The implementation of AV-ACC in Aimsun 8.2.1 is limited, as it does not support vehicle platooning functionalities or explicitly model V2V communication latency, which should be considered when interpreting the results. Instead, AV-ACC behavior is represented through individual car-following adjustments such as reaction time, headway and deceleration rate.

A critical factor influencing car-following behavior is the reaction time, which reflects the temporal delay between the perception of a stimulus and the corresponding driving action. In AIMSUN, this parameter is further disaggregated into three distinct types to capture specific traffic conditions and behavioral responses: Reaction time, Reaction time at stop, Reaction time at traffic light. These parameters have a significant influence on traffic flow performance, particularly under congested or signalized conditions.

In the calibration phase, four candidate configurations (C1–C4) were defined to replicate, as accurately as possible, the behavior of regular vehicles (RVs) within the studied complex intersection. This was achieved by adjusting the three reaction time parameters, while the remaining behavioral parameters—related to car-following, lane-changing, and driver characteristics—were set according to values reported in the literature to ensure consistency with empirically observed driving behavior. The GEH statistics were employed as the calibration metric to assess the consistency between simulated and observed traffic volumes at both the entry and exit points of the intersection and are calculated using the following formula:

$$GEH = \sqrt{\frac{2(Q_m - Q_c)^2}{Q_m + Q_c}} \quad (1)$$

where Q_m —output traffic volume from the simulation model (vehicles per hour); Q_c —real-world hourly traffic count (vehicles per hour).

According to widely accepted guidelines, a GEH value less than 5% is generally considered indicative of a good fit between simulated and real-world traffic counts [52,53].

In addition to the GEH statistic, the coefficient of determination R^2 can be calculated to further assess calibration quality. Values for R^2 within the range 0.90–1.00 are commonly considered acceptable for traffic model calibration [54].

Within the simulation framework, the experimental design component was structured to investigate the influence of partially automated vehicles interacting with regular vehicles in the considered scenarios. A series of simulation configurations were defined by systematically varying the penetration rate of AV-ACC (25%, 50%, 75%, and 100%). For each scenario, key performance indicators—including the average values of delay, queue, flow and speed—were extracted to facilitate comparative analysis.

4. Case Study

4.1. Study Area

In the last decade, research on road safety and traffic risk assessment has concluded that identifying locations with frequent accidents is essential to understanding their underlying causes [55].

The case study focuses on the intersection ranked number one in Bucharest in the black spot classification (according to the data from Traffic Police Department): Splaiul Independenței × Bd. Doina Cornea × Orhideelor Street (Grozăvești Intersection) (Figure 2). The Basarab overpass is not included in the analysis.

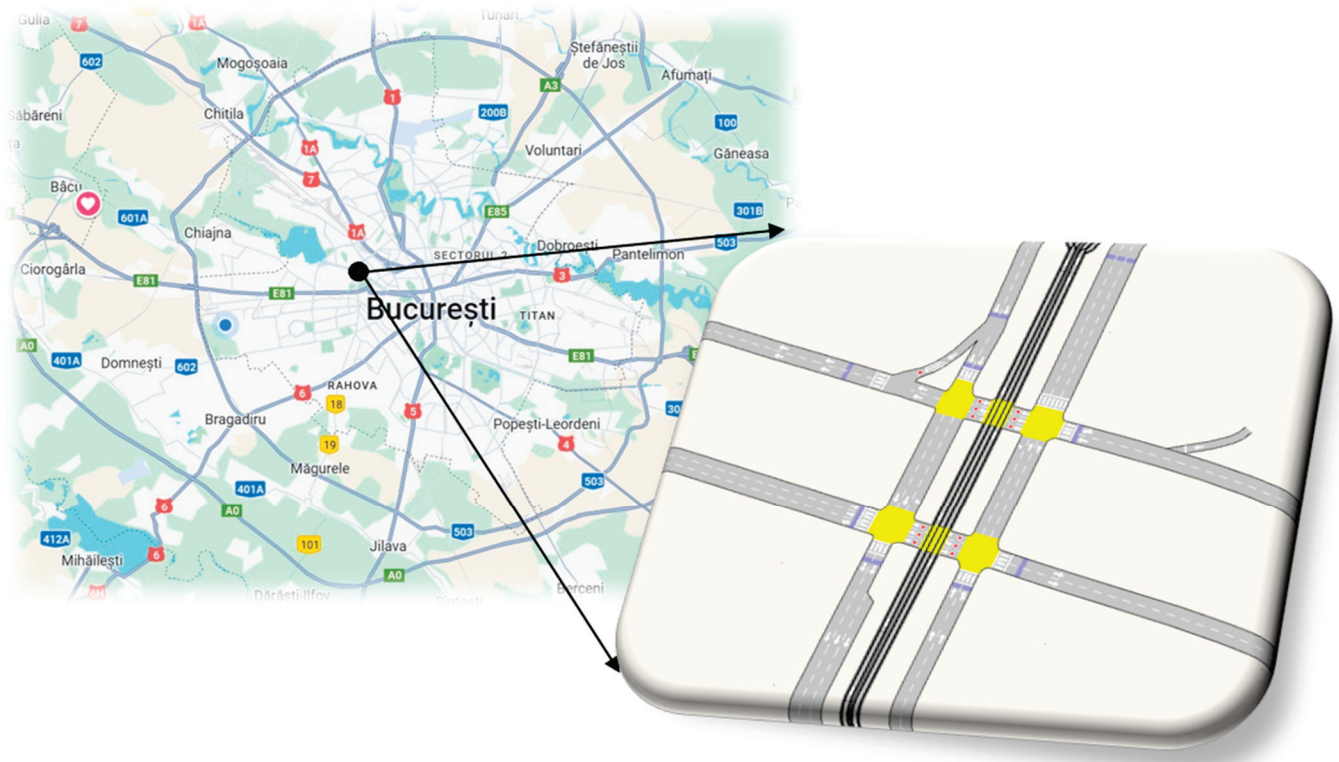


Figure 2. Splaiul Independenței × Bd. Doina Cornea × Orhideelor Street intersection: position and geometry.

The Grozăvești intersection is a complex junction with a central circular island crossed by a tram line. Traffic is signalized and the intersection is classified as Class I in terms of functionality.

The roads that form the intersection are:

- Splaiul Independenței—a category I arterial road, separated by the Dâmbovița River, having 2 traffic lanes per direction (approx. 3.00–3.50 m/lane), excepting the West access that has 3 traffic lanes;
- Bd. Doina Cornea—a category II road, divided by a central tram corridor, having 2 lanes per direction (approx. 3.40–3.50 m/lane),
- Orhideelor Street—a category II road, having 2 lanes per direction (approx. 3.40–3.50 m/lane).

Traffic priority is marked along the East–West axis (Splaiul Independenței).

For the traffic analysis, directional traffic counts were carried out. Traffic counts were performed at the intersection using inductive loop detectors installed at each entry and exit point (Figure 2).

Four measurement zones were defined within the intersection, each equipped with inductive loop detectors. One set of loops was assigned to each traffic conflict subzone, as shown in the plan below (Figure 3). Traffic counts were initially collected using inductive loop detectors over a one-week period (31 March–5 April 2025) for 24 h per day. However, due to a malfunction of one of the loops starting on 1 April, only the data from Monday, 31 March 2025, were used for analysis. Measurements were taken under mild weather conditions (19 °C, partly sunny). Table 2 provides an overview of the traffic flows considered in the analysis.

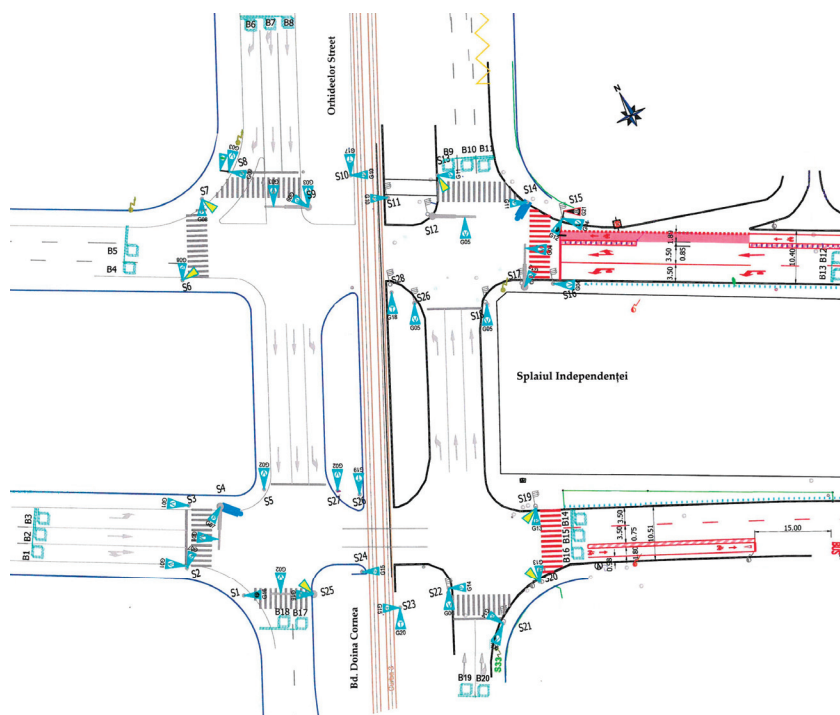


Figure 3. Traffic signal numbering plan and inductive loop coding.

Table 2. Input traffic flows by road segment and turn direction at the studied intersection.

Road Segment	Turn Direction	Toward	%	Vehicles/h	Total Vehicles/h
Splaiul Independenței (West)	Right	to Bd. Doina Cornea	6%	62	1127
	Straight	to Splaiul Independenței (East)	57%	648	
	Left	to Orhideelor Street	37%	417	
Orhideelor Street (North)	Right	to Splaiul Independenței (West)	36%	267	750
	Straight	to Bd. Doina Cornea	31%	231	
	Left	to Splaiul Independenței (East)	34%	252	
Splaiul Independenței (East)	Right	to Orhideelor Street	21%	185	892
	Straight	to Splaiul Independenței (West)	60%	535	
	Left	to Bd. Doina Cornea	19%	172	
Bd. Doina Cornea (South)	Right	to Splaiul Independenței (East)	9%	77	821
	Straight	to Orhideelor Street	71%	585	
	Left	to Splaiul Independenței (West)	19%	159	

From a traffic signalization perspective, the intersection currently operates as shown in the diagram in Figure 4. The traffic signal control plan has a cycle duration of 120 s and consists of four signal groups, each managing a specific direction or movement:

- Signal 1 controls the North (Orhideelor Street)–South (Bd. Doina Cornea) movement, with a green time of 35 s;
- Signal 2 controls the East–West (Splaiul Independenței) movement, with a green time of 33 s;
- Signal 3 manages the central lanes, providing a green time of 63 s;
- Signal 4 is dedicated to public transport (tram line) and has a green time of 15 s.

The diagram shows how these signal phases are interleaved to minimize conflicts and support efficient traffic flow. The overlapping of Signals 1 and 3 indicates the coordination of parallel flows along the NS direction.

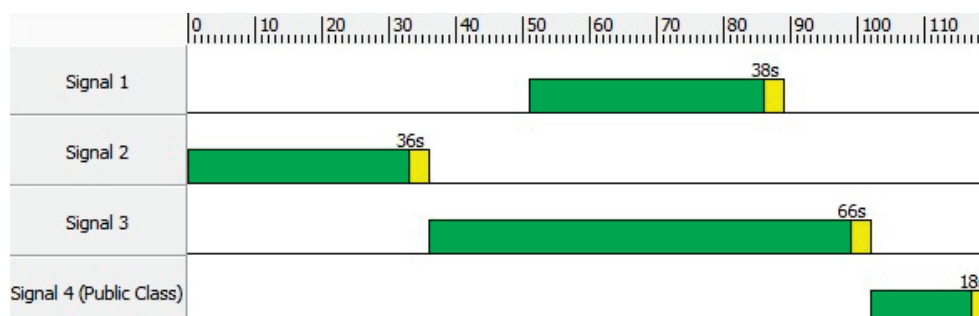


Figure 4. Traffic signal phasing and green time durations at the studied intersection.

To better understand the diagram presented in Figure 4, Figure 5 highlights the physical layout of the intersection, emphasizing the green light phases corresponding to Signals 1 to 4.

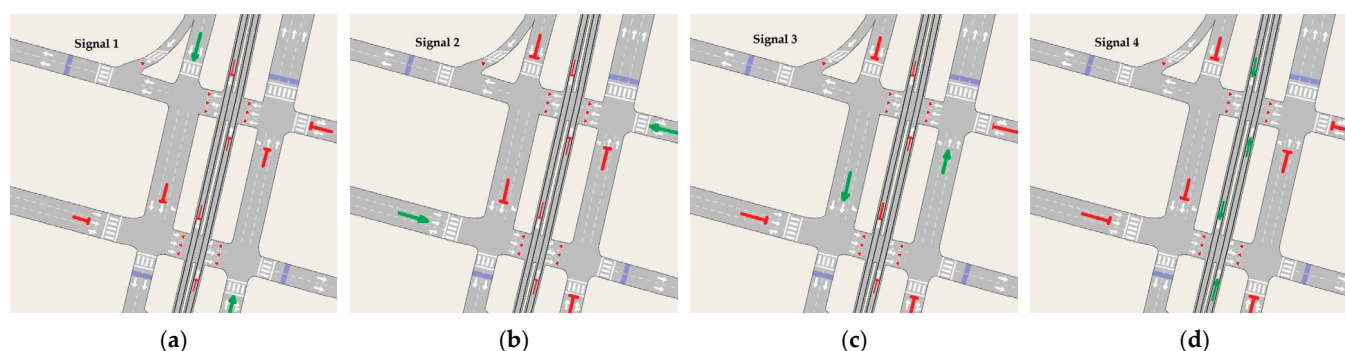


Figure 5. Green light phases corresponding to (a) Signal 1; (b) Signal 2; (c) Signal 3; (d) Signal 4.

The right turn from Orhideelor Street is excluded from signal control and is made via a dedicated slip lane.

The simulation experiments in AIMSUN for the configurations (C1–C4) were conducted under the following assumptions:

- the values assigned to the behavioral parameters are presented in Table 3;
- the warm-up period is set to 15 min, which is considered sufficient to mitigate the effects of the initial conditions and to allow the system to reach a stationary state;
- the simulation step is set to 0.1 s.
- the simulation duration is one hour, corresponding to the morning peak period between 07:00 and 08:00 AM.

For RVs, reaction times were selected based on typical ranges reported in empirical studies, where general reaction times vary between 0.8 and 1.2 s, reaction times at stop range from 1.0 to 1.5 s, and reaction times at traffic lights can reach 1.5 to 2.0 s, reflecting attention lapses and behavioral variability [56,57].

Each configuration was evaluated by comparing simulated and observed traffic volumes using the GEH statistics and the results are summarized in Table 4.

Among the four simulated configurations (C1–C4), only Configuration C4 yielded GEH values below 5% for all entry and exit detectors, thereby satisfying the generally accepted calibration threshold (which require 100% of the GEH values to fall below 10 and at least 85% below 5 [58]) across the network. In contrast, configurations C1 through C3 exhibited multiple GEH values exceeding this threshold, particularly at the western and eastern access points, indicating suboptimal alignment with observed traffic volumes. As such, Configuration C4 was selected as the calibrated base model for subsequent experimental analysis.

Table 3. Input parameters for RV simulation.

Vehicle Attributes	Values Mean $\pm$ SD [Min, Max]			
Length (m)	4.5 $\pm$ 0.3 [4, 5]			
Max. Desired Speed (km/h)	50 $\pm$ 10 [30, 60]			
Speed Acceptance	1.1 $\pm$ 0.1 [0.9, 1.3]			
Max. Acceleration (m/s ²)	2.5 $\pm$ 0.25 [2, 3]			
Normal Deceleration (m/s ²)	3.5 $\pm$ 0.25 [3, 4]			
Maximum Deceleration (m/s ²)	6 $\pm$ 1 [5, 7]			
Headway Aggressiveness	0.3 $\pm$ 0.1 [0.1, 0.5]			
Clearance	1 $\pm$ 0.3 [0.9, 1.3]			
Maximum give-way time (s)	10 $\pm$ 2.5 [5, 15]			
Sensitivity Factor	1 $\pm$ 0.05 [0.9, 1.2]			
Imprudent lane changing cases	Yes			
Maximum Speed Difference (km/h)	40			
	C1	C2	C3	C4
Reaction Time (s)	1	1	1	0.8
Reaction Time at Stop (s)	1.2	1.2	1.5	1.2
Reaction time at traffic light (s)	1.8	2	2	1.8
Queue up speed (m/s)	1.10			
Queue leaving speed (m/s)	3.70			

Table 4. GEH Statistic Values for Input (IN) and Output (OUT) Sections of the Analyzed Area (volumes expressed in veh/h).

Access	Simulation Traffic Volumes				Observed Traffic Volumes	GEH			
	C1	C2	C3	C4		C1	C2	C3	C4
W-IN	902	910	869	1075	1127	7.06	6.80	8.17	1.57
N-IN	751	751	751	751	750	0.04	0.04	0.04	0.04
E-IN	696	671	684	766	892	6.96	7.91	7.41	4.38
S-IN	818	818	818	818	821	0.10	0.10	0.10	0.10
W-OUT	848	834	833	889	988	4.62	5.10	5.14	3.23
N-OUT	1040	1035	1035	1122	1271	6.80	6.95	6.95	4.31
E-OUT	851	857	822	935	983	4.36	4.15	5.36	1.55
S-OUT	407	403	402	427	465	2.78	2.98	3.03	1.80

In addition to the GEH statistic, the coefficient of determination (R^2) was calculated to further assess calibration quality. The resulting value of $R^2 = 0.958$ (Figure 6) indicates a very high level of concordance between observed and simulated traffic volumes. In the literature, values within the range 0.90–1.00 are commonly considered acceptable for traffic model calibration [54].

Combined with the fact that all GEH values fall below 5, this result confirms that Configuration C4 provides a reliable calibrated base model.

To evaluate the robustness of the calibrated model, a validation procedure was carried out using traffic volumes from an independent time period (12:00–13:00). The comparison between observed and simulated volumes was assessed by means of the GEH statistic, which is widely used for traffic modeling validation. The results, summarized in Table 5, indicate that all GEH values are below the commonly accepted threshold of 5, confirming a good level of agreement between the microsimulation outputs and field measurements.

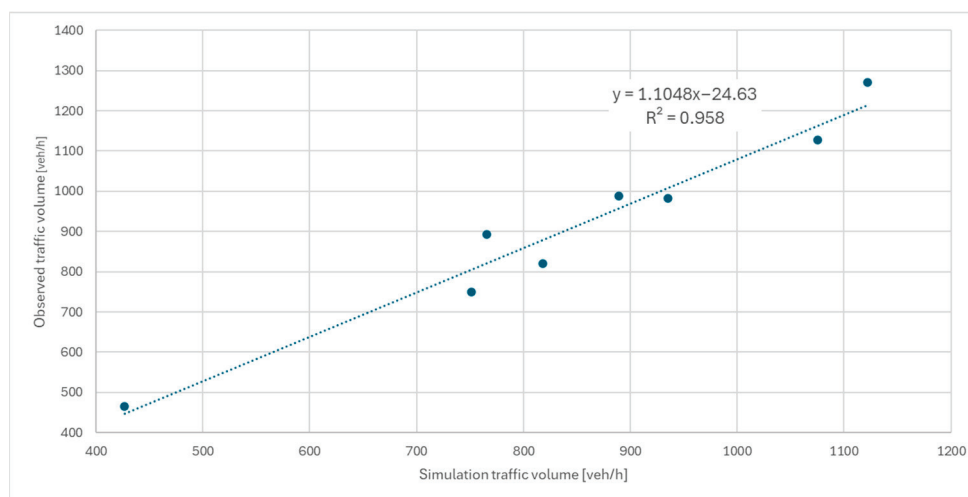


Figure 6. Observed vs. simulated traffic volumes.

Table 5. Validation results based on GEH statistic for traffic volumes (expressed in veh/h).

Access	Simulation Traffic Volumes	Observed Traffic Volumes	GEH
W-IN	1014	1164	4.54
N-IN	725	721	0.15
E-IN	774	810	1.28
S-IN	814	816	0.07
W-OUT	846	881	1.19
N-OUT	1088	1007	2.50
E-OUT	838	982	4.77
S-OUT	517	559	1.81

While the validation presented in Table 5 is based on traffic volumes, future research will include the evaluation of additional performance indicators such as speed, delay, and queue length, in order to provide a more comprehensive assessment of model accuracy.

4.2. Microsimulation Scenarios and Results

The experiment was structured to analyze the dynamic behavior of partially automated vehicles in mixed traffic conditions, focusing on longitudinal motion, lane-changing interactions, and traffic flow stability. Scenarios were developed to reflect varying levels of ACC-equipped vehicle penetration within the traffic stream composition - specifically 25%, 50%, 75%, and 100%. After configuring the simulation environment in AIMSUN and assigning the prepared input data summarized in Tables 2 and 4, simulations were conducted for each case. Key traffic performance indicators, such as intersection throughput, average delay, queue and speed, were collected to evaluate the influence of automation on operational efficiency.

Two test scenarios were examined reflecting different traffic conditions:

- A base scenario: this scenario represents the current traffic conditions.
- A growth scenario: assuming a 20% increase in traffic demand while keeping the network geometry and signal settings unchanged.

For both scenarios, the analysis focused on a one-hour morning peak period (7:00–8:00 a.m.), across five levels of ACC-equipped vehicle penetration (0%, 25%, 50%, 75%, and 100%). Each case incorporated a distinct set of behavioral parameters to differentiate between conventional RVs and partially automated vehicles (ACC-equipped), as

implemented in the AIMSUN microsimulation platform (see Table 6). Each scenario was run for ten stochastic replications. Flows were aggregated in 2-minute bins.

Table 6. Input parameters for partially automated vehicles simulation.

Parameter	Values Mean $\pm$ SD [Min, Max]
Length (m)	4.5 $\pm$ 0.3 [4, 5]
Max. Desired Speed (km/h)	54 $\pm$ 5 [50, 58]
Speed Acceptance	1 $\pm$ 0 [1, 1]
Max. Acceleration (m/s ²)	2.5 $\pm$ 0.25 [2, 3]
Normal Deceleration (m/s ²)	3.25 $\pm$ 0.25 [3, 3.5]
Maximum Deceleration (m/s ²)	5.2 $\pm$ 0.5 [4.5, 6]
Headway Aggressiveness	0.05 $\pm$ 0.05 [0, 0.1]
Clearance	1 $\pm$ 0 [1, 1]
Maximum give-way time (s)	10 $\pm$ 2.5 [5, 15]
Sensitivity Factor	1 $\pm$ 0.05 [0.9, 1.1]
Imprudent lane changing cases	No
Maximum Speed Difference (km/h)	30
Reaction Time (s)	0.7
Reaction Time at Stop (s)	1
Reaction time at traffic light (s)	0.9
Queue up speed (m/s)	1.00
Queue leaving speed (m/s)	3.50

For SAE Level 2 vehicles equipped with Adaptive Cruise Control (ACC), reaction times were set to 0.7 s during car-following, 1.0 s when resuming from a stop, and 0.9 s in response to traffic signals. These values reflect the faster and more consistent response behavior of ACC systems, as documented in studies such as [38,59].

In congested urban environments, RVs tend to exhibit higher queue entry speeds (approximately 1.10 m/s), primarily due to their greater behavioral variability and longer reaction times. In contrast, vehicles equipped with Adaptive Cruise Control (ACC) adopt a more conservative and stable car-following behavior, resulting in lower queue entry speeds of approximately 1.00 m/s. Furthermore, while RVs typically accelerate more abruptly when exiting a queue (≈ 3.70 m/s), ACC-equipped vehicles apply smoother and more controlled acceleration, leading to slightly lower queue exit speeds (around 3.50 m/s), influenced by their shorter reaction times at stop and reduced headway aggressiveness [60]. Headway Aggressiveness value of 0.05 was adopted, reflecting the conservative behavior of commercially available Adaptive Cruise Control (ACC) systems. This parameter setting implies that vehicles maintain relatively large time headways, prioritizing safety and stability over roadway capacity.

The variation in key performance indicators (delay time, queue, flow and speed) is shown in Figures 7–9 for base scenario and in Figures 10 and 11 for growth scenario. In both scenarios we considered the increasing of AV-ACC vehicles penetration (0%, 25%, 50%, 75% and 100%).

In base scenario, traffic conditions are the least efficient for RV. As the proportion of AV-ACC vehicles increases (from 25% to 100%), both the delay time and queue length consistently decrease and both flow and speed display a progressive increase, demonstrating that a higher penetration of AV-ACC vehicles improve traffic performance.

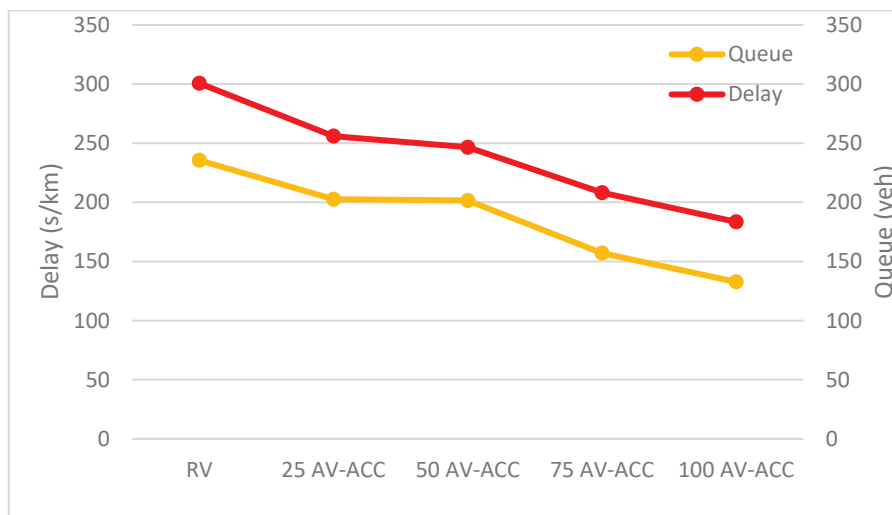


Figure 7. Variation in delay time and mean queue in the base scenario under different AV-ACC penetration rates.

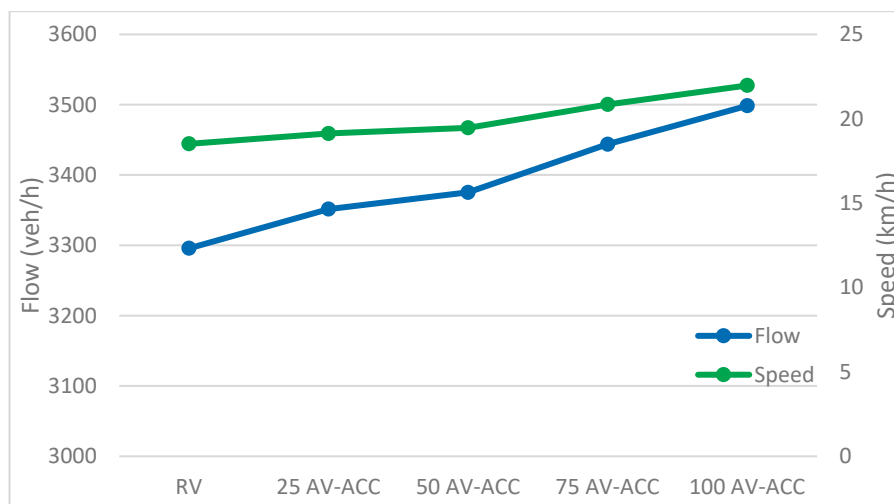


Figure 8. Variation in average speed and traffic flow in the base scenario under different AV-ACC penetration rates.

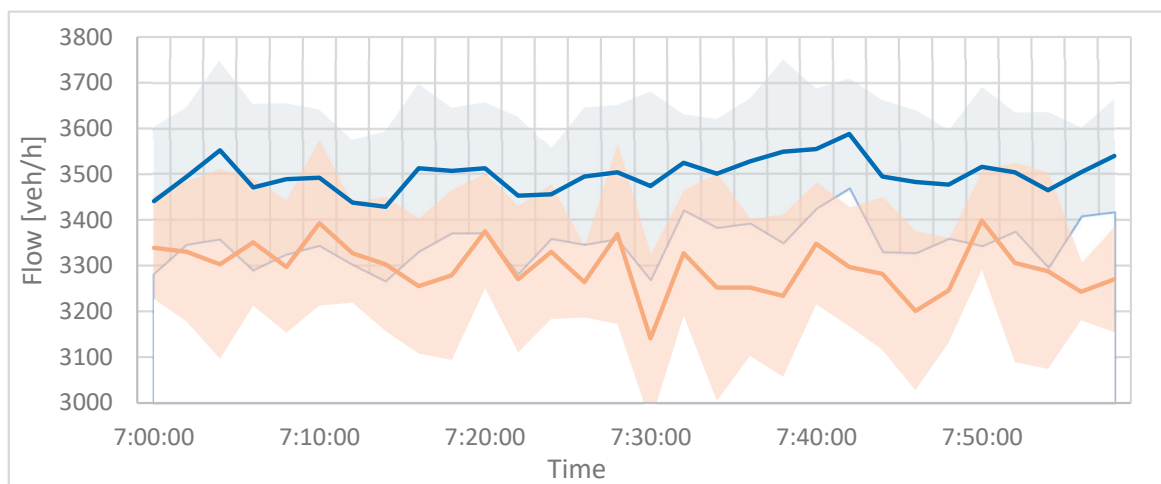


Figure 9. Mean flow $\pm$ SD over time (RV—orange line, AV-ACC—blue line).

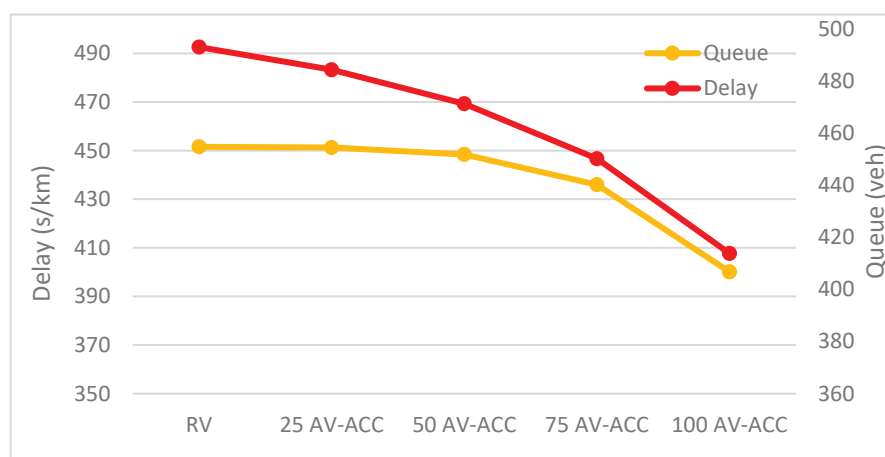


Figure 10. Variation in delay time and mean queue in the growth scenario under different AV-ACC penetration rates.

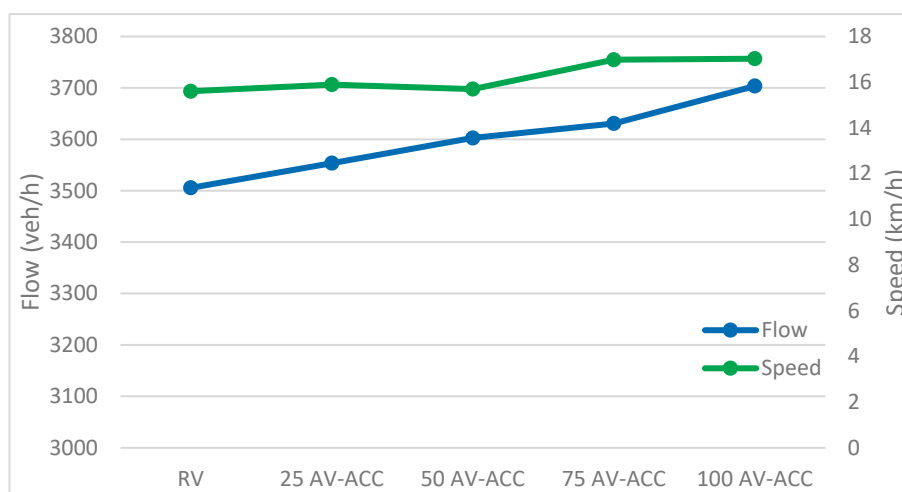


Figure 11. Variation in average speed and traffic flow in the growth scenario under different AV-ACC penetration rates.

Delay time decreases by 17.99% for mixed traffic (50% AV-ACC) and by 38.97% for full AV-ACC traffic conditions. The queue also decreases by 14.47% for mixed traffic and by 43.68% for full AV-ACC traffic. The speed increases by 5.11% for mixed traffic and by 18.70% for full AV-ACC traffic. Significant improvements in traffic performance are observed when AV-ACC penetration exceeds 75%, underlining the impact of higher automation levels in reducing congestion and improving traffic flow efficiency.

The flow increases by 2.41% for mixed traffic and by 6.15% for full AV-ACC traffic.

To evaluate stability, we analyzed traffic flow using variance and time-series methods. We simulated two scenarios—RV and AV-ACC—each with 10 independent replications over a 1-hour window. At scenario level we report mean across replications, standard deviation (SD), variance, coefficient of variation (CV) and 95% confidence intervals (CI). Equality of variances was assessed with Levene's test and differences in means with Welch's *t*-test.

We obtain the following results:

RV: 3295.7 veh/h (SD = 42.8; CV = 1.30%; 95% CI [3265.1, 3326.3]).

AV-ACC: 3498.4 veh/h (SD = 31.6; CV = 0.90%; 95% CI [3475.8, 3521.0]).

The mean increase is +202.7 veh/h, which corresponds to +6.15% for AV-ACC relative to RV.

Levene's test indicates no significant difference in variances (stat = 1.6128, $p = 0.2203$); the RV/AV-ACC variance ratio is 1.832. Welch's *t*-test shows a highly significant difference

in means ($t = 12.0596$, $p = 1.27 \times 10^{-9}$). Mean $\pm$ SD bands are narrow for both scenarios (Figure 9); the $\Delta(t)$ series is consistently positive; and the ~ 10 -minute rolling SD indicates a stable process with no abrupt variability spikes.

Within the simulated conditions, AV-ACC delivers a robust $\approx 6\%$ increase in hourly flow together with lower relative dispersion (CV) and stable temporal behavior.

Figures 10 and 11 show the variation in key performance indicators (delay time, queue, flow, and speed) for the growth scenario, with AV-ACC penetration levels of 0%, 25%, 50%, 75%, and 100%.

In the growth scenario, traffic conditions are also the least efficient for RV. Delay time decreases by 4.41% for mixed traffic and by 16.08% for full AV-ACC traffic conditions. The queue decreases by 0.70% for mixed traffic and by 11.41% for full AV-ACC traffic. For AV-ACC penetration exceeding 50%, traffic flow becomes more homogeneous, allowing coordinated vehicle behavior that leads to reduced queues. The speed increases only by 1% under mixed traffic conditions and by 9.12% under full AV-ACC traffic. The flow increases by 2.76% for mixed traffic and by 5.65% for full AV-ACC traffic.

The evolution of queue length with increasing AV-ACC penetration does not follow a strictly monotonic pattern at lower adoption levels ($\leq 50\%$). This nonlinear variation can be attributed to the interaction mechanisms between heterogeneous driver types. At low penetration rates, AV-ACC vehicles are still strongly influenced by the surrounding regular vehicles, which reduces their stabilizing potential and may even introduce local fluctuations in queue dynamics. Only when penetration levels exceed 50% does the stabilizing influence of AV-ACC become dominant, resulting in a more consistent reduction in both delay time and queue length.

All reported performance indicators (delay time, queue length, flow and speed) were calculated as network-wide averages aggregated over all approaches of the modeled intersection during the one-hour simulation period.

Under higher traffic volumes (growth scenario), variations in traffic performance indicators are smaller compared to those obtained in lower traffic volumes (base scenario).

5. Conclusions

This paper investigates the AV-ACC implementation impacts on the traffic performance in a complex urban intersection, under a mixed traffic environment with RVs. The full or partial implementation of AV-ACC offers several positive impacts that improve the efficiency of the traffic flow in the intersection.

A complex signalized intersection was selected, and field data was collected to use as an input in AIMSUN microsimulation software. In the calibration phase, four configurations were defined to replicate the behavior of RVs. This was achieved by adjusting reaction time in motion, reaction time at stop and reaction time at traffic light, while the remaining behavioral parameters—car-following, lane-changing, driver characteristics—were set according to the values reported in the literature. We chose the configuration with GEH values less than 5% as the calibrated base scenario.

Two traffic demand scenarios—a base scenario and a 20% growth scenario—were analyzed to assess the impact of ACC-equipped vehicles under varying percentages of AV-ACC in traffic composition. The results demonstrated that increasing AV-ACC penetration rates improved traffic performance across both scenarios, with more significant benefits observed under current traffic conditions compared to the growth scenario.

For the base scenario, all traffic performance indicators are improved for mixed traffic. For growth scenarios, mixed traffic for penetration rates below 50% involving automated vehicles with ACC will initially have a slight negative effect on traffic flow. The results

of the simulations show that AV-ACC can improve traffic capacity and stability when the penetration rate exceeds 50%.

It should be noted that the performance improvements reported in this study are specific to the fixed signal control plan applied in all scenarios. This constraint was deliberately imposed to isolate the influence of AV driving behavior from that of signal timing optimization. While this approach enhances the internal validity of the results, it also limits their direct applicability to other signal settings. In practice, the magnitude of benefits may differ under alternative control strategies, such as adaptive or traffic-responsive systems. Future research should therefore investigate how varying AV penetration rates interact with different signal control policies to assess the combined effects on intersection performance.

The study was limited to AIMSUN microsimulation software. It is recommended that further evaluations be conducted with other microsimulation models in order to confirm findings and establish principles for the effective use of these models in traffic engineering. Also, it is recommended that behavioral parameters related to car-following, lane-changing and driver characteristics be measured in future research. Collecting real-world data on these parameters would allow a more accurate calibration and validation of simulation models, thereby enhancing their reliability when applied to complex traffic environments.

Although comparative methods were used to maintain interpretability in a controlled simulation setting, future studies could employ machine learning techniques to capture potential nonlinearities and complex interactions among AV behavioral parameters.

Future research should also extend the analysis to multiple intersections and corridor-level networks, investigating how signal coordination and varying traffic patterns influence the performance and stability of mixed traffic with AV-ACC vehicles.

Finally, this study is limited by the unavailability of detailed vehicle trajectory data in the current simulation framework, which prevented the analysis of stability indicators such as acceleration/deceleration distributions or vehicle stops. Future work will address these aspects using extended simulation capabilities.

Author Contributions: Conceptualization, M.A.R. and F.C.O.; methodology, M.A.R.; software, M.A.R. and F.C.O.; formal analysis, V.D.; resources, O.M.D. and S.B.; writing—original draft preparation, M.A.R., F.C.O. and V.D.; writing—review and editing, I.C.; project administration, M.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: The research work contained in this paper was supported within the framework of the grant no. 168/04.12.2023 from the National Program for Research of the National Association of Technical Universities—GNAC ARUT 2023.

Data Availability Statement: Additional data can be obtained from the authors.

Acknowledgments: This work was supported by a grant from the National Program for Research of the National Association of Technical Universities—GNAC ARUT 2023.

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

References

1. Porru, S.; Misso, 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]
2. Allam, Z.; Bibri, S.E.; Jones, D.S.; Chabaud, D.; Moreno, C. Unpacking the “15-Minute City” via 6G, IoT, and Digital Twins: Towards a New Narrative for Increasing Urban Efficiency, Resilience, and Sustainability. *Sensors* **2022**, *22*, 1369. [CrossRef]
3. Allam, Z.; Sharifi, A. Research Structure and Trends of Smart Urban Mobility. *Smart Cities* **2022**, *5*, 539–561. [CrossRef]
4. Christidis, P.; Ibáñez Rivas, J.N. *Measuring Road Congestion*; Practical Guidance Report EUR 25550 EN; Joint Research Centre, Institute for Prospective Technological Studies: Seville, Spain, 2012; ISBN 978-92-79-27015-4. ISSN 1831-9424. Available online: https://www.researchgate.net/publication/339712526_Measuring_road_congestion (accessed on 12 August 2025).

5. Counterpoint Technology Market Research. *Two-Thirds of Passenger Vehicles Sold in Europe in Q1 2025 Came with ADAS: Chinese OEMs Democratizing the Market*; Counterpoint Technology Market Research: London, UK, 2025; Available online: https://www.counterpointresearch.com/en/insights/two-thirds-of-pv-sold-in-europe-in-q1-2025-came-with-adas-chinese-oems-democratizing-market?utm_source=chatgpt.com (accessed on 12 August 2025).
6. Biyik, C.; Abareshi, A.; Paz, A.; Arce Ruiz, R.; Battarra, R.; Rogers, C.D.F.; Lizarraga, C. Smart Mobility Adoption: A Review of the Literature. *J. Open Innov. Technol. Mark. Complex.* **2021**, *7*, 146. [CrossRef]
7. Goumiri, S.; Yahiaoui, S.; Djahel, S. Smart Mobility in Smart Cities: Emerging challenges, recent advances and future directions. *J. Intell. Transp. Syst.* **2024**, *29*, 81–117. [CrossRef]
8. Medina-Tapia, M.; Robusté, F. Exploring paradigm shift impacts in urban mobility: Autonomous Vehicles and Smart Cities. *Transp. Res. Procedia* **2018**, *33*, 203–210. [CrossRef]
9. Golbabaee, F.; Yigitcanlar, T.; Bunker, J. The role of shared autonomous vehicle systems in delivering smart urban mobility: A systematic review of the literature. *Int. J. Sustain. Transp.* **2021**, *15*, 731–748. [CrossRef]
10. Manfreda, A.; Ljubi, K.; Groznik, A. Autonomous vehicles in the smart city era: An empirical study of adoption factors important for millennials. *Int. J. Inf. Manag.* **2021**, *58*, 102050. [CrossRef]
11. Richter, M.A.; Hagenmaier, M.; Bandte, O.; Parida, V.; Wincent, J. Smart cities, urban mobility and autonomous vehicles: How different cities needs different sustainable investment strategies. *Technol. Forecast. Soc. Change* **2022**, *184*, 121857. [CrossRef]
12. Pérez-Moure, H.; Lampon, J.; Velando-Rodriguez, M.E.; Rodriguez-Comesana, L. Revolutionizing the road: How sustainable, autonomous, and connected vehicles are changing digital mobility business models. *Eur. Res. Manag. Bus. Econ.* **2023**, *29*, 100230. [CrossRef]
13. Di, X.; Shi, R. A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning. *Transp. Res. C* **2021**, *125*, 103008. [CrossRef]
14. Campisi, T.; Severino, A.; Al-Rashid, M.A.; Pau, G. The Development of the Smart Cities in the Connected and Autonomous Vehicles (CAVs) Era: From Mobility Patterns to Scaling in Cities. *Infrastructures* **2021**, *6*, 100. [CrossRef]
15. Medina-Tapia, M.; Robusté, F. Implementation of Connected and Autonomous Vehicles in Cities Could Have Neutral Effects on the Total Travel Time Costs: Modeling and Analysis for a Circular City. *Sustainability* **2019**, *11*, 482. [CrossRef]
16. Cugurullo, F.; Acheampong, R.A.; Gueriau, M.; Dusparic, I. The Transition to Autonomous Cars, the Redesign of Cities and the Future of Urban Sustainability. *Urban Geogr.* **2021**, *42*, 833–859. [CrossRef]
17. Saravanos, A.; Pissadaki, E.K.; Singh, W.S.; Delfino, D. Gauging Public Acceptance of Conditionally Automated Vehicles in the United States. *Smart Cities* **2024**, *7*, 913–931. [CrossRef]
18. Tettamanti, T.; Varga, I.; Szalay, Z. Impacts of autonomous cars from a traffic engineering perspective. *Period. Polytech. Transp. Eng.* **2016**, *44*, 244–250. [CrossRef]
19. Seraj, M.; Li, J.; Qiu, Z. Modeling microscopic car-following strategy of mixed traffic to identify optimal platoon configurations for multiobjective decision-making. *J. Adv. Transp.* **2018**, *2018*, 7835010. [CrossRef]
20. SAE J 3016—SAE Standard J3016_202104; Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International Recommended Practice: Warrendale, PA, USA, 2021.
21. SAE Levels of Driving Automation™ Refined for Clarity and International Audience. Available online: <https://www.sae.org/site/blog/sae-j3016-update> (accessed on 12 August 2025).
22. Al-Turki, M.; Ratrou, N.T.; Rahman, S.M.; Reza, I. Impacts of Autonomous Vehicles on Traffic Flow Characteristics under Mixed Traffic Environment: Future Perspectives. *Sustainability* **2021**, *13*, 11052. [CrossRef]
23. Milanés, V.; Shladover, S.E.; Spring, J.; Nowakowski, C.; Kawazoe, H.; Nakamura, M. Cooperative Adaptive Cruise Control in Real Traffic Situations. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 296–305. [CrossRef]
24. Calvert, S.C.; Schakel, W.J.; van Lint, J.W.C. Will Automated Vehicles Negatively Impact Traffic Flow? *J. Adv. Transp.* **2017**, *2017*, 3082781. [CrossRef]
25. Wessling, B. Mercedes Rolls Out Level 3 Autonomous Driving Tech in Germany. Available online: <https://www.therobotreport.com/mercedes-rolls-out-level-3-autonomous-driving-tech-in-germany/> (accessed on 26 November 2023).
26. Dent, S. Mercedes Becomes the First Automaker to Sell Level 3 Self-Driving Vehicles in California. Available online: <https://www.engadget.com/mercedes-becomes-the-first-automaker-to-sell-level-3-self-driving-vehicles-in-california-103504319.html> (accessed on 1 December 2023).
27. Zhao, C.; Li, L.; Pei, X.; Li, Z.; Wang, F.-Y.; Wu, X. A comparative study of state-of-the-art driving strategies for autonomous vehicles. *Accid. Anal. Prev.* **2021**, *150*, 105937. [CrossRef] [PubMed]
28. Makahleh, H.Y.; Ferranti, E.J.S.; Dissanayake, D. Assessing the Role of Autonomous Vehicles in Urban Areas: A Systematic Review of Literature. *Future Transp.* **2024**, *4*, 321–348. [CrossRef]
29. D’Emidio, M.; Delfaraz, E.; Di Stefano, G.; Frittella, G.; Vittoria, E. Route Planning Algorithms for Fleets of Connected Vehicles: State of the Art, Implementation, and Deployment. *Appl. Sci.* **2024**, *14*, 2884. [CrossRef]

30. Kassens-Noor, E.; Dake, D.; Decaminada, T.; Kotval-K, Z.; Qu, T.; Wilson, M.; Pentland, B. Sociomobility of the 21st century: Autonomous vehicles, planning, and the future city. *Transp. Policy* **2020**, *99*, 329–335. [CrossRef]
31. Negash, N.M.; Yang, J. Driver Behavior Modeling toward Autonomous Vehicles: Comprehensive Review. *IEEE Access* **2023**, *11*, 22788–22821. [CrossRef]
32. Zheng, F.; Liu, C.; Liu, X.; Jabari, S.E.; Lu, L. Analyzing the Impact of Automated Vehicles on Uncertainty and Stability of the Mixed Traffic Flow. *Transp. Res. Part C Emerg. Technol.* **2020**, *112*, 203–219. [CrossRef]
33. Tian, Z.; Feng, T.; Timmermans, H.J.P.; Yao, B. Using autonomous vehicles or shared cars? Results of a stated choice experiment. *Transp. Res. C* **2021**, *128*, 103117. [CrossRef]
34. Rey, D.; Levin, M.W. Blue phase: Optimal network traffic control for legacy and autonomous vehicles. *Transp. Res. B* **2019**, *130*, 105–129. [CrossRef]
35. Lazar, D.A.; Biyik, E.; Sadigh, D.; Pedarsani, R. Learning how to dynamically route autonomous vehicles on shared roads. *Transp. Res. C* **2021**, *130*, 103258. [CrossRef]
36. Wang, J.; Wang, W.; Ren, G.; Yang, M. Worst-case traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles by factoring in the uncertain link capacity. *Transp. Res. C* **2022**, *140*, 103703. [CrossRef]
37. Chaudhry, A.; Sha, H.; Boghani, H.; Thomas, P.; Quddus, M.; Brackstone, M.; Tympakianaki, A.; Bin, H.; Glaser, S.; Papazikou, E.; et al. *Behavioural Parameters for Connected and Automated Vehicles Within the LEVITATE Project*; LEVITATE Consortium: Loughborough, UK, 2021.
38. Makridis, M.; Mattas, K.; Borio, D.; Giuliani, R.; Ciuffo, B. Estimating reaction time in Adaptive Cruise Control System. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; IEEE: New York, NY, USA, 2018; pp. 1312–1317. [CrossRef]
39. Laval, J.A. Traffic Flow as a Simple Fluid: Toward a Scaling Theory of Urban Congestion. *Transp. Res. Rec.* **2023**, *2678*, 376–386. [CrossRef]
40. Lee, E.H.; Lee, E. Congestion Boundary Approach for Phase Transitions in Traffic Flow. *Transp. B Transp. Dyn.* **2024**, *12*, 2379377. [CrossRef]
41. Liu, Z.; Ding, F.; Dai, Y.; Li, L.; Chen, T.; Tan, H. Spatial-Temporal Graph Convolution Network Model with Traffic Fundamental Diagram Information Informed for Network Traffic Flow Prediction. *Expert Syst. Appl.* **2024**, *249*, 123543. [CrossRef]
42. Oikonomou, M.G.; Ziakopoulos, A.; Chaudhry, A.; Thomas, P.; Yannis, G. From Conflicts to Crashes: Simulating Macroscopic Connected and Automated Driving Vehicle Safety. *Accid. Anal. Prev.* **2023**, *187*, 107087. [CrossRef]
43. Liu, W.; Hua, M.; Deng, Z.; Meng, Z.; Huang, Y.; Hu, C.; Song, S.; Gao, L.; Liu, C.; Shuai, B.; et al. Systematic Survey of Control Techniques and Applications in Connected and Automated Vehicles. *IEEE Internet Things J.* **2023**, *10*, 21892–21916. [CrossRef]
44. Ahmed, H.U.; Ahmad, S.; Yang, X.; Lu, P.; Huang, Y. Safety and Mobility Evaluation of Cumulative-Anticipative Car-Following Model for Connected Autonomous Vehicles. *Smart Cities* **2024**, *7*, 518–540. [CrossRef]
45. Lu, Q.; Tettamanti, T.; Hörcher, D.; Varga, I. The Impact of Autonomous Vehicles on Urban Traffic Network Capacity: An Experimental Analysis by Microscopic Traffic Simulation. *Transp. Lett.* **2020**, *12*, 540–549. [CrossRef]
46. Chen, S.; Hu, J.; Shi, Y.; Peng, Y.; Fang, J.; Zhao, R.; Zhao, L. Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G. *IEEE Commun. Stand. Mag.* **2017**, *1*, 70–76. [CrossRef]
47. Gupta, M.; Benson, J.; Patwa, F.; Sandhu, R. Secure V2V and V2I communication in intelligent transportation using cloudlets. *IEEE Transp. Serv. Comput.* **2020**, *15*, 1912–1925. [CrossRef]
48. Cascetta, E.; Gallo, M.; Montella, B. Models and algorithms for the optimization of signal settings on urban networks with stochastic assignment models. *Ann. Oper. Res.* **2006**, *144*, 301–328. [CrossRef]
49. Milakis, D.; Snelder, M.; van Arem, B.; van Wee, B.; Correia, G.H.A. Development and transport implications of automated vehicles in the Netherlands: Scenarios for 2030 and 2050. *Eur. J. Transp. Infrastruct. Res.* **2017**, *17*, 63–85. [CrossRef]
50. Gipps, P.G. A behavioural car-following model for computer simulation. *Transp. Res. Part B Methodol.* **1981**, *15*, 105–111. [CrossRef]
51. Gipps, P.G. A model for the structure of lane-changing decisions. *Transp. Res. Part B Methodol.* **1986**, *20*, 403–414. [CrossRef]
52. Federal Highway Administration (FHWA). *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software Publication No FHWA-HRT-04-040*; U.S. Department of Transportation: Washington, DC, USA, 2004.
53. Transport for London (TfL). *Traffic Modelling Guidelines: Version 4.0*; Boice, R., Radia, D., Biggs, J., Finnie, A., Eds.; TfL Traffic Manager and Network Performance Best Practice; Transport for London: London, UK, 2021.
54. Duraku, R.; Atanasova, V. Traffic Synthetic Model Development and Calibration in Anamorava Region. *Mach. Technol. Mater.* **2019**, *13*, 173–177.
55. Costescu, D.; Roşca, M.; Burciu, S.; Ruscă, F. On accident prediction functions for urban road intersections. *UPB Sci. Bull. Ser. D Mech. Eng.* **2016**, *78*, 55–64.
56. Körber, M.; Gold, C.; Lechner, D.; Bengler, K. The influence of age on the take-over of vehicle control in highly automated driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *50*, 55–68. [CrossRef]

57. Yang, X.; Rakha, H.A.; Ala, M. Modeling driver reaction time to the traffic signal change using queue discharge observations. *J. Adv. Transp.* **2015**, *49*, 795–814. [CrossRef]
58. Department for Transport. TAG Unit M3.1—Highway Assignment Modelling. In *Transport Analysis Guidance*; Department for Transport: London, UK, 2024. Available online: <https://assets.publishing.service.gov.uk/media/67ed0e54e9c76fa33048c634/tag-m3-1-highway-assignment-modelling.pdf> (accessed on 12 August 2025).
59. Varotto, S.F.; Mons, C.; Hogema, J.H.; Christoph, M.; van Nes, N.; Martens, M.H. Do adaptive cruise control and lane keeping systems make the longitudinal vehicle control safer? Insights into speeding and time gaps shorter than one second from a naturalistic driving study with SAE Level 2 automation. *Transp. Res. Part C Emerg. Technol.* **2022**, *141*, 103756. [CrossRef]
60. Lioris, J.; Pedarsani, R.; Tascikaraoglu, F.Y.; Varaiya, P. Platoons of connected vehicles can double throughput in urban roads. *Transp. Res. Part C Emerg. Technol.* **2015**, *77*, 292–305. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Tariff-Sensitive Global Supply Chains: Semi-Markov Decision Approach with Reinforcement Learning

Duygu Yilmaz Eroglu

Department of Industrial Engineering, Faculty of Engineering, Bursa Uludag University, Gorukle Campus, Bursa 16059, Turkey; duygueroглу@uludag.edu.tr

Abstract: Global supply chains often face uncertainties in production lead times, fluctuating exchange rates, and varying tariff regulations, all of which can significantly impact total profit. To address these challenges, this study formulates a multi-country supply chain problem as a Semi-Markov Decision Process (SMDP), integrating both currency variability and tariff levels. Using a Q-learning-based method (SMART), we explore three scenarios: (1) wide currency gaps under a uniform tariff, (2) narrowed currency gaps encouraging more local sourcing, and (3) distinct tariff structures that highlight how varying duties can reshape global fulfillment decisions. Beyond these baselines we analyze uncertainty-extended variants and targeted sensitivities (quantity discounts, tariff escalation, and the joint influence of inventory holding costs and tariff costs). Simulation results, accompanied by policy heatmaps and performance metrics, illustrate how small or large shifts in exchange rates and tariffs can alter sourcing strategies, transportation modes, and inventory management. A Deep Q-Network (DQN) is also applied to validate the Q-learning policy, demonstrating alignment with a more advanced neural model for moderate-scale problems. These findings underscore the adaptability of reinforcement learning in guiding practitioners and policymakers, especially under rapidly changing trade environments where exchange rate volatility and incremental tariff changes demand robust, data-driven decision-making.

Keywords: global supply chain; SMDP; reinforcement learning; tariff; exchange rate

1. Introduction

Managing multi-country production systems involves dealing with random production times, transportation times, uncertain exchange rates, and tariff costs. A typical approach is to formulate a Semi-Markov Decision Process (SMDP) [1,2] and solve it under the average reward criterion. Pontrandolfo et al. [3] demonstrated how a reinforcement learning (RL) technique can be applied to a GSCM context. Early studies further emphasized the strategic importance of integrating distributed supply chain activities under global conditions, with a particular focus on tariffs and currency gaps [3,4].

In this study, we build on the global supply chain management (GSCM) framework proposed by Pontrandolfo et al. [3], which employed an RL technique within an SMDP context to coordinate production, distribution, and sourcing across multiple countries. Our primary objective is to investigate how shifts in exchange rates and tariff policies influence both overall profitability and the allocation of orders among various production and distribution options. Leveraging an average-reward SMART learning framework that is cross-validated with a Deep Q-Network, we combine a baseline-versus-uncertainty scenario exploration with targeted sensitivity tests on quantity discounts, tariff escalation,

and the joint influence of inventory holding costs and duties; our core contribution is to quantify—under realistic parameters and in light of prior studies—how tariffs in the 25–30% range interact with exchange rate movements to establish a practical ceiling beyond which cross-border procurement rarely remains economically attractive.

To achieve this, we consider a three-country GSCM setting, where each country exhibits distinct exchange rate behavior modeled through random distributions, and shipping options include both within-country and cross-border modes (the latter featuring slow or fast transit choices). We assign typical cost parameters reflecting production expenses, pipeline and inventory costs, and a late delivery penalty if total lead times exceed a threshold. Additionally, when goods cross national borders, a tariff is imposed to capture international trade costs. Both production and transportation lead times follow probability distributions, reflecting real-world variability in manufacturing and logistics processes.

Within this broad setup, we define three main scenarios to systematically explore the impact of altering currency and tariff conditions. In the first scenario, we closely match the conditions and parameter settings originally proposed by Pontrandolfo et al. [3], effectively replicating their base case. Next, we narrow the currency gaps among countries to see how more closely aligned exchange rates affect sourcing strategies. Finally, we vary tariff structures so that each country has a different duty rate, revealing how policy interventions can drive substantial changes in global supply chain decisions. Through these scenarios, we highlight how small or large policy levers—such as incremental tariff adjustments or exchange rate shocks—lead to different strategies in production location, shipping mode selection, and inventory management. In doing so, we aim to offer insights for both practitioners (facing rapidly evolving trade conditions) and policymakers (examining broader effects of tariff changes).

Following the spirit of Pontrandolfo et al. [3], we model customer-demand arrivals in each country as an Erlang (Gamma) process with shape $k = 5$ and scale $\beta = 1$ (medium variability, $CV \approx 0.45$); the uncertainty-extended scenarios re-run the model under $k = 2$ (heavy-tailed, $CV \approx 0.71$) and $k = 7$ (light-tailed, $CV \approx 0.38$) to stress test policy robustness against demand variance. Unlike the original study—which compared local heuristics (LH, BH) and an RL-based SMART policy under multiple uncertainty levels—we focus exclusively on a Q-learning-based (SMART) SMDP in our core experiments. At the same time, to validate the model and confirm that the learned policy aligns with a more advanced, neural network approach, we also implement a Deep Q-Network (DQN). The DQN is thus deployed primarily as a verification step, showing that the tabular RL outcomes remain robust under a known problem size, one widely used in the literature to capture tariff effects. Consequently, while DQN results corroborate the viability of our Q-learning solution, we continue with Q-learning for the main analyses in order to more clearly illustrate how tariff variations impact the near-optimal sourcing and shipping decisions.

2. Literature Review

2.1. Reinforcement Learning for Logistics Optimization

Efforts to operationalize learning algorithms for supply chain optimization have been extensive. Researchers indicate that reinforcement learning can facilitate coordination in competitive supply chains, yet achieving convergence to truly optimal policies across distributed settings remains challenging [5]. A reinforcement learning-based model was applied to the beer game to improve coordination and reduce inventory costs across supply chain members, showing significant performance advantages over traditional ordering policies [6]. Multi-agent reinforcement learning has been applied to supply chain coordination problems. Zhao and Sun [7] developed a coordination framework for multi-agent systems, focusing on improving supply chain decision-making through real-time learning.

Puskás et al. (2020) applied reinforcement learning to optimize vehicle platoon coordination in a physical internet-based supply chain, demonstrating cost-efficient logistics through dynamic reconfiguration [8]. A comprehensive survey by Rolf et al. [9] reviewed RL in supply chain management (SCM) and underscored the gap between research prototypes and industrial-scale implementations. Zou et al. [10] further extended logistics-oriented RL by integrating a graph-attention network into a deep RL framework that jointly minimized carbon taxes, cooling costs, and travel distance for fresh food distribution under tight delivery windows, demonstrating the sustainability potential of DRL-based routing. Liu [11] developed a reinforcement learning-based model to optimize coordination and negotiation in cross-border e-commerce supply chains, improving efficiency and reducing decision time.

Recent work further highlights how pricing, financing, and inventory decisions interact under uncertainty. Li and Zheng [12] analyzed an online dynamic pricing problem with unknown, information-driven demand, proving tight revenue bounds relative to clairvoyant and deterministic benchmarks. Complementing that stream, Gurkan et al. [13] showed that a static (R, S, p) replenishment pricing policy can rival the profit-optimal dynamic (s, S, p) rule when demand variability is moderate. Capital constraints add another layer of complexity: Li and Li [14] designed a revenue-sharing and buy back (RSBB) contract that coordinates a cash-strapped new-energy-vehicle supply chain, with the revenue share decreasing in the retailer's acceptable bankruptcy risk, whereas the buy back price moves in the opposite direction. Beyond single-product settings, Bergemann et al. [15] characterized the welfare frontier of third-degree price discrimination, establishing that any surplus split between consumers and a monopolist is attainable within well-defined bounds. Finally, industry-specific studies such as that by Gomes et al. [16] demonstrate how tailored scheduling models (e.g., for fabric dyeing) can reduce inventories and set-up times, reinforcing the managerial value of jointly optimizing operational and pricing levers.

2.2. Reinforcement Learning for Inventory Management and Supply Chain Security

Beyond mainstream production and distribution concerns, specialized domains within supply chains have also benefited from RL. Lu and Wu [17] applied multi-agent reinforcement learning to optimize logistics information systems, focusing on improving stability, security, and decision efficiency in supply chain data processing. Kavididevi et al. (2024) [18] utilized IoT-integrated reinforcement learning to enhance the efficiency and reliability of refrigerated transport in cold chain logistics, enabling real-time environmental monitoring and dynamic decision-making to reduce spoilage and energy use. Zhou et al. [19] highlight RL's capacity to address volatile supply and demand, demonstrating how policy iterations outperform traditional methods in uncertain inventory settings. Liu et al. [20] applied a deep reinforcement learning approach to multi-echelon inventory management, showing that multi-agent learning architectures can effectively reduce total costs and respond dynamically to demand fluctuations. Meanwhile, Aboutorab et al. [21] integrated natural language processing with RL to proactively uncover and mitigate global supply chain disruptions, underscoring the importance of automated, data-driven resilience strategies. Focusing on a capital-intensive domain, Ma et al. [22] proposed a hierarchical RL framework for upstream crude oil scheduling, revealing superior performance relative to conventional operation research techniques in reducing penalties and transportation costs under dynamic supply–demand conditions. Complementing these efforts, Piao et al. [23] formulated a civil aircraft spares network as a partially observable MDP and achieved a 45% efficiency gain via dual-policy multi-agent RL, while Ma et al. [24] coupled Soft Actor–Critic with a blockchain-based zero-trust architecture to optimize inventory under demand uncertainty and ensure tamper-proof traceability. Collectively, these studies il-

illustrate how RL—especially when coupled with domain-specific heuristics, hierarchical structures, or real-time analytics—can flexibly adapt to operational complexities, yielding robust decision-making tools even under high degrees of uncertainty and scale.

2.3. Deep RL and Multi-Agent/Process Control Applications

Reinforcement learning has also seen significant advances through Deep Q-Networks (DQN), introduced initially by Mnih et al. [25], wherein a neural network function approximator replaced the traditional Q-table. Subsequent studies have extended DQN to handle more advanced RL challenges; for instance, Wang and Wang [26] applied it in international logistics for dynamic decision-making under uncertainty, while Double DQN [27], Dueling DQN [28], and Prioritized Experience Replay [29] have further improved stability and performance. Zhu et al. [30] provided a comprehensive survey on transfer learning within deep reinforcement learning, showing how knowledge from external tasks can accelerate learning when faced with new or complex scenarios. Padhye and Lakshmanan [31] focused on a ranking problem and proposed an actor–critic framework that utilizes deep networks to handle large action spaces (i.e., extensive item sets) and reduce variance via advanced policy-gradient techniques. Complementing these algorithmic advances is the safety-critical perspective highlighted by Norouzi et al. [32], who investigated “safe” deep reinforcement learning to manage diesel engine emissions. Their work incorporated a safety filter to enforce output constraints during training, thus reducing the risk of unacceptable solutions. Collectively, these studies underscore the growing sophistication of deep reinforcement learning approaches, whether through improved transfer-learning mechanisms, better handling of large-scale ranking tasks, or embedding explicit safety filters for real-world control systems. In this work, we primarily build our solution using a Q-learning-based SMDP framework; however, we also employ DQN as a secondary method to validate the outcomes of our Q-table approach. By comparing the two, we demonstrate that the Q-learning model’s decisions and results are consistent with and, in some scenarios, closely matched by a deep RL alternative, thus reinforcing confidence in the proposed tabular solution under the given problem constraints.

2.4. Broader Cyber-Physical and Data-Centric Applications of Deep Reinforcement Learning

Recent studies have illustrated the growing applicability of reinforcement learning (RL) in diverse, real-world domains. In mobile robotics, for example, DRL-based solutions have significantly enhanced autonomous navigation strategies and sensor integration, leading to more flexible, self-controlled robots under dynamic conditions [33]. Within sports analytics, offline RL approaches can now leverage large-scale event and tracking data to predict in-game outcomes or optimal decisions, outperforming traditional methods in both accuracy and efficiency [34]. Medical interventions, such as cardiovascular guidewire navigation, benefit from zero-shot RL techniques that generalize to unseen vascular anatomies without retraining [35]. Meanwhile, the financial sector has adopted RL frameworks to handle the challenges of dynamic datasets and market variability, enabling more robust and data-centric trading strategies [36]. Yu et al. [37] extended DQN and DDPG to dynamic edge-computing off-loading for fault diagnosis, showcasing DRL’s ability to orchestrate resource allocation across distributed cyber-physical nodes. These examples underscore RL’s versatility and its ongoing expansion into areas requiring adaptive, data-driven decision-making.

2.5. Research Gap

Taken together, this body of work highlights a rapidly enlarging frontier in which reinforcement learning—spanning both tabular and deep variants—consistently improves coordination, adaptability, and overall performance across an ever-broader array of

operational settings. The same versatility positions RL as a particularly promising tool for tariff-sensitive global supply chains that must cope with compounded uncertainties such as currency shocks, shifting trade policies, and environmental pressures; by learning adaptively and scaling to higher complexities, RL frameworks can deliver robust, data-driven policies in precisely these volatile contexts. To the best of our knowledge, despite the breadth of applications reviewed, no prior study simultaneously integrated exchange rate volatility, heterogeneous tariff regimes, and variable decision intervals within a semi-Markov RL framework validated by both tabular and deep methods. The present study fills this gap by applying average-reward SMART learning and validating the resulting policy against a Deep Q-Network in a tariff-sensitive global sourcing setting. Furthermore, we embed the proposed SMART-DQN framework in a two-tier experimental program that extends well beyond static parameter settings. First, a baseline vs. uncertainty-extended scenario exploration introduces demand surges and random transport disruptions to test policy robustness under operational noise. Second, a suite of sensitivity analyses—spanning quantity discounts, systematic tariff escalation, and the joint inventory holding–tariff cost trade-off—demonstrate that the 25–30% tariff band consistently emerges as the tipping point at which cross-border sourcing loses its economic appeal. These layers of validation strengthen the methodological novelty of our study and confirm the managerial relevance of the identified policy threshold.

3. Methodology

In this section, we first formulate the Semi-Markov Decision Process (SMDP) model tailored to global supply chain management (GSCM) problems. We then describe the reinforcement learning (RL) approach, focusing on the SMART algorithm (a Q-learning-based method) to obtain near-optimal solutions under variable transition times. Finally, we introduce Deep Q-Networks (DQNs) as a powerful extension for dealing with larger or more complex state spaces. Although our core experiments relied on a tabular Q-learning approach, we employed DQN primarily as a validation tool, confirming that the policy learned by the SMART algorithm aligns well with results from a deep neural network function approximator in moderate-scale scenarios.

All experiments were executed on the same personal computer (Intel® Core™ i5-6200U @ 2.30 GHz, 8 GB RAM) running Microsoft Windows 10 Pro (22H2). The computational environment comprised Python 3.10.18, JupyterLab 4.3.4, NumPy 1.26.4, SimPy 4.1.1, and Matplotlib 3.10.3.

3.1. Semi-Markov Decision Processes (SMDP)

A Semi-Markov Decision Process (SMDP) is defined by a state space S , an action set A , transition probabilities, rewards, and random transition times. Unlike a standard Markov Decision Process (MDP) that operates in fixed time steps, an SMDP allows variable intervals between decisions. Let $\{(X_m, T_m)\}$ be a Markov renewal process, where $X_m \in S$ denotes the state of the system at the m -th decision epoch and T_m is the time of the m -th epoch. The defining property of the SMDP is expressed as

$$\begin{aligned} \Pr(X_{m+1} = j, t \leq T_{m+1} - T_m \mid X_0, \dots, X_m, T_0, \dots, T_m) \\ = \Pr(X_{m+1} = j, t \leq T_{m+1} - T_m \mid X_m), \end{aligned} \quad (1)$$

where t represents the random duration until the next decision epoch. In this framework, a policy π is a rule prescribing which action $a \in A$ to choose when the system is in a particular state $s \in S$. The goal is typically to optimize a performance criterion—commonly the long-

run average reward in semi-Markov settings—by selecting the actions that maximize the expected return over time, given the variable transition intervals.

In other words, while (1) shows how the evolution of states and times satisfies the semi-Markov (memoryless) property, the policy π indicates the action to be taken once the system enters state s . Although the variable time intervals do not explicitly appear in the notation for π , they are inherently accounted for in the process dynamics via the random variable t , as well as in the resulting long-run objective function (e.g., average reward) under time-varying transitions.

3.2. Reinforcement Learning (RL)

Reinforcement Learning (RL) offers a model-free framework for solving SMDPs by forgoing explicit transition probabilities and instead learning policies through trial-and-error interaction with the environment [2]. Concretely, an RL agent observes states s , selects actions a , and receives rewards r over many episodes. As it gathers experience, the agent refines its action-value function $Q(s, a)$ —the estimated long-term value of choosing action a in state s —toward the optimal $Q^*(s, a)$.

During learning, the agent balances exploration (trying actions that are uncertain) with exploitation (selecting high-value actions found so far). As more transitions are sampled, $Q(s, a)$ converges to accurate value estimates, guiding the agent to near-optimal decisions under varying state transitions and rewards. This model-free property is especially advantageous in SMDPs with uncertain or complex dynamics since it dispenses with having to pre-specify detailed transition probabilities or reward structures.

3.3. Smart Algorithm (Average Reward Q-Learning)

In this study, we employed a value iteration-based RL variant for average reward SMDPs, often referred to as the SMART (Semi-Markov Average Reward Technique) algorithm. Unlike discount-based Q-learning, which optimizes the discounted sum of rewards, average reward Q-learning focuses on maximizing ρ , the long-run average reward. The general update rule is

$$Q(s, a) \leftarrow (1 - \alpha) Q(s, a) + \alpha \left[r - \rho t + \max_{b \in A} Q(s', b) \right], \quad (2)$$

where

- α is the learning rate,
- r is the immediate reward received after taking action a in state s ,
- ρ is the current estimate of the average reward,
- t is the elapsed time for the transition,
- s' is the next state, and
- b iterates over all possible actions from s' .

This update is repeated until convergence, at which point $Q(s, a)$ approximates the maximum achievable average reward and the policy $\pi(s) = \arg \max_a Q(s, a)$ is nearly optimal.

3.4. Deep Q-Networks (DQNs)

Deep Q-Networks (DQNs) were first introduced by Mnih et al. [25] as a powerful extension of classical Q-learning, enabling reinforcement learning (RL) to scale to larger or more complex state spaces via deep neural network function approximation. In their pioneering work, the authors demonstrated that a convolutional neural network (CNN) trained end-to-end could learn effective policies from raw images in Atari 2600 games, surpassing many previous hand-engineered approaches. This breakthrough showed that combin-

ing standard Q-learning updates [38] with deep function approximators could achieve human-level (and sometimes super-human) performance in high-dimensional tasks.

DQN builds on tabular Q-learning by introducing three main concepts:

- **Neural Network Function Approximation:** A deep neural network $Q_\theta(s, a)$, parameterized by weights θ , approximates the action-value function $Q(s, a)$. This approach avoids the exponential growth of a tabular Q-learning when the state space becomes large or continuous.
- **Experience Replay:** Transition tuples (s, a, r, s') collected from the environment are stored in a replay buffer. During training, mini-batches are sampled randomly from this buffer, breaking the strong correlations often present in sequential data and improving the stability of training.
- **Target Networks:** DQN maintains a separate set of network parameters θ^- for the target Q function, which is updated more slowly (e.g., every fixed number of steps) to reduce oscillations and divergence in learning.

In small to medium-sized problems, DQN often achieves good results quickly without requiring exhaustive enumeration of all states, making it especially suitable for preliminary or proof-of-concept experiments in RL-based supply chain models. As problem dimensions increase (e.g., due to larger stock level ranges, additional countries, or more complex dynamics), DQN's capacity for function approximation becomes even more critical, allowing it to generalize where a tabular Q table would become intractably large.

3.5. Implementation Outline

A high-level summary of the SMART-based procedure applied to this GSCM model is as follows:

1. **Initialize:** Set $Q(s, a)$ to some initial value for all states s and actions a . Initialize ρ , the average reward estimate, to 0 or a small number.
2. **Observe State:** Determine the current system state (i, s) .
3. **Action Selection:** Choose an action a (e.g., shipping mode) using an ϵ -greedy or softmax policy based on the current Q -values.
4. **Simulate:** Generate a random production time and shipping time, apply the chosen action, and calculate the realized reward r .
5. **Update:** Compute the transition time t . Update $Q((i, s), a)$ using (2). Periodically adjust ρ via $\rho \leftarrow \rho + \beta[r - \rho t]$, where β is a small step size.
6. **Transition:** Observe the next state (i', s') . Repeat until convergence or until a predefined number of iterations.

Once the algorithm converges, the resulting Q-table (or function) captures the near-optimal policy dictating how orders should be allocated among shipping modes and production locations. In the following sections, we apply this methodology to three main scenarios differing in exchange rate distributions and tariff structures.

3.6. Modeling Assumptions

The SMART-based simulation rests on the following assumptions, which match or extend the settings used by Pontrandolfo et al. [3] and enable tractable average-reward learning:

- **Exchange rate processes.** Currency conversion factors e_i for the three countries are independently and stationarily drawn from the scenario-specific uniform intervals. Cross-country correlations are neglected to isolate tariff effects.
- **Demand arrivals.** Customer orders follow a stationary Erlang (Gamma) arrival process with parameters (k, β) ; demand is therefore independent of exchange rate fluctuations and inventory levels.

- Lead times. Production and transportation times are mutually independent and identically distributed (i.i.d.) across episodes, each following the uniform distributions.
- Inventory costs. Holding costs are linear in time and quantity, with no economies of scale; stockouts are not back-ordered but incur lost sales penalties implicit in the reward function. A late arrival penalty of $L = 3$ monetary units per unit shipped late.
- Capacity. Warehouse capacities are finite and fixed (five equivalent stock-keeping units in all runs); production capacity is assumed to be sufficient to meet demand on average, consistent with [3].
- Tariffs and prices. Tariff rates and selling prices are deterministic within each scenario.

Robustness checks. To test the impact of relaxing the above assumptions, Section 6 introduces (i) uncertainty-extended scenarios with demand surges and random transport disruptions and (ii) three sensitivity suites: quantity discounts, systematic tariff escalation, and the joint tariff–inventory cost grid. All robustness experiments confirm that the main finding remains stable under these perturbations.

4. Experimental Results

In this section, we illustrate our proposed GSCM framework in the context of a three-country network adapted from Pontrandolfo et al. [3]. We first describe the network structure and then detail the key cost components, parameter settings, and action choices in our SMDP model.

4.1. Network Structure

As depicted in Figure 1, each country i ($i = 1, 2, 3$) includes

- A production plant (PP_i),
- An export warehouse (EW_i),
- An import warehouse (IW_i),
- A final market (M_i).

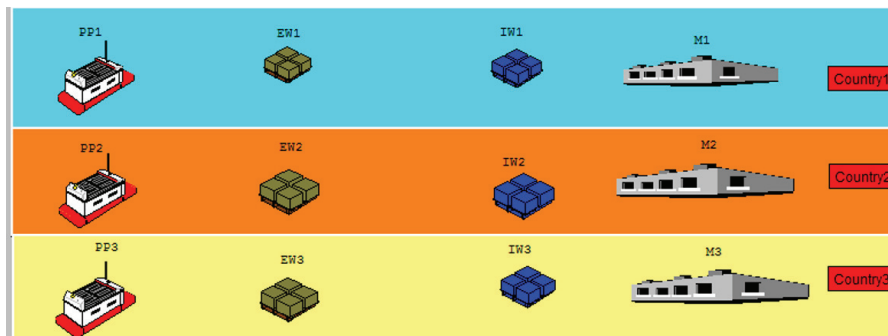


Figure 1. Illustration of the three-country supply chain (adapted from [3]).

Customer demands originate at each market M_i . When a demand arrives, the decision-maker must choose between two broad fulfillment routes:

1. Domestic route—ship directly from the same country's export warehouse EW_i ; this option avoids tariffs and involves only short, low-cost transport, but is limited by the available stock in EW_i .
2. Cross-border route—import the item from another country's export warehouse EW_j ($j \neq i$). This choice triggers additional transportation time, tariff charges at the destination country, and extra pipeline inventory costs.

Both routes are further constrained by finite warehouse capacities; once a warehouse is full, incoming production must wait, and once it is empty, certain routing options become infeasible. These capacity limits are explicitly captured in the stock component of the

SMDP state, making inventory management an integral part of the reinforcement learning decision process.

4.2. Cost Formulation and Parameter Settings

To capture the principal trade-offs in this global network, we modeled production costs, transportation costs, inventory holding costs, and late delivery penalties, alongside tariff expenses for cross-border movements. The total time to fulfill an order encompassed both production lead time and transportation time, each of which followed a specified probability distribution. Our baseline exchange rate intervals and the uniform 15% tariff were taken directly from the benchmark study of Pontrandolfo et al. [3], so that subsequent scenario variations remained comparable to an established data set.

Below, we summarize the primary parameters and their role in our SMDP model:

- Exchange Rates (e_i). Each country's currency converts into the reference monetary unit (MU) via

$$e_1 \sim U(0.4, 0.6), \quad e_2 \sim U(0.9, 1.1), \quad e_3 \sim U(1.9, 2.1).$$

- Sale Price (P). The list price is fixed at $P = 500$ MU (monetary unit); selling in country j therefore yields $e_j P$.
- Production Cost (M). A unit produced in country i costs $e_i M$ in MU with $M = 300$ LU (local currency units).
- Inventory Holding Cost (I_0). Baseline carrying charge $I_0 = 10$ MU per unit-time, rescaled by the country-specific storage factors $IT_i \in \{0.1715, 0.3430, 0.6860\}$ and the sensitivity multiplier κ introduced in Section 6.
- Transportation Cost and Time. Let S_{ij} be the shipping charge for an order routed from origin i to destination j and s_t be its transit time:
 1. Domestic ($i = j$): $S_{ij} = 30$, $s_t \sim U(1, 3)$.
 2. Cross-border, slow ($i \neq j$): $S_{ij} = 45$, $s_t \sim U(15, 17)$.
 3. Cross-border, fast ($i \neq j$): $S_{ij} = 60$, $s_t \sim U(3, 5)$.
- Pipeline Cost Rate (C). A running charge of $C = 6$ MU per time unit is applied while the shipment is in transit.
- Tariff Rate (τ). A uniform ad-valorem duty $\tau e_j P$ applies to every cross-border movement ($i \neq j$). Baseline $\tau = 0.15$ for Scenarios 1 and 2; Scenario 3 adopts country-specific rates $[\tau_1, \tau_2, \tau_3] = [0.10, 0.15, 0.30]$.
- Late-Delivery Penalty (L). If total lead time $p_t + s_t > 10$ tu, a penalty of $e_j L$ with $L = 3$ MU is incurred.
- Production Lead Time (p_t). Drawn i.i.d. from $U(1, 3)$ time units.

Revenue/Reward Calculation. Suppose an item is produced in country i and shipped to country j . The net gain in the implementation is derived by the following equation. In (3) the large bracket groups together (i) the production cost, (ii) the pre- and post-production inventory charges $\kappa I_0 e_i IT_i$ and $\kappa I_0 e_j IT_j$, (iii) the fixed freight fee S_{ij} , and (iv) the time-proportional pipeline cost $C(p_t + s_t)$ incurred while the item is in transit. The two indicator functions $1_{\{i \neq j\}}$ and $1_{\{p_t + s_t > 10\}}$ evaluate to 1 when the shipment is, respectively, cross-border and late; they are 0 otherwise. All symbols appearing in the reward expression are summarized in Table 1.

$$\begin{aligned}
 r = & \underbrace{e_j P}_{\text{revenue}} - \left(\underbrace{e_i M}_{\text{production}} + \underbrace{\kappa I_0 e_i IT_i}_{\text{pre-prod. holding}} + \underbrace{S_{ij}}_{\text{shipping}} + \underbrace{C(p_t + s_t)}_{\text{pipeline time}} \right. \\
 & \left. + \underbrace{\kappa I_0 e_j IT_j}_{\text{post-ship. holding}} \right) - \underbrace{\tau e_j P 1_{\{i \neq j\}}}_{\text{tariff}} - \underbrace{e_j L 1_{\{p_t + s_t > 10\}}}_{\text{late penalty}}.
 \end{aligned} \tag{3}$$

Table 1. Notation used in the landed-cost and reward expressions.

Symbol	Meaning
e_i, e_j	Origin (i) and destination (j) exchange rates, quoted as local-currency (LU) units per monetary unit (MU).
P	List selling price (MU per unit).
M	Unit production cost in origin currency (LU).
I_0	Baseline inventory holding charge (MU · tu ^{−1}).
κ	Multiplier applied to I_0 for sensitivity runs ($\kappa \in \{0.5, 1.0, 2.0\}$).
IT_i, IT_j	Country-specific storage factors $\{0.1715, 0.3430, 0.6860\}$.
C	In-transit cost rate (MU · tu ^{−1}).
p_t, s_t	Realized production and shipping lead times (tu).
τ	Ad valorem customs duty rate applied when $i \neq j$.
L	Late delivery penalty (MU) if $p_t + s_t > 10$ tu.
$1_{\{\cdot\}}$	Indicator function (equals 1 when the stated condition holds; otherwise, 0).

Illustrative Calculation

Using the baseline values $M = 300$, $I_0 = 10$, $IT = (0.1715, 0.3430, 0.6860)$, $C = 6$, $P = 500$, $\tau = 0.15$, and $\kappa = 1.0$ and the instance $i = 0$ with $e_0 = 0.50$, $j = 2$ with $e_2 = 2.00$, $p_t = 2$, and $s_t = 4$ tu, while assuming that the shipment arrives on time ($p_t + s_t \leq 10$), (3) becomes

$$\begin{aligned}
 r &= e_2 P - \left(e_0 M + \kappa I_0 e_0 IT_0 + C(p_t + s_t) + \kappa I_0 e_2 IT_2 \right) - \tau e_2 P \\
 &= 2.00 \times 500 - \left(0.50 \times 300 + 1.0 \times 10 \times 0.50 \times 0.1715 + 6 \times (2 + 4) + 1.0 \times 10 \times 2.00 \times 0.6860 \right) \\
 &\quad - 0.15 \times 2.00 \times 500 \\
 &= 1000 - (150 + 0.86 + 36 + 13.72) - 150 \\
 &= 649.42.
 \end{aligned}$$

Even after accounting for production, two holding phases, pipeline cost, and a 15% tariff, the high destination exchange rate ($e_2 = 2.00$) still yields a healthy profit of ≈ 649 MU for this single order.

4.3. SMDP States and Actions

Destination Country Logic:

Let i be the originating country and a the chosen action. We define the destination country j as follows:

1. Action 0: Local Shipping.

$$j = i.$$

2. If $i = 0$ (Country 1):

- Action 1 or 2 $\rightarrow$ shipping from Country 2 ($j = 1$), where Action 1 indicates slow shipping and Action 2 indicates fast shipping.

- Action 3 or 4 $\rightarrow$ shipping from Country 3 ($j = 2$), where Action 3 is slow and Action 4 is fast.
3. If $i = 1$ (Country 2):
 - Action 1 or 2 $\rightarrow$ shipping from Country 1 ($j = 0$), with Action 1 for slow and Action 2 for fast.
 - Action 3 or 4 $\rightarrow$ shipping from Country 3 ($j = 2$), with Action 3 for slow and Action 4 for fast.
 4. If $i = 2$ (Country 3)::
 - Action 1 or 2 $\rightarrow$ shipping from Country 1 ($j = 0$), with Action 1 for slow and Action 2 for fast.
 - Action 3 or 4 $\rightarrow$ shipping from Country 2 ($j = 1$), with Action 3 for slow and Action 4 for fast.

4.4. Scenario Setup

Using these parameters and definitions, we investigated how an RL-based SMDP model adapted under different global conditions. Specifically, we examined three key scenarios:

1. Replication of Pontrandolfo et al. [3]—adopting their original exchange rate distributions and a uniform tariff of 0.15.
2. Narrowed Currency Gap—modifying the exchange rate ranges to be more closely aligned among the three countries.
3. Differentiated Tariffs—imposing distinct duty rates per country to assess how tariff disparities shape sourcing decisions.

By contrasting these scenarios, we demonstrate how even moderate shifts in exchange rates or import duties can significantly alter global supply chain strategies. In the next section, we detail our RL-based methodology (SMART) for deriving near-optimal policies within this SMDP framework.

4.5. Setup and Implementation

We began our analysis with a three-country baseline in which each country's demand process followed the same Erlang distribution with moderate variability. Specifically, we set the shape parameter $k = 5$ and scale $\beta = 1$, yielding an average inter-arrival time of $k\beta = 5$ and a coefficient of variation $CV \approx 0.45$. For the robustness experiments introduced later, we additionally explored “heavy-tailed” demand with $k = 2$ ($CV \approx 0.71$) and “light” demand with $k = 7$ ($CV \approx 0.38$); the implementation merely replaced the shape parameter k when launching those runs. The code leveraged the `simpy` simulation library to generate demand arrivals, applied a Q-learning (RL) update (SMART-style) for each incoming demand, and ran the simulation for `SIM_TIME = 100,000` time units.

Each demand arrival triggers a state lookup: we identified the $(country, stock)$ state; selected an action via ϵ -greedy; computed the resulting reward based on production and shipping costs, tariffs, and inventory holding; and then updated the Q-table.

5. Comparisons and Overall Discussion

We investigated three scenarios, each reflecting a different combination of exchange rate distributions and tariff structures. Below, we recap the findings from each scenario and then discuss the overarching insights, including a validation experiment using Deep Q-Networks (DQNs).

All simulation notebooks, trained models, and auxiliary artifacts underlying the present work are archived on Zenodo (Yilmaz Eroglu (2025)) [39].

5.1. Validation

Although our main solution utilized a Q-learning method, we also employed a Deep Q-Network (DQN) to validate that the learned policy was robust, aligning with more advanced function-approximation approaches [25]. DQN replaced the Q-table with a neural network and used experience replay and target networks to stabilize training [27–29]. We tested the same three scenarios using both Q-learning and DQN under identical cost structures and demand processes, comparing their average rewards (ρ) over simulation time.

The tabular-versus-DQN run reported below was executed under Scenario 1. We fixed the demand process at $\Gamma(k = 5, \beta = 1)$ and ran each learner for a 500 tu horizon.

Quantitative comparison (Table 2).

Table 2 brings together six performance indicators for the stock-level grid $|S| \in \{5, 10, 20, 30\}$. The SMART lookup table (ρ_{Tab}) out-earns its neural counterpart (ρ_{DQN}) at every size, with absolute gaps of 4–7 MU. Overlap% is the share of the 450 state–action pairs whose greedy action under DQN matched the SMART optimum; even when the reward gap was large, DQN still agreed with SMART on at least one state in eight and on almost 60 % of the states at $|S| = 30$. MSE quantifies the mean-squared deviation between the two Q-tables (reported in 10^3 MU^2); its sharp fall from 21.6 to 1.2 confirmed that the network approximation improved as the state grid became denser. Both learners converged within a single 10 k epoch episode (Conv. ep.) across the full range, indicating efficient learning for problems up to $|S| = 30$.

Table 2. Quantitative comparison of tabular Q-learning vs. DQN (500 tu horizon, ≈ 300 arrivals).

$ S $	ρ_{Tab}	ρ_{DQN}	Overlap (%)	MSE ($\times 10^3$)	Conv. Ep. (T/D)
5	20.82	16.10	13.3	21.6	1 / 1
10	25.25	21.05	40.0	5.1	1 / 1
20	20.63	13.93	11.7	2.2	1 / 1
30	21.09	15.60	57.8	1.2	1 / 1

Dynamic behavior (Figure 2).

Figure 2 tracks the on-line estimate of the long-run average reward $\rho(t)$ over the first 300 demand arrivals (≈ 500 simulation time units). Solid lines show SMART; dashed lines show the corresponding DQN runs. All trajectories exhibit an initial exploration spike followed by damped oscillations and stabilization. For every $|S|$, the SMART curve settled earlier and higher, illustrating its sample-efficiency advantage: each interaction updated the exact state–action cell, whereas the DQN needed to rely on mini-batch gradients to shape its value function.

Scalability snapshot (Figure 3).

Figure 3 distills the end-point data of Figure 2 into a single scalability plot: final ρ versus $|S|$. SMART dominated across the entire range, peaking at $|S| = 10$ and losing only four MU when the state space tripled to $|S| = 20$. DQN peaked at the same point but its profit dropped sharply (by more than 7 MU) at $|S| = 20$ and recovered only marginally at $|S| = 30$. Hence, within realistic warehouse capacities ($|S| \leq 30$) a tabular policy was simultaneously more interpretable and more profitable than a deeper, data-hungry alternative.

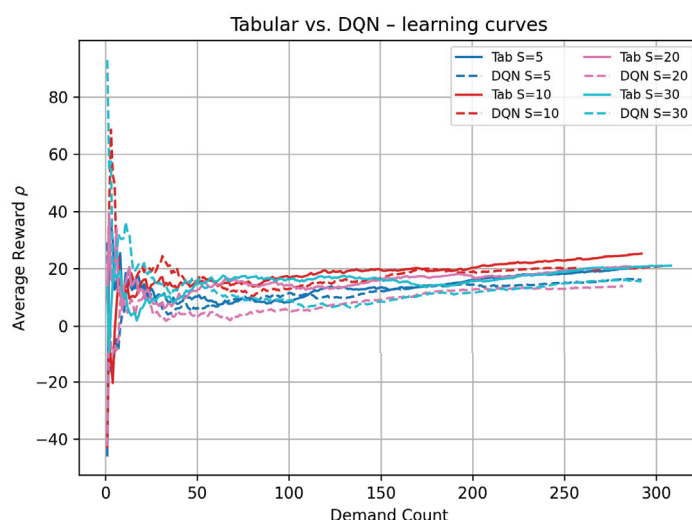


Figure 2. Learning curves of tabular Q-learning (solid) versus DQN (dashed) for stock level sizes $|S| = 5, 10, 20, 30$. Horizontal axis: cumulative Demand Count; vertical axis: on-line estimate of $\rho(t)$.

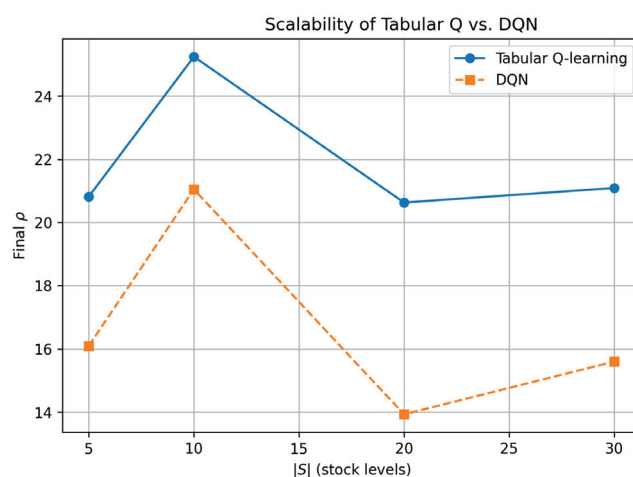


Figure 3. Scalability of tabular Q vs. DQN: final average reward ρ after 500 tu as a function of $|S| \in \{5, 10, 20, 30\}$.

The long-run average rewards in Table 2 and Figures 2 and 3 are purposely higher (about $20\text{--}26 \text{ MU} \cdot \text{tu}^{-1}$ for tabular) than those reported in the subsequent scenario analyses (about $8\text{--}10 \text{ MU} \cdot \text{tu}^{-1}$). This is because the DQN-vs.-tabular benchmark was run for only 500 simulated time units (≈ 300 demand arrivals) under Scenario 1's wide exchange rate gap and a modest 15% tariff, with no demand surges or transport disruptions. The short horizon limited inventory holding and late delivery penalties, while the large FX arbitrage inflated gross margins, together yielding a “best-case, warm-up snapshot” that was well suited for algorithmic comparison but not indicative of the steadier, lower-margin values observed over the 60,000–100,000 tu horizons analyzed later.

Given that warehouse policy limited stock discretization to $S \leq 30$ and since tabular Q-learning outperformed DQN throughout this range, we retained the tabular SMART agent as the primary optimizer in all subsequent tariff and scenario analyses. DQN remained a secondary verification tool and a scalability reserve for potential future extensions where the state space would grow beyond the current operational bounds.

We analyzed three progressively restrictive policy experiments that varied exchange rate dispersion and tariff intensity in order to isolate their individual and joint effects on routing decisions. Scenario 1 adopted Pontrandolfo et al. [3]'s original settings—wide exchange rate bands (0.4–0.6, 0.9–1.1, 1.9–2.1) and a uniform 15% tariff—so that subsequent

results were directly comparable to the seminal benchmark. Scenario 2 retained the 15% tariff but “compressed” the currency bands to (0.8–1.0, 0.9–1.1, 0.7–0.9), mirroring the $\pm 10\%$ purchasing power parity spreads typical of OECD and upper-middle-income currencies in the 2024 Big-Mac Index [40]. Frequent tariff interventions during the 2018–2019 U.S.–China dispute clustered around 10% and 15%, with 25% emerging as the upper end of actual policy moves [41]. Scenario 3 therefore retained 10% and 15% as “realistic” tiers and introduced a 30% ceiling—just above the observed 25%—to stress-test the model under a plausibly extreme, yet policy-relevant, shock.

Figure 4 shows that the tabular SMART learner reached a stable, long-run average reward under each exchange rate/tariff scenario before any policy comparisons were made, thus complementing the earlier DQN cross-validation and ensuring that subsequent results were based on steady-state, not transient, behavior. Figure 4 tracks the long-run average reward ρ per episode, where one episode was defined as 10,000 consecutive decision epochs (production–shipment cycles). Bundling raw epochs into fixed-length episodes served two purposes: (i) it produced a smooth, interpretable curve that was independent of the highly variable physical time a single epoch may cover in a semi-Markov setting; (ii) it allowed us to apply a pragmatic stopping rule—training was halted once the absolute change in ρ between two successive episodes fell below 1 % of the previous value.

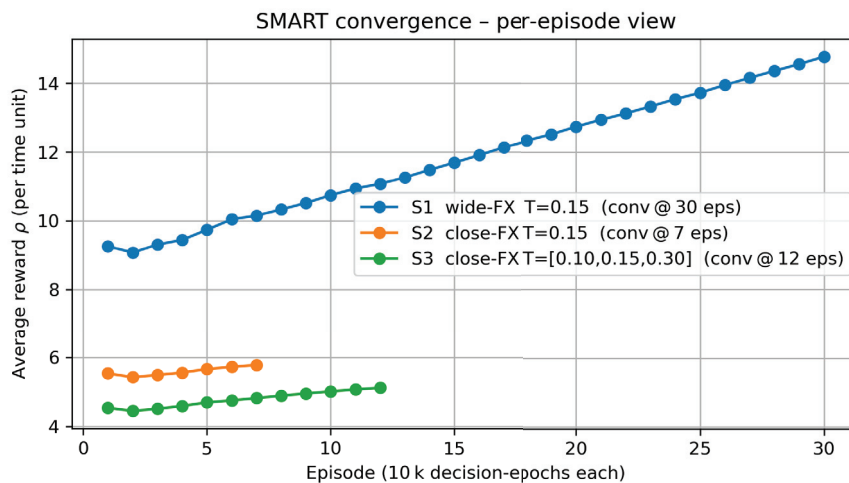


Figure 4. Convergence behavior of the SMART learner. Each point represents the long-run average reward ρ after one episode of 10,000 decision epochs (arrivals $\rightarrow$ actions $\rightarrow$ stochastic lead times). Convergence was declared once the drift between two consecutive episodes fell below 1 %.

To put the 100,000 simulated time-unit (tu) horizon used in the main experiments into perspective, recall that customer demand in each country followed an Erlang process with shape–scale parameters $(k, \beta) = (5, 1)$; hence, the mean inter-arrival time was $k\beta = 5$ tu. Over 100,000 tu, each country therefore generated on average $100,000/5 = 20,000$ arrivals. Because the three national markets operated independently, the system as a whole experienced approximately $3 \times 20,000 \approx 60,000$ decision epochs, where one decision epoch was defined as “arrival $\rightarrow$ action choice $\rightarrow$ stochastic lead-time realization” and triggered a single Q -value update. For monitoring purposes, we bundled these raw epochs into fixed episodes of 10,000 decision epochs each; consequently, 100,000 tu corresponded to about $60,000/10,000 \simeq 6$ episodes. This back-of-the-envelope calculation justified the six-episode convergence windows shown in Figure 4.

Under these conditions the SMART learner converged in 7, 12, and 30 episodes for Scenarios 2, 3, and 1, respectively. Scenario 1, which featured large exchange rate differentials, exhibited a prolonged upward drift because the agent continued to exploit profitable

cross-border arbitrage as it received more experience. By contrast, Scenario 2 and Scenario 3 stabilized much earlier, reflecting a flatter cost landscape and the dampening effect of tariffs on exploratory re-routing. The steady-state reward hierarchy $\rho_{S1} > \rho_{S2} > \rho_{S3}$ visible on the last points of each curve corroborated the policy performance comparisons. Note that the largest run (Scenario 1, 30 episodes) still fit well within the global simulation horizon of 100,000 time units adopted throughout this study, confirming that all performance metrics were obtained from a fully converged Q-table.

5.2. Scenario 1: Erlang Demand with Uniform Tariffs

In this baseline scenario, countries faced wide exchange rate gaps (e.g., 0.4–0.6, 0.9–1.1, 1.9–2.1) but a single uniform tariff ($T = 0.15$). The resulting Q-table is presented in Table 3 and a heatmap appears in Figure 5.

Table 3. Scenario 1 Q-table results (uniform tariff, wide exchange gaps).

State	Country	Stock	Action 0	Action 1	Action 2	Action 3	Action 4	BestAct
0	1	0	55,797.67	55,735.26	56,134.91	56,585.01	56,802.81	4
1	1	1	56,053.55	56,018.52	56,154.06	56,641.59	56,521.91	3
2	1	2	55,626.69	55,563.01	56,178.62	56,126.73	56,432.10	4
3	1	3	55,458.11	55,748.21	55,629.78	56,249.75	56,312.47	4
4	1	4	55,678.57	55,817.28	56,178.32	56,066.94	56,461.56	4
5	2	0	37,992.47	37,960.22	37,831.13	38,307.13	38,576.20	4
6	2	1	38,293.11	37,701.19	37,679.09	38,309.32	38,805.14	4
7	2	2	38,255.27	37,643.85	38,124.26	38,588.06	38,751.09	4
8	2	3	38,375.37	37,801.73	38,149.97	38,532.25	38,754.07	4
9	2	4	38,426.73	37,919.21	38,195.35	38,182.92	38,945.80	4
10	3	0	27,146.37	26,108.66	26,296.62	26,358.95	26,486.41	0
11	3	1	27,184.81	26,079.84	26,235.35	26,250.30	26,524.80	0
12	3	2	27,225.82	26,082.50	26,412.65	26,339.11	26,418.35	0
13	3	3	27,220.88	25,991.03	26,483.73	26,419.68	26,695.00	0
14	3	4	27,238.16	26,195.35	26,409.91	26,465.04	26,579.72	0

In Table 3, the final column labeled `bestAction` gives $\arg \max_{a \in \{0, \dots, 4\}} Q(s, a)$, that is, the index of the action that yielded the highest long-run average reward for the corresponding *(country, stock)* state. Action 0 denoted local shipping, whereas Actions 1–4 denoted cross-border shipping with slow (1, 3) or fast (2, 4) transit, as defined in the previous subsection. Hence, `bestAction` 3 or 4 indicated that the agent preferred to import from the high-exchange-rate country via the specified mode. Figure 5 visualizes the same Q-learning table as a heatmap. A prominent block-wise gradient appears from rows 0–4 to rows 10–14: bright yellow bands correspond to Country 1 (the most favorable exchange rate range), magenta tones to Country 2, and deep indigo to Country 3. Within each five-row block (fixed country), color shades are almost uniform, confirming that, in this scenario, stock level had only a second-order effect on the learned Q-values; the dominant gradient was the block-wise shift caused by the large exchange rate differences across countries.

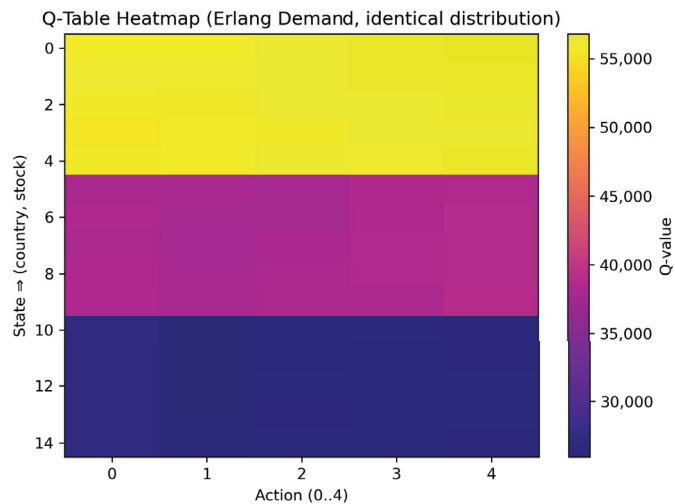


Figure 5. Scenario 1 heatmap. Each row corresponds to one (country, stock) state; columns are actions {0, 1, 2, 3, 4}. High values (yellow) favor shipping from or to high-exchange-rate regions.

Our major observations were

- Country 1 and Country 2 frequently selected actions associated with cross-border shipping from Country 3 (e.g., bestAction = 3 or 4) to exploit Country 3’s high exchange rate, offsetting the uniform tariff cost.
- Country 3 predominantly used action 0, indicating that local fulfillment was most profitable when exchange rate advantages were already in its favor.
- The agent achieved an average reward consistent with the previous literature’s wide-gap scenarios, validating the RL approach in this baseline setting.

5.3. Scenario 2: Closer Exchange Rates

Here, we narrowed the currency ranges to (0.8–1.0, 0.9–1.1, 0.7–0.9) while keeping the tariff rate at $T = 0.15$. As displayed in Table 4 and Figure 6, the net-revenue advantage of cross-border shipments was markedly smaller than in Scenario 1.

Table 4. Scenario 2 Q-table results (closer exchange rates, uniform tariff).

State	Country	Stock	Action 0	Action 1	Action 2	Action 3	Action 4	BestAction
0	1	0	11,669.72	11,427.81	11,593.90	11,417.57	11,548.73	0
1	1	1	11,712.63	11,481.70	11,607.85	11,407.87	11,452.70	0
2	1	2	11,648.60	11,452.73	11,683.33	11,392.84	11,502.95	2
3	1	3	11,685.69	11,497.28	11,599.06	11,442.73	11,493.52	0
4	1	4	11,675.88	11,521.64	11,718.05	11,409.78	11,584.26	2
5	2	0	12,138.90	11,938.90	12,008.87	11,825.58	11,948.29	0
6	2	1	12,200.02	11,853.55	11,954.85	11,817.45	12,040.80	0
7	2	2	12,239.07	11,878.84	12,105.46	11,924.77	12,014.54	0
8	2	3	12,311.24	11,967.88	12,151.85	11,952.27	12,022.63	0
9	2	4	12,387.64	12,088.60	12,240.48	11,918.42	12,221.33	0
10	3	0	10,296.33	10,125.91	10,257.62	10,175.23	10,325.42	4
11	3	1	10,299.94	10,109.84	10,235.71	10,142.84	10,347.86	4
12	3	2	10,359.88	10,121.03	10,316.26	10,186.36	10,279.27	0
13	3	3	10,332.04	10,081.90	10,337.05	10,204.45	10,390.45	4
14	3	4	10,330.06	10,164.82	10,313.79	10,212.58	10,368.00	4

Table 4: The Q-values here lay in a much tighter band—roughly 10,100 to 12,400 monetary units—showing that exchange rate differences contributed less to overall profitability. The bestAction column revealed a strong shift toward local fulfillment: Country 1 selected

Action 0 in three of five stock states; Country 2 chose local shipping in all states; Country 3 overwhelmingly chose the fast cross-border alternative (Action 4) in four out of five stock states; the sole exception was the mid-stock level ($s = 2$), where local fulfillment (action 0) became marginally more rewarding.

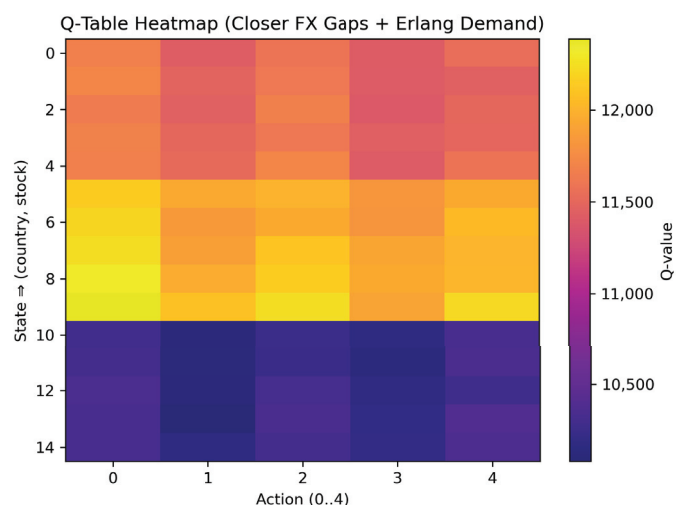


Figure 6. Scenario 2 heatmap (closer exchange rates).

In Figure 6, the color scale ranged from around 10,100 (deep indigo) to 12,400 (bright yellow). Compared with Figure 5, the horizontal block-wise contrast across countries was less pronounced; instead, subtle vertical striations indicate that the choice between slow (actions 1, 3) and fast (actions 2, 4) cross-border modes was here cost-neutral in many states. Overall, the flattened Q-value landscape confirms that narrowing the currency gap reduced the economic incentive for cross-border sourcing, thereby encouraging more local or balanced strategies across the three-country network.

5.4. Scenario 3: Distinct Tariffs with Closer Exchange Rates

Finally, we introduced distinct tariffs for each origin country, e.g., $d_1 = 0.10$, $d_2 = 0.15$, $d_3 = 0.30$, while retaining the narrower currency intervals.

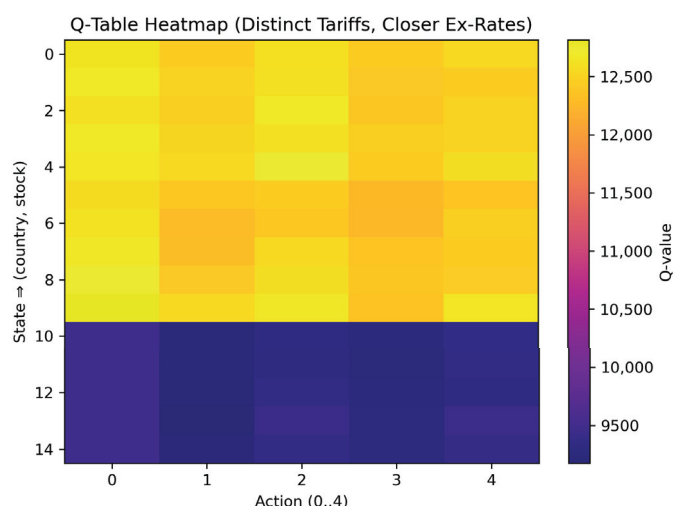
As summarized in Table 5 and visualized in Figure 7, the color scale now ran from roughly 9100 (deep indigo) to 12,750 (bright yellow) monetary units, representing a wider band than in Scenario 2. This broader spread confirmed that the introduction of heterogeneous tariffs—rather than the already-narrow exchange rate differences—became the dominant source of Q-value variation. The bestAction column shows a strong preference for local fulfillment (Action 0) in Countries 1 and 2, with only two fast cross-border selections (Action 2) for Country 1 at stock levels $s = 2$ and $s = 4$. Country 3 never chose a cross-border option; its 30% export tariff fully canceled any residual currency edge. Correspondingly, rows 0–9 in Figure 7—representing Countries 1 and 2—display warm yellow or orange cells under Action 0, whereas rows 10–14 (Country 3) remain uniformly violet, highlighting the profitability gap induced by the tariff differential. Vertical shading differences between slow (Actions 1 and 3) and fast (Actions 2 and 4) cross-border modes are faint, indicating that these intra-origin transit-time effects were secondary to the tariff impact.

Table 5. Scenario 3 Q-table results (closer exchange rates, distinct tariffs).

State	Country	Stock	Action 0	Action 1	Action 2	Action 3	Action 4	BestAction
0	1	0	12,635.99	12,435.77	12,588.43	12,422.53	12,538.60	0
1	1	1	12,680.25	12,491.59	12,602.34	12,413.40	12,439.54	0
2	1	2	12,615.54	12,461.90	12,680.22	12,398.07	12,491.98	2
3	1	3	12,673.93	12,526.96	12,608.40	12,472.45	12,500.39	0
4	1	4	12,650.28	12,543.50	12,730.19	12,420.97	12,584.04	2
5	2	0	12,559.73	12,380.48	12,431.44	12,264.04	12,370.90	0
6	2	1	12,622.40	12,292.89	12,376.09	12,255.44	12,467.23	0
7	2	2	12,663.00	12,317.49	12,531.50	12,365.93	12,439.72	0
8	2	3	12,737.03	12,409.21	12,579.32	12,394.63	12,447.82	0
9	2	4	12,815.02	12,533.38	12,671.67	12,359.32	12,652.52	0
10	3	0	9464.18	9217.06	9341.78	9262.61	9371.38	0
11	3	1	9479.41	9208.13	9323.17	9230.65	9386.80	0
12	3	2	9482.04	9199.65	9376.28	9251.11	9338.48	0
13	3	3	9477.38	9176.32	9406.39	9287.30	9448.74	0
14	3	4	9493.40	9243.05	9375.91	9285.56	9398.39	0

Table 5 and Figure 7 demonstrate how imposing heterogeneous origin tariffs reshaped the learned policy: the Q-value landscape was here segmented primarily by tariff level rather than by the modest exchange rate differences, leading to the following country-specific patterns.

- Country 1 showed significant variation between local (Action 0) and cross-border (particularly Action 2 or 4) shipping. The interplay of moderate exchange gain versus partial tariff led to nuanced decisions.
- Country 2 mostly remained local as the distinct tariff structure seldom warranted cross-border shipping unless a modest inventory shortage arose.
- Country 3, facing the highest tariff outflow (0.30), rarely found cross-border shipping profitable; nearly all states yielded bestAct = 0, preserving local supply.

**Figure 7.** Scenario 3 heatmap (distinct tariffs, closer exchange rates).

6. Discussion and Conclusions

6.1. Baseline Design: Three Currency–Tariff Scenarios

Through the three scenarios presented, we demonstrate how a reinforcement learning (RL) approach (SMART) adapts its fulfillment strategy in response to varying exchange rate gaps and tariff rules.

- Wide exchange gaps + uniform tariffs (Scenario 1): Encourage cross-border shipping from the highest-exchange-rate country (Country 3), provided inventory or lead-time constraints do not override the advantage.
- Closer exchange rates (Scenario 2): Reduce the incentive for cross-border shipments, promoting more local or balanced sourcing while showing how a smaller currency gap can shift strategies away from a single dominant supplier.
- Distinct tariffs (Scenario 3): Further complicate the cost landscape by discouraging exports from countries with steep tariffs (e.g., 0.30 for Country 3), although local vs. global choices may still hinge on inventory levels and lead-time risks.

Although the baseline three-scenario matrix already captured the most salient cost drivers, exchange rate gaps, and tariff heterogeneity, real-world supply chains face additional stochastic shocks such as demand surges and in-transit disruptions. To verify that our RL controller remains effective under such noise, we augmented each baseline configuration with a carefully chosen set of uncertainty-extended variants. The resulting six-scenario suite, documented in Table 6 and visualized in Figure 8, allowed us to confirm that all qualitative insights obtained from the lean three-scenario design continued to hold once these extra uncertainties were switched on.

6.2. Scenario Exploration: Baseline vs. Uncertainty-Extended Variants

To probe the robustness of the SMART reinforcement learning controller, we expanded the original three scenarios (wide-gap exchange rates, closer exchange rates, and distinct tariffs) with three uncertainty-extended counterparts that injected (i) stochastic supply disruptions and (ii) demand-arrival variance via alternative $\Gamma(k, \beta = 1)$ parameters. The complete six-scenario design is summarized in Table 6.

All six scenarios shared the same cost structure, yet three extended variants injected additional randomness along two orthogonal axes:

1. Demand $\Gamma(k, \beta = 1)$. We modeled inter-arrival times with a Gamma distribution whose shape k directly controlled variance while leaving the mean unchanged.
 - $k = 5$ (baseline): moderate variability.
 - $k = 7$ ("light"): thinner tail $\rightarrow$ smoother workload.
 - $k = 2$ ("heavy"): thicker tail $\rightarrow$ bursty arrivals, mimicking demand shocks.
2. Supply disruption (10 tu delay, $p = 0.10$). With a probability of 0.10, every shipment received an additional 10 time-unit transit penalty, representing port congestion, customs holds, or transport breakdowns. The penalty was large enough to trigger the model's late delivery cost in many episodes, thereby stressing the learning agent.

Why these extensions matter. They broadened the uncertainty set from purely price-driven shocks (exchange rates and tariffs) to operational and market shocks: variable customer demand and stochastic pipeline delays. The red, violet, and brown trajectories in Figure 8 demonstrate that despite noticeably higher volatility in ρ , the ordering of profit performance ($\rho_{S1} > \rho_{S2} > \rho_{S3}$) remained intact. Hence, the core managerial insights reported for the lean three-scenario design were robust to these extra real-world complications, justifying our decision to keep the subsequent policy experiments focused on Scenarios 1–3 only.

Table 6. Definition of the six simulation scenarios.

Scenario Label	Exchange Rate Range	Tariff Rule	Demand $\Gamma(k, \beta=1)$	Supply Disruption
Scenario 1	(0.4–0.6, 0.9–1.1, 1.9–2.1)	uniform 0.15	$k=5$ (baseline)	–
Scenario 2	(0.8–1.0, 0.9–1.1, 0.7–0.9)	uniform 0.15	$k=5$	–
Scenario 3	same as Sc. 2	[0.10, 0.15, 0.30]	$k=5$	–
Scenario 1 (+Ext)	same as Sc. 1	uniform 0.15	$k=2$ (heavy)	10 tu delay, $p=0.10$
Scenario 2 (+Ext)	same as Sc. 2	uniform 0.15	$k=7$ (light)	10 tu delay, $p=0.10$
Scenario 3 (+Ext)	same as Sc. 2	[0.10, 0.15, 0.30]	$k=2$	10 tu delay, $p=0.10$

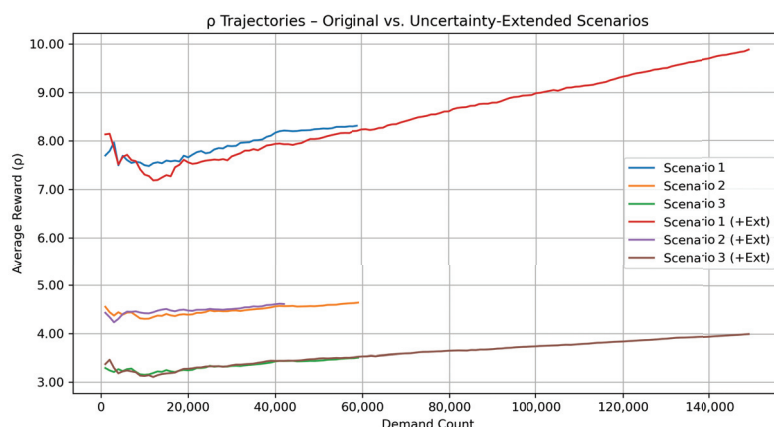


Figure 8. Long-run average reward (ρ) under six scenarios. Each curve tracks ρ versus cumulative demand arrivals. Scenarios 1–3 (solid blue, orange, green) reproduced the baseline exchange rate/tariff settings; their uncertainty-extended counterparts (red, violet, brown) added heavy-tailed or light-tailed demand and a 10 time-unit (tu) disruption that occurred with a probability of 0.10.

What the trajectories reveal.

Figure 8 confirms three consistent patterns.

- Currency advantage dominated. After injecting heavy demand variance and 10 tu supply delay shocks, Scenario 1 (+Ext) still produced the highest long-run reward. Country 3 retained the strongest currency (exchange rate interval 1.9–2.1), so shipping from Country 3 remained profitable enough to outweigh both the uniform 0.15 tariff and the extra disruptions. The extended demand process ($k=2$) simply lengthened the horizontal axis—more demands were realized—thereby stretching the trajectory far to the right and letting ρ climb above 9.7.
- Narrower gaps compressed profit. Both baseline and extended versions of Scenario 2 plateaued at around $\rho \approx 4.5$. Here, the three exchange rate intervals overlapped (0.8–1.1), so cross-border arbitrage was weak; the lighter arrival variance case ($k=7$) only added marginal noise without changing the level.
- Steep tariffs remained binding. Scenario 3 trajectories settled roughly one unit below Scenario 2. The country-specific tariffs [0.10, 0.15, 0.30] penalized exports from the high-currency regions, and the added disruption risk in the extended variant reduced ρ a little further.

Because the extended variants preserved the qualitative ordering $\rho_{S1} > \rho_{S2} > \rho_{S3}$, we retained the lean three-scenario set to streamline the policy experiments, while Figure 8 shows that the conclusions were robust to additional real-world uncertainties.

6.3. Sensitivity Analysis

6.3.1. Sensitivity to Quantity Discounts

Real-world procurement contracts often feature volume-based rebates; to examine whether such price incentives altered our policy insights, we re-ran the baseline exchange

rate/tariff design under a 10% quantity discount that was triggered once on-hand inventory exceeded two days of expected demand. The threshold $= 2 \times \mathbb{E}[\text{daily demand}] \simeq 2$ units, matching the Erlang- $(k = 5, \beta = 1)$ arrival rate. The six resulting trajectories—three original scenarios with and without a discount—are plotted in Figure 9. Across all cases, the discount depressed the long-run average reward ρ by roughly 15–20% yet preserved the ordering $\rho_{S1} > \rho_{S2} > \rho_{S3}$. Hence, the managerial conclusions derived from the lean three-scenario framework remained robust; the subsequent sections therefore continued to focus on Scenarios 1–3 without loss of generality.

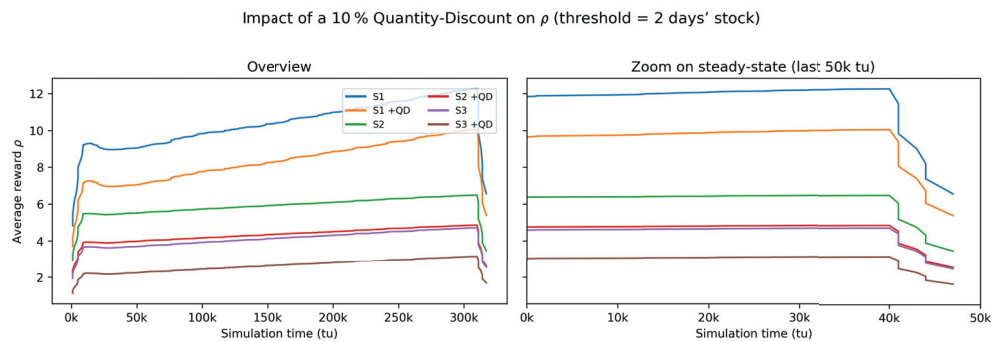


Figure 9. Impact of a 10% quantity discount on the long-run average reward ρ . Each colored pair represents one of the three exchange rate/tariff scenarios: the first entry (S1–S3) is the baseline, whereas the matching label with “+QD” shows the outcome when the 10% discount was triggered once on-hand stock exceeded two days of expected demand. The **(left panel)** depicts the entire 300 k tu horizon; the **(right panel)** zooms into the final 50 k tu to highlight steady-state differences between the baseline and discount runs. The symbol “k tu” abbreviates thousand time units (e.g., 50 k tu = 50,000 simulation time units).

6.3.2. Sensitivity to Tariff

To quantify how strongly the SMART policy reacted to tariff shocks, we re-ran the baseline exchange rate design while systematically varying a uniform ad valorem tariff τ applied to every cross-border shipment. The sweep covered tariff grid $\tau \in \{0, 0.05, 0.10, \dots, 0.40\}$; each run spanned 250 k tu (steady state reached $\gtrsim 80\%$ of the horizon). $\tau = 0$ to 0.40 in steps of 0.05 for all three scenarios:

- Scenario S1—wide: large exchange rate gaps, uniform baseline tariff $\tau = 0.15$.
- Scenario S2—close: compressed exchange rates, uniform tariff.
- Scenario S3—tariff-vec: same exchange rates as S2 but country-specific baseline tariffs; when swept, the same τ was imposed on every lane for comparison.

Figure 10 tracks the share of cross-border shipments ($|\text{international orders}|/\text{total}$), whereas Figure 11 displays the resulting long-run average reward ρ under the same tariff grid.

Figure 10, cross-border share under rising tariffs: The plot tracks the share of orders that the SMART policy fulfilled via cross-border shipments as the ad valorem tariff increased from 0 to 40%. For the wide-gap exchange rate setting (Scenario 1, solid blue) the currency advantage was so strong that the policy kept more than 80% of flows international until tariffs exceeded about 25%; only beyond that point did the curve slope down more steeply. By contrast, the closer-gap baseline (Scenario 2, orange) and the distinct tariff case (Scenario 3, green) were far more sensitive: their curves crossed the “half-and-half” line (50% cross-border share) around the $\tau = 0.30$ mark, confirming the reviewer’s intuition that a 30% duty was a decisive threshold for reshaping fulfillment toward local sources.

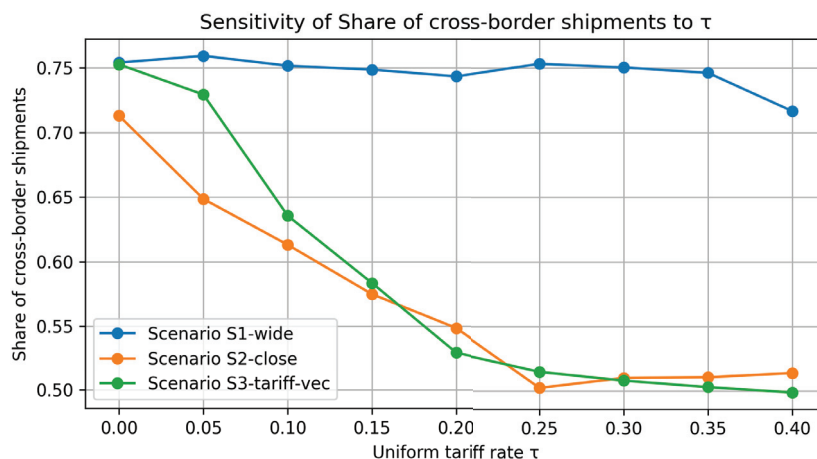


Figure 10. Tariff sensitivity of the cross-border shipment share. All curves slope downward: higher uniform tariffs discouraged global fulfillment. S1 (large currency gaps) remained the most internationalized ($\approx 75\%$ at $\tau=0$; still $> 70\%$ at $\tau=0.40$), while S2 and especially S3 pivoted sharply toward domestic sourcing once $\tau \geq 0.20$.

Figure 11—profit impact of tariff escalation: The panel converts the uniform tariff sweep into long-run average reward ρ (MU tu^{-1}). All three scenarios followed an almost linear downward trend, yet with clearly distinct slopes. Scenario 1, protected by the wide exchange rate gap, still posted the highest margins but surrendered about 6 MU of ρ for every ten-percentage-point (10 pp) increase in the ad valorem duty. By contrast, Scenarios 2 and 3—whose currency bands were compressed—lost roughly 2.5–3 MU per 10 pp; the slightly steeper gradient of Scenario 3 reflects the extra asymmetry introduced by its country-specific baseline tariffs. Taken together, the sourcing-mix plot (Figure 10) and the ρ curve shown here demonstrate that both network routing and overall profitability pivoted decisively once tariffs crossed the 25–30% band, thereby pinpointing the policy threshold at which cross-border procurement ceased to be economically attractive.

Key take-aways. (i) Tariffs reshape fulfilment choices primarily by throttling the volume of cross-border moves (Figure 10); profit effects (Figure 11) are a second-order consequence. (ii) Scenario ordering $\rho_{S1} > \rho_{S2} \approx \rho_{S3}$ persists for $\tau \leq 0.25$, but narrows considerably beyond that point. (iii) A uniform tariff of $\tau \gtrsim 0.30$ is sufficient to push the system toward near-autarky in otherwise balanced currency settings, underscoring how trade policy can overwrite exchange rate incentives.

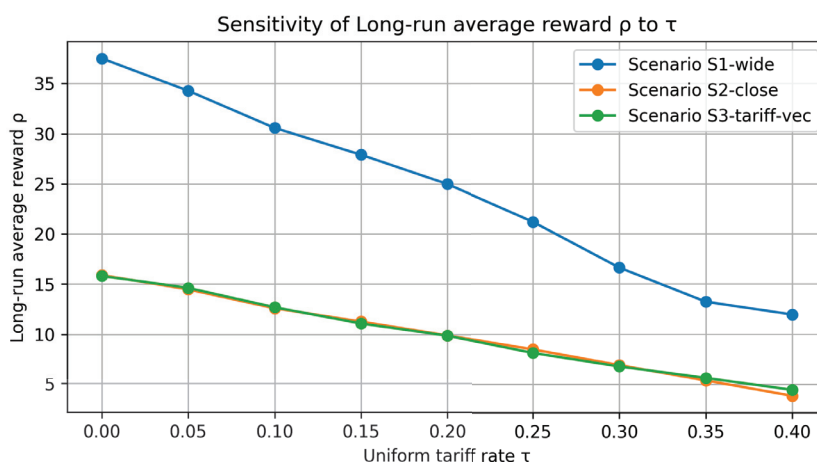


Figure 11. Tariff sensitivity of the long-run average reward ρ . Profit declined roughly linearly with τ . Wide currency advantages (S1) buffered the loss, yet ρ dropped by $\sim 65\%$ over the sweep. S2 and S3 converged, showing that once currency gaps were narrow, any tariff materially eroded margins.

6.3.3. Joint Impact of Inventory Holding Cost and Tariff Level

Inventory carrying charges and customs duties are two of the most salient cost drivers in global fulfillment. To quantify their joint influence, we multiplied the baseline per-unit holding rate $I_0 = 10 \text{ MU/tu}$ by $\kappa \in \{0.5, 1.0, 2.0\}$ and swept the uniform ad valorem tariff across $\tau \in \{0, 0.10, 0.20, 0.30\}$, keeping all other parameters identical to Table 6. These two scalars entered the landed-cost expression already introduced in Equation (3): e_i, e_j : origin/destination exchange rates; $M = 300 \text{ MU}$: unit production cost; $IT_i \in \{0.1715, 0.3430, 0.6860\}$: country-specific storage factors; $C = 6 \text{ MU per tu}$: in-transit cost rate; $P = 500 \text{ MU}$: list price; $L = 3 \text{ MU}$: late delivery penalty. The indicator $1_{\{i \neq j\}}$ activated the customs duty only for cross-border routes; $1_{\{p_t + s_t > 10\}}$ triggered the late arrival charge whenever the realized pipeline time exceeded ten time units.

- Inventory multiplier κ rescaled both holding terms in Equation (3), i.e., the pre-production component $\kappa I_0 e_i IT_i$ and the post-shipment component $\kappa I_0 e_j IT_j$.
- Tariff rate τ linearly scaled the customs-duty term $\tau e_j P 1_{\{i \neq j\}}$.

Each (κ, τ) pair was simulated for 3×10^4 time units with average-reward Q-learning; the long-run reward ρ was averaged over three macro-replications to attenuate noise. The results are summarized in Figure 12.

Key observations. (i) Doubling or halving κ shifted the entire ρ -curve almost in parallel, leaving its slope with respect to τ virtually unchanged, indicating that customs duties dominated stockholding charges in shaping landed cost. (ii) Increasing τ steeply reduced profit irrespective of κ , underscoring the primary role of duties in cross-border sourcing decisions. (iii) For every (κ, τ) tested, the ordering $\rho_{S1\text{-wide}} > \rho_{S2\text{-close}} > \rho_{S3\text{-tarvec}}$ remained intact, confirming that currency gaps and tariff structures were the principal levers, whereas inventory costs acted as a measurable—but second-order—modifier.

Managerial Insights: From a practical standpoint, global supply chain managers should continuously monitor tariff regulations and exchange rate fluctuations across potential sourcing countries. When tariffs spike or currency gaps narrow, previously optimal cross-border options might become suboptimal, emphasizing the need for agile decision-making and frequent policy updates. Having robust estimates for pipeline costs, lead-time risks, and inventory holding expenses further helps identify whether the gain from a cheaper supplier in another country still outweighs potential delays or tariff penalties.

To strengthen confidence in the learned policy, we also conducted a verification experiment using a Deep Q-Network (DQN). Although the tabular Q-learning (SMART) approach was simpler and highly interpretable for moderate-scale problems, the DQN results closely matched those of the tabular method, indicating that both methods respond similarly to changes in exchange rates and tariffs.

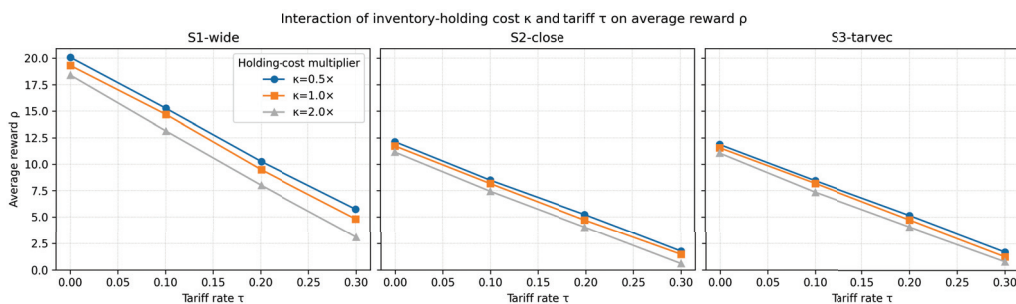


Figure 12. Interaction of inventory holding cost κ and tariff τ on average reward ρ . Columns correspond to the three exchange rate scenarios, while the three colored curves in each panel represent $\kappa = 0.5\times, 1.0\times$, and $2.0\times$ the baseline holding rate. Markers denote the simulated τ levels; lines are visual guides.

From Prototype to Shop-Floor Decision Support: While the foregoing results confirm the analytical value of the SMART learner, a key managerial question is whether such a policy can be embedded in day-to-day planning systems.

- **Data pipeline.** Most ERP suites already expose inventory levels and open purchase orders and inbound transit times. These feeds can populate the learner’s state vector in real time or on an hourly batch schedule.
- **Training and inference layer.** The SMART algorithm is lightweight enough to train overnight on a single CPU node. We therefore recommend a two-tier micro-service: (i) a nightly “trainer” container that ingests the previous day’s transactions and updates the Q-table; (ii) a stateless “inference” endpoint that is called by the ERP’s MRP run or by a /suggestRoute button in the buyer’s user interface.
- **Power BI dashboard.** Writing the live Q-table and KPIs (average reward, cross-border share, lead-time breach rate) to a cloud data-lake allows instant visualization in Power BI, Tableau, or Looker. Planners can “what-if” tariffs or exchange rates via slicers, triggering on-the-fly policy simulations without touching the production ERP.
- **Governance and override.** A simple traffic-light override is advisable: if the recommended action deviates from the incumbent rule or if the implied delivery date violates a service level agreement, the suggestion is pushed to an approval queue rather than executed automatically.

In short, the SMART learner can be surfaced as a thin decision-support layer—trained off-line, served on-line—that complements, rather than replaces, existing ERP logic. This architecture minimizes integration risk while giving managers a transparent, continuously improving view of how tariffs, exchange rate shifts, and pipeline costs jointly shape optimal fulfillment choices.

6.4. Limitations and Future Work

- Modeling Scope.** The prototype optimizes a single-product, single-echelon network without explicit capacity limits. Extending SMART to multiple SKUs (Stock-Keeping Units), regional DC (Distribution Center) layers, and finite production or transport capacities would require a constrained MDP formulation or a hierarchical multi-agent RL design. Future extensions could also account for sustainability metrics (e.g., CO₂ cost per lane) and dynamic forecasting signals that update the MDP in real time.
- Stochastic Environment.** Customer demand arrivals follow a stationary Erlang $\Gamma(k, \beta)$ process, while all unit-cost coefficients (production, holding, pipeline, tariff, late penalty rates) are deterministic; uncertainty in the reward therefore arises only from demand timing, the stochastic production and transit times, and exchange rate fluctuations. Future research could relax these assumptions by introducing non-stationary demand patterns, volatile freight-rate schedules, and time-dependent quantity discount schemes.
- Computational Scalability.** Tabular Q-learning suffices for the current 450 state–action pairs but does not scale to large state spaces. Preliminary Deep-Q results match the tabular policy; moving to $|S| > 10^4$ will necessitate deeper networks, experience replay prioritization, and distributed training to maintain convergence speed.

Funding: This research received no external funding.

Data Availability Statement: All data generated and analyzed in this study are openly available on Zenodo (Yilmaz Eroglu (2025)) [39].

Conflicts of Interest: The author declares no conflicts of interest.

References

- Howard, R. *Dynamic Probabilistic Systems, Volume II: Semi-Markov and Decision Processes*; John Wiley & Sons, Inc.: New York, NY, USA, 1971.
- Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; John Wiley & Sons: New York, NY, USA, 2014.
- Pontrandolfo, P.; Gosavi, A.; Okogbaa, O.G.; Das, T.K. Global supply chain management: A reinforcement learning approach. *Int. J. Prod. Res.* **2002**, *40*, 1299–1317. [CrossRef]
- Günay, E.E.; Park, K.; Okudan Kremer, G.E. Integration of product architecture and supply chain under currency exchange rate fluctuation. *Res. Eng. Des.* **2021**, *32*, 331–348. [CrossRef]
- van Tongeren, T.; Kaymak, U.; Naso, D.; van Asperen, E. Q-learning in a competitive supply chain. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montréal, QC, Canada, 7–10 October 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1211–1216.
- Chaharsooghi, S.K.; Heydari, J.; Zegordi, S.H. A reinforcement learning model for supply chain ordering management: An application to the beer game. *Decis. Support Syst.* **2008**, *45*, 949–959. [CrossRef]
- Zhao, X.; Sun, X. A multi-agent reinforcement learning approach for supply chain coordination. In Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Qingdao, China, 15–17 July 2010; pp. 341–346.
- Puskás, E.; Budai, Á.; Bohács, G. Optimization of a physical internet based supply chain using reinforcement learning. *Eur. Transp. Res. Rev.* **2020**, *12*, 47. [CrossRef]
- Rolf, B.; Jackson, I.; Müller, M.; Lang, S.; Reggelin, T.; Ivanov, D. A review on reinforcement learning algorithms and applications in supply chain management. *Int. J. Prod. Res.* **2023**, *61*, 7151–7179. [CrossRef]
- Zou, Y.; Gao, Q.; Wu, H.; Liu, N. Carbon-Efficient Scheduling in Fresh Food Supply Chains with a Time-Window-Constrained Deep Reinforcement Learning Model. *Sensors* **2024**, *24*, 7461. [CrossRef]
- Liu, J. Simulation of Cross-border E-commerce Supply Chain Coordination Decision Model Based on Reinforcement Learning Algorithm. In Proceedings of the 2023 International Conference on Networking, Informatics and Computing (ICNETIC), Palermo, Italy, 29–31 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 429–433.
- Li, X.; Zheng, Z. Dynamic pricing with external information and inventory constraint. *Manag. Sci.* **2024**, *70*, 5985–6001.
- Gurkan, M.E.; Tunc, H.; Tarim, S.A. The joint stochastic lot sizing and pricing problem. *Omega* **2022**, *108*, 102577. [CrossRef]
- Li, X.; Li, Y. NEV's supply chain coordination with financial constraint and demand uncertainty. *Sustainability* **2022**, *14*, 1114. [CrossRef]
- Bergemann, D.; Brooks, B.; Morris, S. The limits of price discrimination. *Am. Econ. Rev.* **2015**, *105*, 921–957. [CrossRef]
- Gomes, U.T.; Pinheiro, P.R.; Saraiva, R.D. Dye schedule optimization: A case study in a textile industry. *Appl. Sci.* **2021**, *11*, 6467. [CrossRef]
- Lu, C.; Wu, Y. Optimization of Logistics Information System based on Multi-Agent Reinforcement Learning. In Proceedings of the 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Lalitpur, Nepal, 18–19 January 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 421–426.
- Kavididevi, V.; Monikapreethi, S.; Rajapriya, M.; Juliet, P.S.; Yuvaraj, S.; Muthulekshmi, M. IoT-Enabled Reinforcement Learning for Enhanced Cold Chain Logistics Performance in Refrigerated Transport. In Proceedings of the 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 10–12 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 379–384.
- Zhou, T.; Xie, L.; Zou, C.; Tian, Y. Research on supply chain efficiency optimization algorithm based on reinforcement learning. *Adv. Contin. Discret. Model.* **2024**, *2024*, 51. [CrossRef]
- Liu, X.; Hu, M.; Peng, Y.; Yang, Y. Multi-agent deep reinforcement learning for multi-echelon inventory management. *Prod. Oper. Manag.* **2022**, 10591478241305863. [CrossRef]
- Aboutorab, H.; Hussain, O.K.; Saberi, M.; Hussain, F.K.; Prior, D. Adaptive identification of supply chain disruptions through reinforcement learning. *Expert Syst. Appl.* **2024**, *248*, 123477. [CrossRef]
- Ma, N.; Wang, Z.; Ba, Z.; Li, X.; Yang, N.; Yang, X.; Zhang, H. Hierarchical Reinforcement Learning for Crude Oil Supply Chain Scheduling. *Algorithms* **2023**, *16*, 354. [CrossRef]
- Piao, M.; Zhang, D.; Lu, H.; Li, R. A supply chain inventory management method for civil aircraft manufacturing based on multi-agent reinforcement learning. *Appl. Sci.* **2023**, *13*, 7510. [CrossRef]
- Ma, Z.; Chen, X.; Sun, T.; Wang, X.; Wu, Y.C.; Zhou, M. Blockchain-based zero-trust supply chain security integrated with deep reinforcement learning for inventory optimization. *Future Internet* **2024**, *16*, 163. [CrossRef]
- Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
- Wang, Y.; Wang, J. Research on international logistics supply chain management strategy based on deep reinforcement learning. *Appl. Math. Nonlinear Sci.* **2024**, *9*, 13. [CrossRef]

27. van Hasselt, H.; Guez, A.; Silver, D. Deep Reinforcement Learning with Double Q-Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
28. Wang, Z.; Schaul, T.; Hessel, M.; van Hasselt, H.; Lanctot, M.; de Freitas, N. Dueling Network Architectures for Deep Reinforcement Learning. In Proceedings of the 33rd International Conference on Machine Learning (ICML), New York, NY, USA, 19–24 June 2016; pp. 1995–2003.
29. Schaul, T.; Quan, J.; Antonoglou, I.; Silver, D. Prioritized Experience Replay. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, PR, USA, 2–4 May 2016.
30. Zhu, Z.; Lin, K.; Jain, A.K.; Zhou, J. Transfer learning in deep reinforcement learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 13344–13362. [CrossRef]
31. Padhye, V.; Lakshmanan, K. A deep actor critic reinforcement learning framework for learning to rank. *Neurocomputing* **2023**, *547*, 126314. [CrossRef]
32. Norouzi, A.; Shahpour, S.; Gordon, D.; Shahbakhti, M.; Koch, C.R. Safe deep reinforcement learning in diesel engine emission control. *Proc. Inst. Mech. Eng. Part J. Syst. Control Eng.* **2023**, *237*, 1440–1453. [CrossRef]
33. Prasuna, R.G.; Potturu, S.R. Deep reinforcement learning in mobile robotics—A concise review. *Multimed. Tools Appl.* **2024**, *83*, 70815–70836. [CrossRef]
34. Rahimian, P.; Mihalyi, B.M.; Toka, L. In-game soccer outcome prediction with offline reinforcement learning. *Mach. Learn.* **2024**, *113*, 7393–7419. [CrossRef]
35. Scarponi, V.; Duprez, M.; Nageotte, F.; Cotin, S. A zero-shot reinforcement learning strategy for autonomous guidewire navigation. *Int. J. Comput. Assist. Radiol. Surg.* **2024**, *19*, 1185–1192. [CrossRef]
36. Liu, X.Y.; Xia, Z.; Yang, H.; Gao, J.; Zha, D.; Zhu, M.; Wang, C.D.; Wang, Z.; Guo, J. Dynamic datasets and market environments for financial reinforcement learning. *Mach. Learn.* **2024**, *113*, 2795–2839. [CrossRef]
37. Yu, L.; Guo, Q.; Wang, R.; Shi, M.; Yan, F.; Wang, R. Dynamic offloading loading optimization in distributed fault diagnosis system with deep reinforcement learning approach. *Appl. Sci.* **2023**, *13*, 4096. [CrossRef]
38. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 1998; Volume 1.
39. Yilmaz Eroglu, D. *RL for Multi-Country Supply-Chains (Q-Learning & Deep Q-Network)*; Zenodo: Geneva, Switzerland, 2025. [CrossRef]
40. The Economist. The Big Mac Index—January 2024. Available online: <https://github.com/TheEconomist/big-mac-data/releases/tag/2024-01> (accessed on 15 July 2025).
41. Jeanne, O.; Son, J. To what extent are tariffs offset by exchange rates? *J. Int. Money Financ.* **2024**, *142*, 103015. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Automated Parcel Locker Configuration Using Discrete Event Simulation

Eugen Rosca, Floriana Cristina Oprea, Anamaria Ilie, Stefan Burciu and Florin Rusca *

Transport, Traffic and Logistics Department, National University of Science and Technology Politehnica Bucharest, 313 Splaiul Independentei, 060042 Bucharest, Romania; eugen.rosca@upb.ro (E.R.); cristina.oprea@upb.ro (F.C.O.); anamaria.ilie@upb.ro (A.I.); stefan.burciu@upb.ro (S.B.)

* Correspondence: florin.rusca@upb.ro

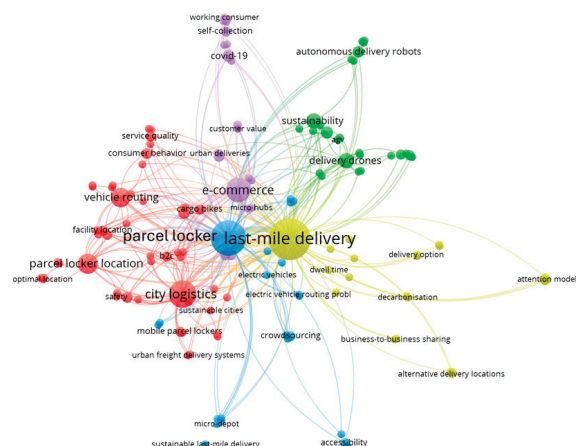
Abstract: Automated parcel lockers (APLs) are transforming urban last-mile delivery by reducing failed distributions, decoupling delivery from recipient availability, optimizing carrier routes, reducing carbon foot-print and mitigating traffic congestion. The paper investigates the optimal design of APLs systems under stochastic demand and operational constraints, formulating the problem as a resource allocation optimization with service-level guarantees. We proposed a data-driven discrete-event simulation (DES) model implemented in ARENA to (i) determine optimal locker configurations that ensure customer satisfaction under stochastic parcel arrivals and dwell times, (ii) examine utilization patterns and spatial allocation to enhance system operational efficiency, and (iii) characterize inventory dynamics of undelivered parcels and evaluate system resilience. The results show that the configuration of locker types significantly influences the system's ability to maintain high customers service levels. While flexibility in locker allocation helps manage excess demand in some configurations, it may also create resource competition among parcel types. The heterogeneity of locker utilization gradients underscores that optimal APLs configurations must balance locker units with their size-dependent functional interdependencies. The Dickey–Fuller GLS test further validates that postponed parcels exhibit stationary inventory dynamics, ensuring scalability for logistics operators. As a theoretical contribution, the paper demonstrates how DES combined with time-series econometrics can address APLs capacity planning in city logistics. For practitioners, the study provides a decision-support framework for locker sizing, emphasizing cost–service trade-offs.

Keywords: automated parcel locker; city logistics; last-mile delivery; batch queueing; capacity optimization; lockers configuration; discrete event simulation; ARENA simulation

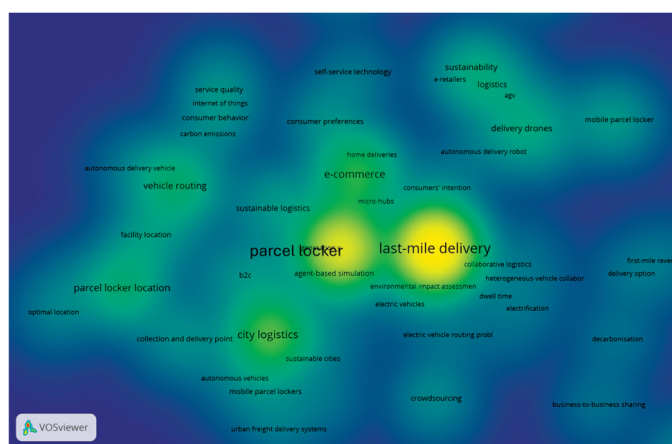
1. Introduction

Automated parcel lockers (APLs) are self-service kiosks that allow people to pick up and drop off packages at any time. As highlighted by the World Economic Forum [1], APLs have emerged as a globally scalable sustainable solution for last-mile delivery, addressing critical externalities while enhancing operational efficiency. To map the research landscape, a bibliometric analysis was conducted using VOSviewer 1.6 [2], examining 453 publications from MDPI and Scopus scientific databases, published between 2000 and 2025. The literature was filtered using the keywords “parcel locker” OR “urban locker” OR “logistics boxes”, revealing five dominant thematic clusters (Figure 1). The first cluster addresses modern city logistics solutions, facility location, vehicle routing, and service quality. The

strategic placement of APLs is paramount for effective urban logistics networks [3–5]. Research demonstrates that optimally located lockers can reduce last-mile delivery distances by 30–50% [6].



(a) key-words network



(b) key-words density

Figure 1. VOSviewer bibliometric analysis on APLs (source: authors).

Placing lockers in the right urban locations is key to making them useful and efficient. Good spots for these lockers include busy areas like shopping centers, transport nodes, and office buildings, where many people pass by daily. Residential neighborhoods also benefit from lockers, as they save delivery trucks from making multiple stops. Other smart locations are near supermarkets, pharmacies, or gas stations, where people already go regularly (Figure 2). Safety is important too, so lockers should be in well-lit, secure areas, possibly with surveillance cameras.



(a) besides commercial area

(b) on street

(c) in residential area

Figure 2. Different APL sites in urban areas (source: authors).

Geographic information systems and machine learning algorithms enable dynamic capacity planning while maintaining consumer accessibility by analyzing population density, transportation hubs, and commercial infrastructure [7,8]. APLs significantly improve routing efficiency through delivery consolidation. Instead of individual home deliveries,

carriers can service multiple locker locations on optimized routes [9–11]. Studies show this reduces vehicle kilometers traveled by 40–60% and decreases CO₂ emissions by 0.5 kg per parcel [12].

The second cluster is constructed around the integration of APLs in last-mile delivery concepts. Users can collect parcels at their convenience within typically 3–5 days, with 24/7 access at most locations [13]. This flexibility benefits both consumers and carriers by decoupling delivery from recipient availability.

The third cluster highlights the role of e-commerce in driving APL adoption. Modern locker systems employ modular designs to accommodate diverse parcel sizes within compact footprints. Advanced systems dynamically assign compartment sizes, improving space utilization by 30–50% compared to traditional methods [14]. Consumers' attitudes toward innovation are pivotal in the adoption of the newest last-mile delivery solutions and is also an effective mediator of consumers' beliefs towards innovation and turning them from passive to active consumers in the last logistic chain [15]. Wang et al. [16] identify compatibility (alignment with needs), complexity (lifestyle fit), and trialability (ease of testing, values) as key determinants in adopting the new delivery systems.

The fourth cluster encompasses sustainable issues. They contribute to sustainability by reducing vehicle emissions and cutting last-mile trips, lower congestion costs, estimated at EUR 50–150 per ton of CO₂ in urban areas, reduces failed deliveries by 15–25% of total attempts [17–19].

The integration of electric and autonomous vehicles into last-mile delivery systems represents the fifth critical cluster of modern urban logistics solutions. As e-commerce volumes continue to grow globally, these advanced vehicle technologies demonstrate significant potential to reduce the environmental impact of delivery operations while maintaining service efficiency [20,21].

A wide bibliometric investigation on urban last-mile logistics [22] revealed some similar findings. The six thematic clusters explored a diverse range of concepts, including the development of supply chain evaluation tools, the study of parcel lockers and delivery system attributes, the assessment of innovative vehicles and infrastructures such as electric vehicles and urban consolidation centers, the application of technology in logistics, and the proposal of collaborative-based business models. Additionally, three categories of solutions were identified: new vehicles (e.g., drones), operational solutions (e.g., off-peak deliveries), and organizational solutions (e.g., crowdsourced logistics).

Although present among the researchers' interest, simulation is still under-utilized in investigating automated parcel lockers area. Iwan, Kijewska and Lemke [6] used a macroscopic agent-based model for predicting consumer demand and behavior. Double reasons are outlined for the efficiency in using APLs: cost reduction for companies and availability and localization for customers. The most important expectations of users regarding localization include close location to home, on the way to work, and availability of parking spaces. A system dynamics simulation and multi-objective decision support were employed to evaluate the sustainability performance of distribution channel options [23]. The results show that the integration of the players into a distributed network strategy based on a crowd logistics concept is the most viable and sustainable option. Che, Chiang and Luo [24] framed APL localization as a multi-objective optimization problem with three key criteria: coverage maximization, overlap minimization, and total idle capacity minimization. This hybrid approach integrates simulation and optimization to enhance decision-making in urban logistics. While their methodology provides valuable insights, it does not account for uncertain consumer demand or incorporate elastic weighted goals. The study notably emphasizes the importance of establishing minimum utilization thresholds to prevent resource underutilization. Other studies used an agent-based model to estimate

future demand based on socio-economic parameters, and a facility location problem (FLP) optimization model to determine the optimal number and placement of lockers [25,26]. An agent-based model is designed to determine the performance of the APL network and to estimate the future demand based on a few socio-economic parameters. Rabe et al. [27] formulated a multi-period CFLP model to provide the optimal number of APLs to be installed and Monte Carlo simulation was used to estimate the cost and reliability level for different scenarios with random demands. Izco Berastegui et al. [28] used an integrated dynamic simulation to forecast population and e-commerce demand and Pareto Frontier to trade-off between the number of lockers and the desired demand satisfaction of the decision makers. Iannaccone, Marcucci, and Gatta [29] proposed a discrete choice model that allows both to quantify the monetary value of parcel lockers attributes (willingness to pay) and to estimate the potential demand. Using macro-level data, Savik [30] provides an in-depth examination of the potential of automated parcel lockers, capillary distribution, and crowd-shipping as developing solutions to last-mile delivery. Understanding personal values and consumer traits play a significant role in APL success. Utilitarian and hedonic attitudes toward using innovative technologies are determinant through the consumer traits (self-service efficacy) [31]. De Oliveira et al. [32] proved that location, delivery time, information and traceability, cost of transportation, and the willingness to use self-service technologies are determinant factors in estimating the use of APLs. Cieřla [33] found that dedicated application, adjusting the size of the package to the size of the box, placing the parcel in a specific box, security, and parking next to APLs are among attributes that provide customers satisfaction. Lai et al. [34], using the service quality (SERVQUAL) and logistics service quality (LSQ) model showed that timeliness is the strongest predictor that positively impacts customer satisfaction with APLs, followed by reliability, security responsiveness and tangibility.

Existing research on automated parcel lockers predominantly adopts macro- or mezo-scale perspectives, emphasizing network-level optimization, such as facility location, regional/local demand allocation, or system-wide throughput. However, limited attention has been paid to micro-scale operational challenges, particularly on (i) optimal capacity dimensioning (the minimal capacity to handle stochastic parcel arrivals and heterogeneous size distributions), (ii) configuration optimization (strategic selection of locker-type ratios under space and cost constraints), and (iii) dynamic utilization (real-time allocation policies to mitigate capacity shortage during demand fluctuations). This gap is critical, as sub-optimal station-level design directly impacts user satisfaction (e.g., failed deliveries) and operational costs (e.g., overprovisioning, excess inventories). While macro-level studies assume homogeneous demand, at micro-level APLs face nonstationary and stochastic demand patterns—necessitating granular modeling and simulation approaches. Building on the identified gap, this study addresses the following research questions:

1. How can automated parcel locker stations be optimally configured—in terms of the number and mix of locker types (small, medium, large)—to accommodate stochastic demand flows characterized by random arrival rates, heterogeneous parcel size distributions, stochastic dwell time, and operational flexibility (dynamic parcel-to-locker allocation), while maximizing service-level performance (delivery success rate)?
2. Do delivery quality standards (on-time delivery rate) affect the lockers' feasible configuration?
3. How do the key performance metrics for companies (locker utilization rate, variation of undelivered parcels inventory) balance with customers' satisfaction and experience?

The main contributions of this study are as follows:

- It develops a methodological framework and simulation tool to optimize APL configuration.
- It demonstrates how stochastic demand and heterogeneous parcel sizes limit the possible locker configurations while balancing resource capacity and customer satisfaction in terms of on-time successful delivery.
- It integrates discrete event simulation and time-series econometrics to investigate the company's operational performance metrics (including locker utilization rates and fluctuations in undelivered parcel inventory).
- It provides actionable recommendations for improving capacity planning for APL last-mile deliveries.

Section 2 presents a discrete event simulation (DES) model developed in ARENA software, which captures microscopic operational dynamics of APLs. Section 3 completes experimental computation using experimental test data. Section 4 analyzes simulation outcomes. The on-time delivery probability is the key performance measure to assess the system's effectiveness. Results identify feasible APL configurations that maintain on-time delivery probability within target thresholds. Furthermore, time-series analysis of postponed parcel flows confirms stationarity, suggesting that inventory overflow does not critically disrupt operations. Conclusions and recommendations are summarized in the last section.

2. Modelling and Simulation Methodology

Discrete event simulation is a useful technique in transportation and logistics, enabling businesses and governments to optimize operations, reduce costs, and enhance efficiency. By simulating real-world scenarios before implementation, organizations can make smarter decisions and build more resilient systems. It focuses on individual events that occur at specific points in time, making it ideal for modeling dynamic processes such as vehicle movements [35], terminal [36] and warehouse operations [37], supply chain flows [38,39], last-mile delivery [40], and e-commerce logistics [41]. Testing automated parcel lockers through computer simulations before installing them in real life helps cities and delivery companies make smarter decisions. By simulating different capacity options first, cities can install locker systems that work efficiently every day while still managing busy periods—all without wasting space or money. This smart planning leads to the following:

- Preventing overcrowding (predicts how many compartments are necessary at each location, helps avoid full lockers that annoy customers);
- Saving space in busy areas (determines the smallest possible locker size that still meets demand, important for dense urban areas where space is limited);
- Improving delivery efficiency (shows the right balance between small, medium, and large compartments, reduces wasted space from wrong-sized compartments);
- Saving money (prevents buying oversized lockers that are not fully used, avoids the cost of adding more lockers later);
- Matching local needs (customizes capacity based on neighborhood characteristics, residential areas might need different setups than business districts).

The developed simulating model for an automated parcel locker contains four entity types and three resource (server) types (Table 1).

Table 1. Entities and resources in the simulation model.

Item	Attribute	Type	Description
Entities	Serial number	Number	Entity ID.
	Size	Text	The size of the parcel to match the right locker. There are three types of parcels: large, medium, small.
	Creation time	Time stamp	The generation time.
	Dwell time	Number	The time interval a parcel is stored into a locker.
Delivery person	Delivery moment	Time stamp	The moment the delivery process starts.
Resources	Size	Text	There are three types of lockers: large, medium, small. The conventional unit locker has the dimensions of a small locker. The medium locker is equivalent to two conventional lockers, and the large locker with four conventional units.
	State	Logical (Empty/Occupied)	Empty lockers are available for parcel storage, and a locker remains occupied for the duration of the parcel's dwell time.

The parcels are passive entities that are generated and placed into lockers, while the delivery person is an active entity that assesses the state of the system (the number of parcels to be delivered, the state of the lockers) and takes actions (deliver the parcel to a specific locker or postpone the delivery). There are three types of lockers, according to their dimensions (Figure 2). Similarly, parcels are classified into three types, with each type fitting into the corresponding locker or a larger one. Large parcels are accommodated solely by large lockers. Medium parcels are prioritized for medium lockers (high priority) but may be placed in available large lockers if necessary (low priority comparing to large parcels). Small parcels are prioritized for small lockers but can be accommodated in both medium or large lockers if available and required (low priority compared to medium and large parcels). Every day, a specific number of parcels are generated, matching client demand. The automated parcel locker acts as a batch queuing system, with deliveries scheduled at a specific time of day (Figure 3). At this designated time, all parcels that fit appropriately within the lockers are deposited simultaneously, and the lockers are subsequently released after an individual dwell time, corresponding to their pickup by customers. The availability of compartments over time depends on the customers' pickup behavior that is stochastic. If a parcel is delivered on the same day as is generated, the delivery is considered on time. Otherwise, undelivered parcels must wait at the central warehouse until the next available delivery opportunity (on subsequent days).

The parcel processing flow-chart is depicted in Figure 3.

The activity of the automated parcel locker model is summarized as follows:

Parameters:

L —the number of conventional lockers;

L_l, L_m, L_s —the number of large/medium/small lockers;

$N_{l,j}, N_{m,j}, N_{s,j}$ —the number of large/medium/small parcels generated on day j ($j \geq 1$);

$n_{l,j}, n_{m,j}, n_{s,j}$ —the number of large/medium/small parcels delivered on day j ;

$n_{l,j,k}, n_{m,j,k}, n_{s,j,k}$ —the number of large/medium/small parcels delivered on day j that are picked up after k days ($k = 0$ picking the day of delivery, $k = 1$ picking after one day, $k = 2$ picking after two days from delivery);

$n_{m-l,j}$ —the number of medium parcels stored in large lockers on day j ;

$n_{s-m,j}$, $n_{s-l,j}$ —the number of small parcels stored in medium/large lockers on day j ;
 $n_{m-l,j,k}$ —the number of medium parcels stored in large lockers on day j that are picked up after k days;
 $n_{s-m,j,k}$, $n_{s-l,j,k}$ —the number of small parcels stored in medium/large lockers on day j that are picked up after k days;
 $I_{l,j}$, $I_{m,j}$, $I_{s,j}$ —the inventory of undelivered large/medium/small parcels at the central warehouse at the end of day j ;
 $L_{l,j}^f$, $L_{m,j}^f$, $L_{s,j}^f$ —the number of free large/medium/small lockers at the beginning of day j ;
 $L_{m-l,j}^f$ —the number of empty large lockers available for storing medium parcels on day j ;
 $L_{s-l,j}^f$, $L_{s-m,j}^f$ —the number of empty large/medium lockers available for storing small parcels on day j ;
 $P(k) = [P(0), P(1), P(2)]$ —the probabilities vector for the parcels' dwell time ($k = 0$ the parcel picked up the same day as the delivery day, $k = 1$ the parcel is picked up on the subsequent day, $k = 2$ the parcel is picked up two days after the delivery day).

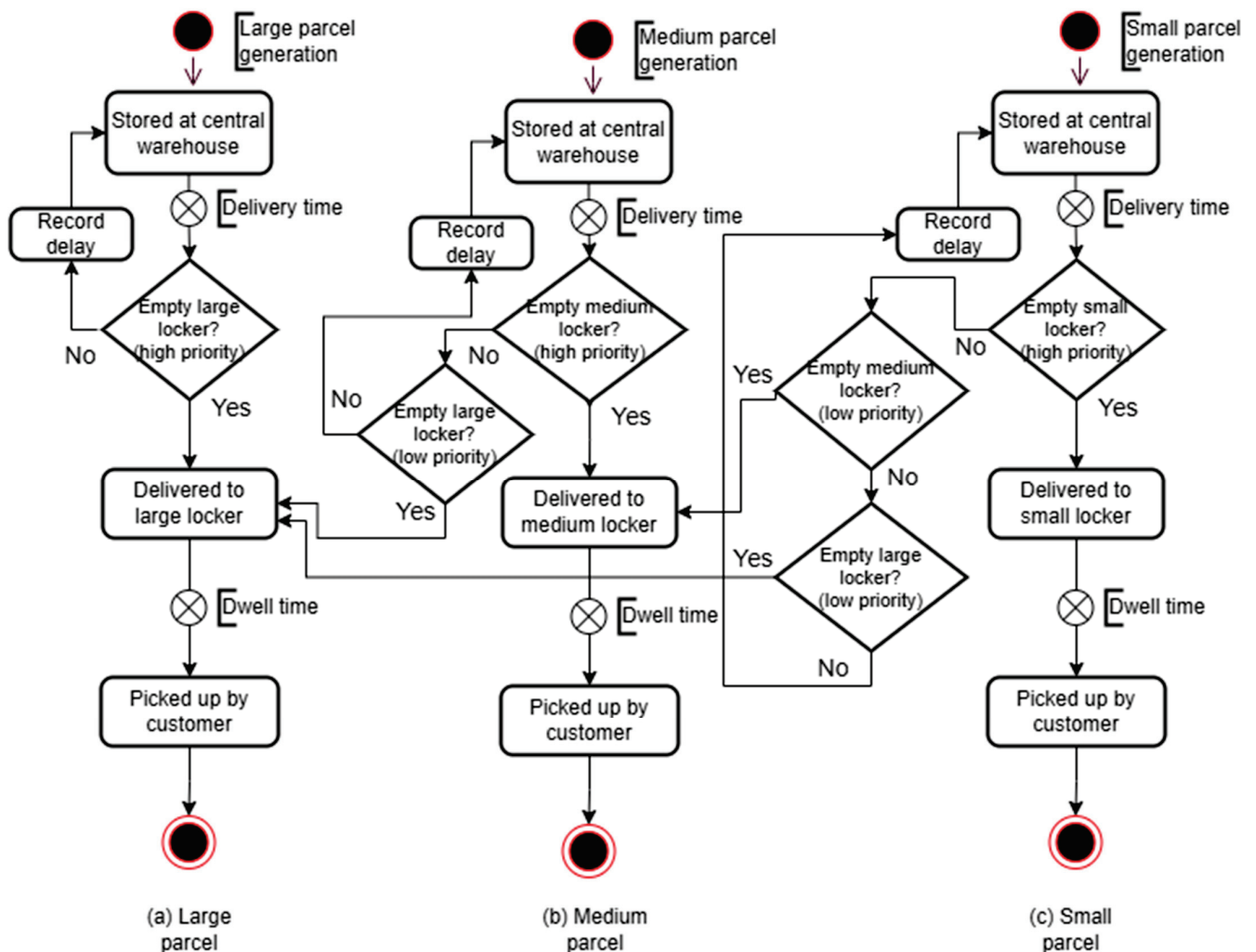


Figure 3. Parcel processing flow-chart (source: authors).

System process equations:

$$n_{l,j} = \min(I_{l,j-1} + N_{l,j}, L_{l,j}^f) \quad (1)$$

$$L_{l,j}^f = L_l - n_{l,j-1,1} - n_{l,j-1,2} - n_{l,j-2,2} - n_{m-l,j-1,1} - n_{m-l,j-1,2} - n_{m-l,j-2,2} - n_{s-l,j-1,1} - n_{s-l,j-1,2} - n_{s-l,j-2,2} \quad (2)$$

$$I_{l,j} = I_{l,j-1} + N_{l,j} - n_{l,j} \quad (3)$$

$$n_{m,j} = \min(I_{m,j-1} + N_{m,j}, L_{m,j}^f + L_{m-l,j}^f) \quad (4)$$

$$L_{m,j}^f = L_m - n_{m,j-1,1} - n_{m,j-1,2} - n_{m,j-2,2} - n_{s-m,j-1,1} - n_{s-m,j-1,2} - n_{s-m,j-2,2} \quad (5)$$

$$L_{m-l,j}^f = L_{l,j}^f - n_{l,j} \quad (6)$$

$$n_{m-l,j} = \max(0, n_{m,j} - L_{m,j}^f) \quad (7)$$

$$I_{m,j} = I_{m,j-1} + N_{m,j} - n_{m,j} \quad (8)$$

$$n_{s,j} = \min(I_{s,j-1} + N_{s,j}, L_{s,j}^f + L_{s-m,j}^f + L_{s-l,j}^f) \quad (9)$$

$$L_{s,j}^f = L_s - n_{s,j-1,1} - n_{s,j-1,2} - n_{s,j-2,2} \quad (10)$$

$$L_{s-m,j}^f = \max(0, L_{m,j}^f - n_{m,j}) \quad (11)$$

$$L_{s-l,j}^f = L_{l,j}^f - n_{l,j} - n_{m-l,j} \quad (12)$$

$$n_{s-m,j} = \min(\max(0, n_{s,j} - L_{s,j}^f), L_{s-m,j}^f) \quad (13)$$

$$n_{s-l,j} = \min(\max(0, n_{s,j} - L_{s,j}^f - L_{s-m,j}^f), L_{s-l,j}^f) \quad (14)$$

$$I_{s,j} = I_{s,j-1} + N_{s,j} - n_{s,j} \quad (15)$$

$$n_{l,j} = n_{l,j,0} + n_{l,j,1} + n_{l,j,2} \quad (16)$$

$$n_{m,j} = n_{m,j,0} + n_{m,j,1} + n_{m,j,2} \quad (17)$$

$$n_{s,j} = n_{s,j,0} + n_{s,j,1} + n_{s,j,2} \quad (18)$$

$$L = 4L_l + 2L_m + L_s \quad (19)$$

Initial values:

$$I_{l,0} = I_{m,0} = I_{s,0} = 0$$

$$n_{l,j \leq 0,k} = n_{m,j \leq 0,k} = n_{s,j \leq 0,k} = 0 \quad (k = 0, 1, 2)$$

Equation (1) determines the quantity of large parcels delivered by calculating the minimum between the sum of the previous day's inventory of large parcels at the central warehouse and the newly arrived large parcels, and the number of available large lockers. The available large lockers (Equation (2)) are calculated by subtracting from the total number of large lockers those currently occupied by large, medium, or small parcels that have not yet been collected (i.e., those with a longer dwell time). Equations (3), (8) and (15) establish the inventory levels of large, medium, and small parcels at the central warehouse at the end of each day, based on the previous day's inventory, the parcels generated daily, and those delivered. Equation (4) calculates the number of medium parcels delivered as the minimum between the available parcels (comprising the previous day's inventory and

newly generated parcels) and the sum of available medium lockers and large lockers that can accommodate medium parcels. The number of available medium lockers is determined by subtracting from the total number of medium lockers those still occupied by uncollected parcels due to their longer dwell time (Equation (5)). The available number of large lockers that can accommodate medium parcels is computed after allocating the large parcels to lockers (Equation (6)). If the number of medium parcels delivered exceeds the number of available medium lockers, the surplus is stored in the available large lockers (Equation (7)). Equation (9) assesses the number of small parcels to be delivered by comparing the existing parcels at the central warehouse with the available small, medium, and large lockers that can accommodate them. The number of small lockers available is the total number of small lockers minus those occupied by parcels that have not yet been collected (Equation (10)). The available number of medium lockers that can accommodate small parcels is derived from the total number of free medium lockers minus those occupied by medium parcels (Equation (11)). Equation (12) calculates the available number of large lockers that can store small parcels as the difference between the number of free large lockers and those used by large and medium parcels. Equations (13) and (14) assess the number of small parcels accommodated by medium and large lockers, respectively. Equations (16)–(18) indicate that the daily delivery of large, medium, and small parcels includes items with different dwell times, as detailed in Table 1. Considering their dimensions, the total number of conventional lockers is equal to the weighted sum of the number of physical lockers (Equation (19)).

The simulation model was developed using ARENA 8.0 simulation software. The model is depicted in Figures 4 and 5.

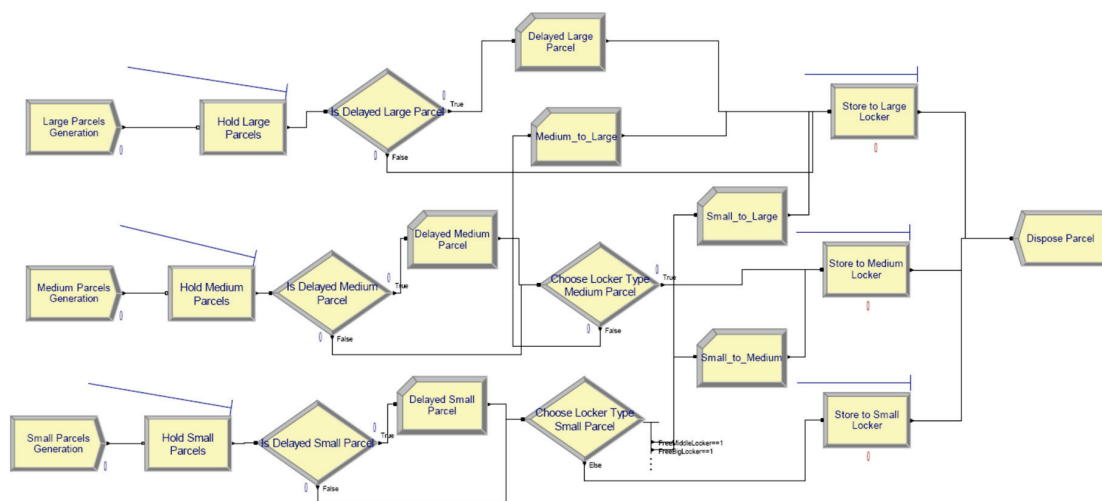


Figure 4. Parcel processing modules (source: authors).

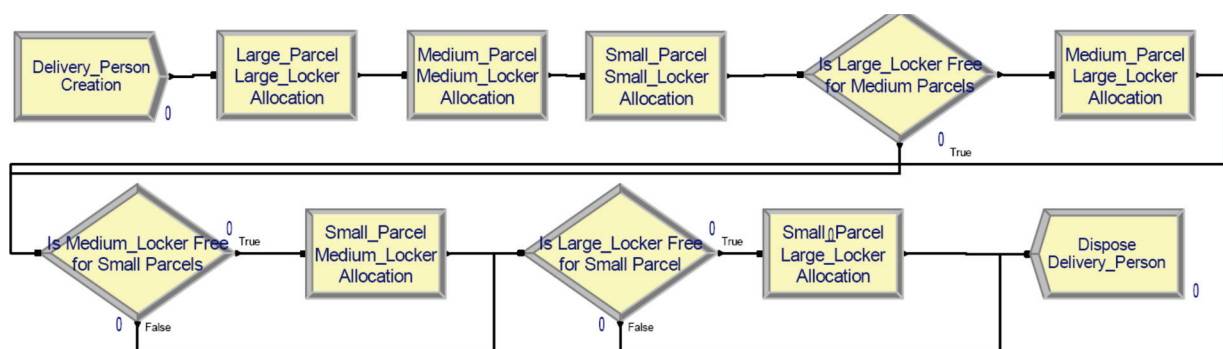


Figure 5. Delivery person acting modules (source: authors).

The modules used in the simulation model are thoroughly described in Table 2.

Table 2. ARENA simulation modules description.

Module	Type	Role and Parameters
Large Parcels Generation	Create	Every day a random number of large parcels ($N_{l,j}$) are generated.
Medium Parcels Generation	Create	Every day a random number of medium parcels ($N_{m,j}$) are generated.
Small Parcels Generation	Create	Every day a random number of small parcels ($N_{s,j}$) are generated.
Hold Large Parcels	Hold	The large parcels are stored at the central warehouse ($I_{l,j-1} + N_{l,j}$) until empty lockers are available. The delivery person controls the parcel releasing process ($n_{l,j}$) by calculating the number of empty large lockers ($L_{l,j}^f$).
Hold Medium Parcels	Hold	The medium parcels are stored at the central warehouse ($I_{m,j-1} + N_{m,j}$) until empty lockers are available. The delivery person controls the parcel releasing process ($n_{m,j}$) by calculating the number of empty medium lockers ($L_{m,j}^f$) or remaining empty large lockers ($L_{m-l,j}^f$) after the accommodation of large parcels.
Hold Small Parcels	Hold	The small parcels are stored at the central warehouse ($I_{s,j-1} + N_{s,j}$) until empty lockers are available. The delivery person controls the parcel releasing process ($n_{s,j}$) by calculating the number of empty small lockers ($L_{s,j}^f$) or remaining empty medium ($L_{s-m,j}^f$) and large lockers ($L_{s-l,j}^f$) after the accommodation of large and medium parcels.
Is Delayed Large Parcel	Decide	It examines whether large parcels are delivered on the day they are generated or if delivery is postponed due to the unavailability of free lockers. It analyzes the time interval between the generation of the parcel and its actual delivery ($TNOW - Entity.CreateTime$).
Is Delayed Medium Parcel	Decide	It examines whether medium parcels are delivered on the day they are generated or if delivery is postponed due to the unavailability of free lockers. It analyzes the time interval between the generation of the parcel and its actual delivery ($TNOW - Entity.CreateTime$).
Is Delayed Small Parcel	Decide	It examines whether small parcels are delivered on the day they are generated or if delivery is postponed due to the unavailability of free lockers. It analyzes the time interval between the generation of the parcel and its actual delivery ($TNOW - Entity.CreateTime$).
Delayed Large Parcel	Record	It records the delayed large parcels by incrementing the counter Delayed_Large_Parcel.
Delayed Medium Parcel	Record	It records the delayed large parcels by incrementing the counter Delayed_Medium_Parcel.
Delayed Small Parcel	Record	It records the delayed large parcels by incrementing the counter Delayed_Small_Parcel.
Choose Locker Type Medium Parcel	Decide	An appropriate locker for the medium parcel is selected based on availability. If an empty medium locker is available, the parcel is stored there (higher priority). Otherwise, the parcel is placed in a free large locker (lower priority competing to large parcels).

Table 2. Cont.

Module		Type	Role and Parameters
Parcels	Choose Locker Type Small Parcel	Decide	An appropriate locker for the small parcel is selected based on availability. If an empty small locker is available, the parcel is stored there (higher priority). Otherwise, the parcel is placed in a free medium or a free large locker (lower priority competing to medium or large parcels).
	Medium to Large	Record	It records the number of medium parcels accommodated in large lockers by incrementing the counter $n_{m-l,j}$.
	Small to Medium	Record	It records the number of small parcels accommodated in medium lockers by incrementing the counter $n_{s-m,j}$.
	Small to Large	Record	It records the number of small parcels accommodated in large lockers by incrementing the counter $n_{s-l,j}$.
	Store to Large Locker	Process	The large locker is seized, and its state becomes occupied until the parcel's dwell time elapses. Following this time, the parcel is picked up by customer and the locker is released.
	Store to Medium Locker	Process	The medium locker is seized, and its state becomes occupied until the parcel's dwell time elapses. Following this time, the parcel is picked up by customer and the locker is released.
	Store to Small Locker	Process	The small locker is seized, and its state becomes occupied until the parcel's dwell time elapses. Following this time, the parcel is picked up by customer and the locker is released.
	Dispose Parcel	Dispose	The parcels are disposed once they are picked up.
Delivery Person	Delivery Person Creation	Create	It creates the delivery person as an entity in the system. The delivery person starts the activity daily at a conventional moment 0.
	Large Parcel Large Locker Allocation	Signal	The large parcels are allowed to be transferred to the large lockers. The signal is transmitted to Hold Large Parcels module and the appropriate number of large parcels are released (see Equation (1)).
	Medium Parcel Medium Locker Allocation	Signal	The medium parcels are allowed to be transferred to the medium lockers. The signal is transmitted to Hold Medium Parcels module and the appropriate number of medium parcels are released (see Equation (4)).
	Small Parcel Small Locker Allocation	Signal	The small parcels are allowed to be transferred to the small lockers. The signal is transmitted to Hold Small Parcels module and the appropriate number of small parcels are released (see Equation (9)).
	Is Large Locker Free for Medium Parcels	Decide	The delivery person checks the availability of unoccupied large lockers ($L_{m-l,j}^f$) to determine whether medium parcels can be accommodated.
	Medium Parcel Large Locker Allocation	Signal	If there are free large lockers, a signal is emitted, and the module Hold Medium Parcels releases the adequate number of medium parcels to be delivered into large lockers (see Equation (7)).
	Is Medium Locker Free for Small Parcels	Decide	The delivery person checks the availability of unoccupied medium lockers ($L_{s-m,j}^f$) to determine whether small parcels can be accommodated.
	Small Parcel Medium Locker Allocation	Signal	If there are free medium lockers, a signal is emitted, and the module Hold Small Parcels releases the adequate number of medium parcels to be delivered into medium lockers (see Equation (13)).

Table 2. Cont.

	Module	Type	Role and Parameters
Delivery Person	Is Large Locker Free for Small Parcels	Decide	The delivery person checks the availability of unoccupied large lockers ($L_{s-1,j}^f$) to determine whether small parcels can be accommodated.
	Small Parcel Large Locker Allocation	Signal	If there are free large lockers, a signal is emitted, and the module Hold Small Parcels releases the adequate number of medium parcels to be delivered into large lockers (see Equation (4)).
	Dispose Delivery Person	Dispose	The delivery person is disposed after it completes its activity.

In its daily operations, the automated parcel locker system functions under stochastic conditions. The daily generation of large, medium, and small parcels follows a random distribution, and the dwell time of parcels is likewise a stochastic variable. The system's efficiency is measured by its ability to ensure prompt parcel delivery, prevent overcrowding, and minimize delivery postpones.

An optimal allocation of locker types (large, medium, and small) is essential to efficiently accommodate parcels, reduce wasted space from improperly sized compartments, and avoid unnecessary expansion. Given a fixed capacity of conventional lockers (i.e., the predefined space allocated to the automated parcel locker system), a series of simulation experiments were conducted to determine the optimal proportion of large, medium, and small lockers. The objective was to ensure that the probability of on-time parcel delivery exceeded a specified threshold, thereby maximizing customer satisfaction.

3. Computational Experiments and Results

The simulation experiments were conducted using the data in Table 3.

Table 3. Simulation model input data.

Item	Data	Type	Values/Range of Variation
Large parcel	Daily generation number	Uniform	[0, 4]
	Dwell time	Stochastic discrete distribution	$p(k) = [0.80; 0.15; 0.05]$, $k = 0, 1, 2$
Medium parcel	Daily generation number	Uniform	[2, 8]
	Dwell time	Stochastic discrete distribution	$p(k) = [0.80; 0.15; 0.05]$, $k = 0, 1, 2$
Small parcel	Daily generation number	Uniform	[20, 36]
	Dwell time	Stochastic discrete distribution	$p(k) = [0.80; 0.15; 0.05]$, $k = 0, 1, 2$
Delivery person	Interval between deliveries	Constant	24 h
Conventional lockers	Capacity (number of lockers)	Constant	70

Parcels are generated daily following uniform distributions with varying limits $U[a,b]$. The parcel pick-up probabilities are as follows: 80% of the parcels are collected on the same day as delivery ($p(0) = 0.80$), 15% on the subsequent day ($p(1) = 0.15$), and 5% two days after delivery ($p(2) = 0.05$). The delivery person dispatches parcels once per day (24 h interval between deliveries). The automated locker system has a total capacity equivalent to

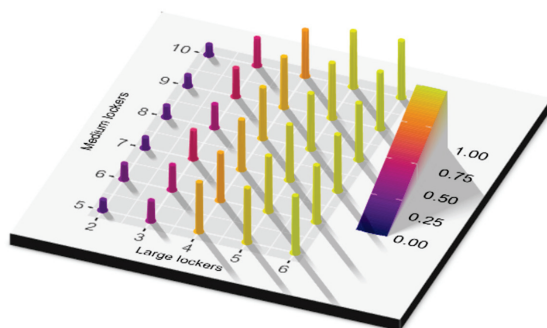
70 conventional lockers, with the number of large, medium, and small lockers constrained by Equation (19):

$$70 = 4L_l + 2L_m + L_s.$$

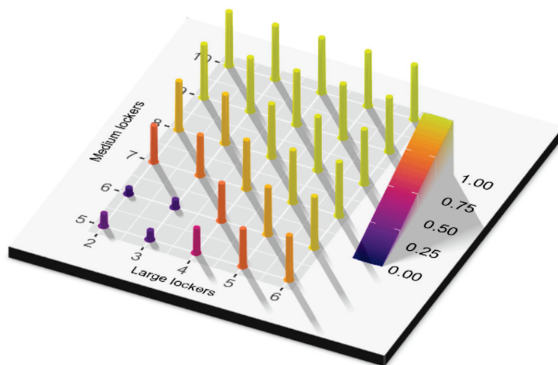
The simulation experiments evaluate the activity of the automated parcel lockers under varying locker-type configurations (i.e., differing proportions of locker sizes). Each simulation runs for a 365-day period. The following outputs are recorded:

- The number of generated parcels per type;
- The parcels' dwell time;
- The number of parcels delivered on-time;
- The undelivered parcel inventory at the end of each day;
- The number of parcels distributed in distinct size lockers.

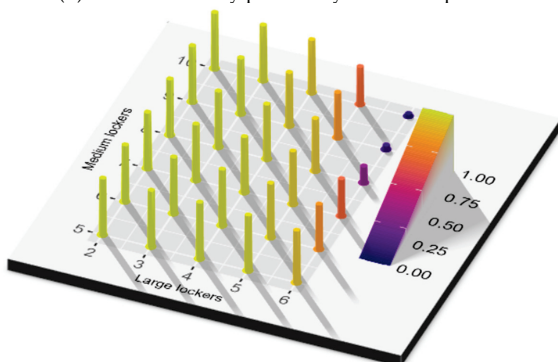
The on-time delivery probability is a key performance measure for both customer satisfaction and company efficiency. Figure 6 illustrates the on-time parcel delivery probabilities across different automated locker configurations $\langle L_l, L_m, L_s \rangle$.



(a) On-time delivery probability—large parcels



(b) On-time delivery probability—medium parcels



(c) On-time delivery probability—small parcels

Figure 6. On-time delivery probabilities (source: authors).

The on-time delivery probability for large parcels falls below 0.5 with 2–3 large lockers but improves significantly (≥ 0.8) when 4 or more large lockers are available (Figure 6a). Medium size parcels can be deposited in both medium and large lockers. For configurations with 2–3 large lockers and 5–6 medium lockers, the on-time delivery probability for medium parcels falls below 0.5 (Figure 6b). However, as the number of available large and medium lockers increases, the on-time delivery probability for medium parcels rises correspondingly. When 6 large lockers and 8–10 medium lockers are available, the on-time delivery probability for small parcels drops below 0.5 (Figure 6c). This probability remains below 0.8 in configurations with either (i) 6 large lockers and 6–7 medium lockers, or (ii) 5 large lockers and 10 medium lockers. In all other analyzed configurations, the probability exceeds 0.8.

The margins of error for the on-time delivery probabilities are computed based on the sample parcels data are

$$\varepsilon = z * \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (20)$$

where

ε is the margin of error of the probability;

$\hat{p}$ —is the estimated probability for on-time delivery;

n —is the sample size;

z —is the z score ($z = 1.96$ for 95% confidence interval).

Considering the above estimated successful delivery probabilities, the margins of error are depicted in Figure 7.

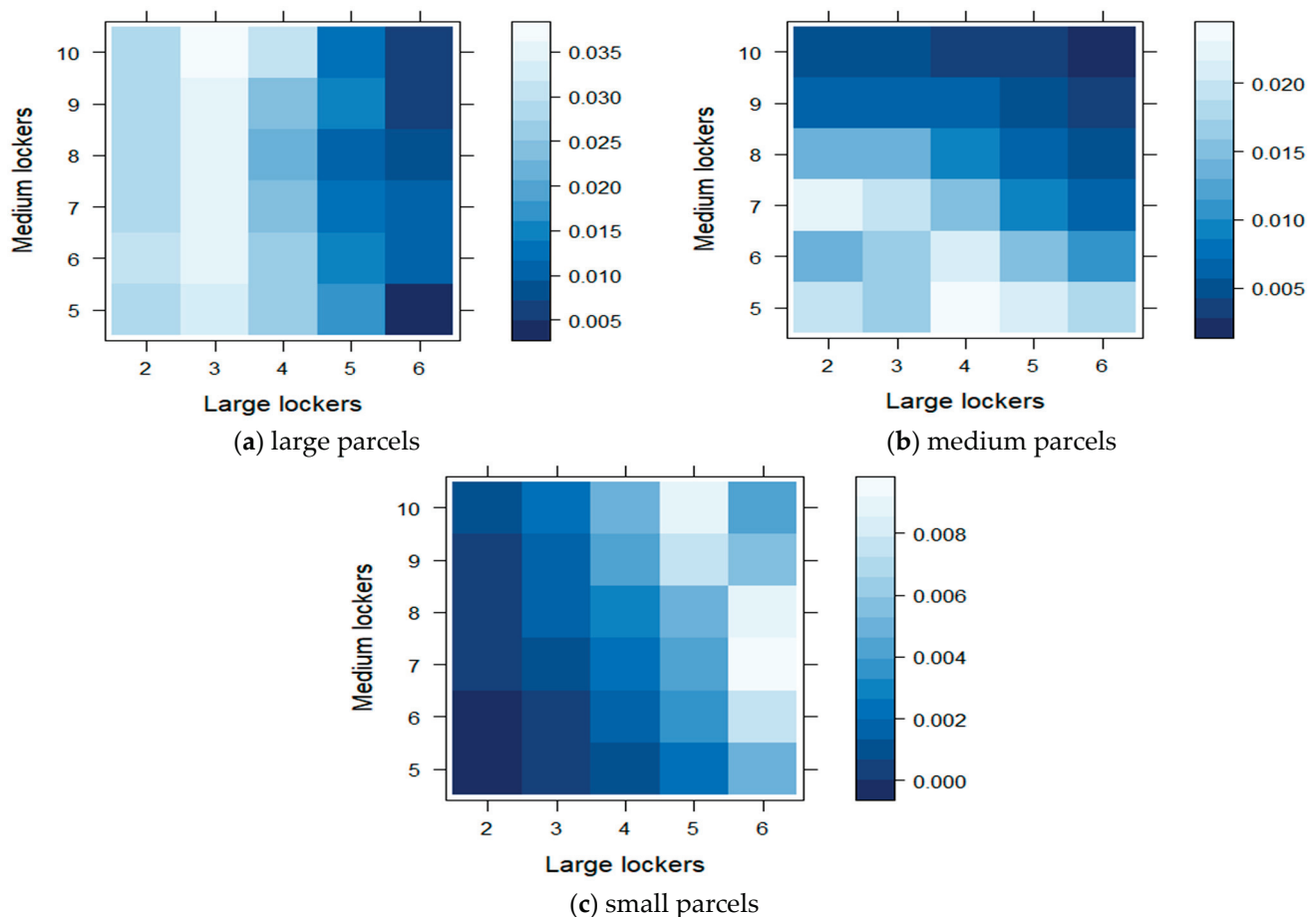


Figure 7. Margins of error for on-time delivery probabilities (source: authors).

Margins of error allow us to accurately assess whether a given APL configuration meets the desired level of successful delivery, accounting for potential fluctuations due to statistical variability. An increase in sample size is associated with a reduction in the margin of error. Specifically, for large parcels, the margin of error typically ranges between 0.5% and 3.5%, whereas for small parcels, it falls below 0.8%. When the successful delivery target falls within the confidence interval of the on-time delivery probability, the decision regarding the feasibility of the configuration becomes inconclusive.

The ability to store parcels in larger lockers enables the system to effectively handle periods of high demand, leading to better utilization of storage capacity and an increased rate of successful deliveries. Figure 8 shows the storage distribution patterns: (a) medium parcels allocated to large lockers, and (b) small parcels distributed across both medium and large lockers. The analysis reveals that configurations with limited medium locker availability (5 units) and substantial large locker capacity (4–6 units) result in over 365 medium parcels being stored in large lockers (Figure 8a). Furthermore, as the number of both medium and large lockers increases, a corresponding rise occurs in the number of small parcels stored across available medium and large lockers (Figure 8b).

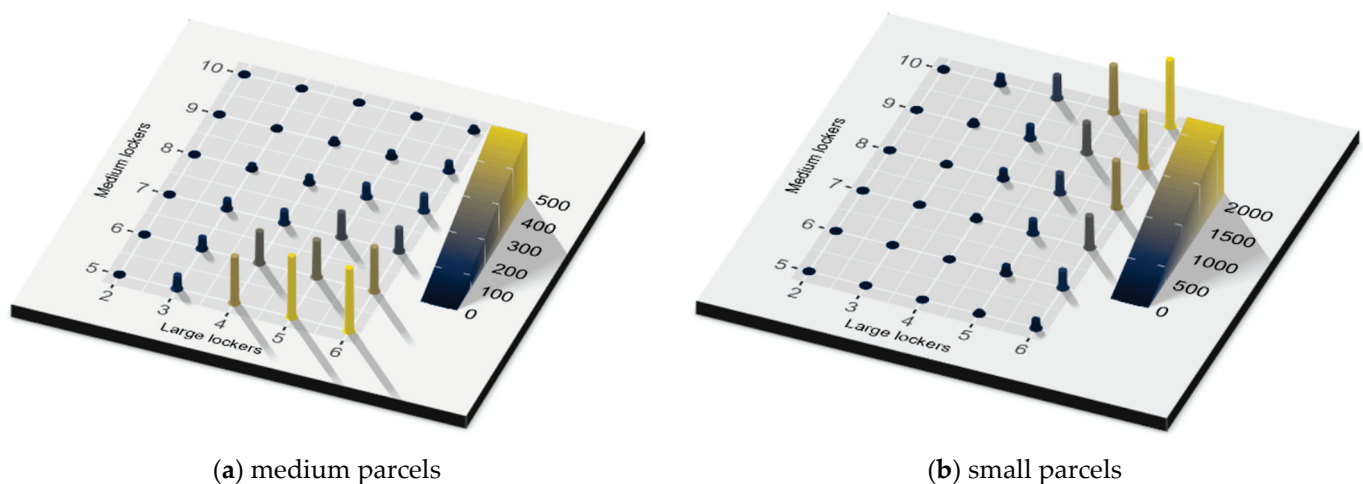


Figure 8. Relocated medium and small parcels to bigger-size lockers (source: authors).

From the perspective of the delivery company, system efficiency is determined by locker utilization rates. Low utilization rates indicate system overcapacity, whereas high rates reflect excessive demand pressure. The ability to allocate parcels across different sized lockers enhances the system's adaptability to demand fluctuations. Based on the input variables, Figure 9 presents the corresponding locker utilization rates. The utilization rate gradient for large lockers exhibits a central-to-peripheral orientation, whereas for medium and small lockers, the gradient is oriented bilaterally (along diagonal axes).

These trends underscore the importance of balanced locker allocation to meet delivery targets, which exhibit flexibility in storage but sensitivity to resource scarcity. By establishing thresholds for on-time delivery probabilities across all parcel types, optimal configurations of automated parcel lockers can be identified (Figure 10). This threshold-based approach enables systematic evaluation of locker system performance under varying operational constraints.

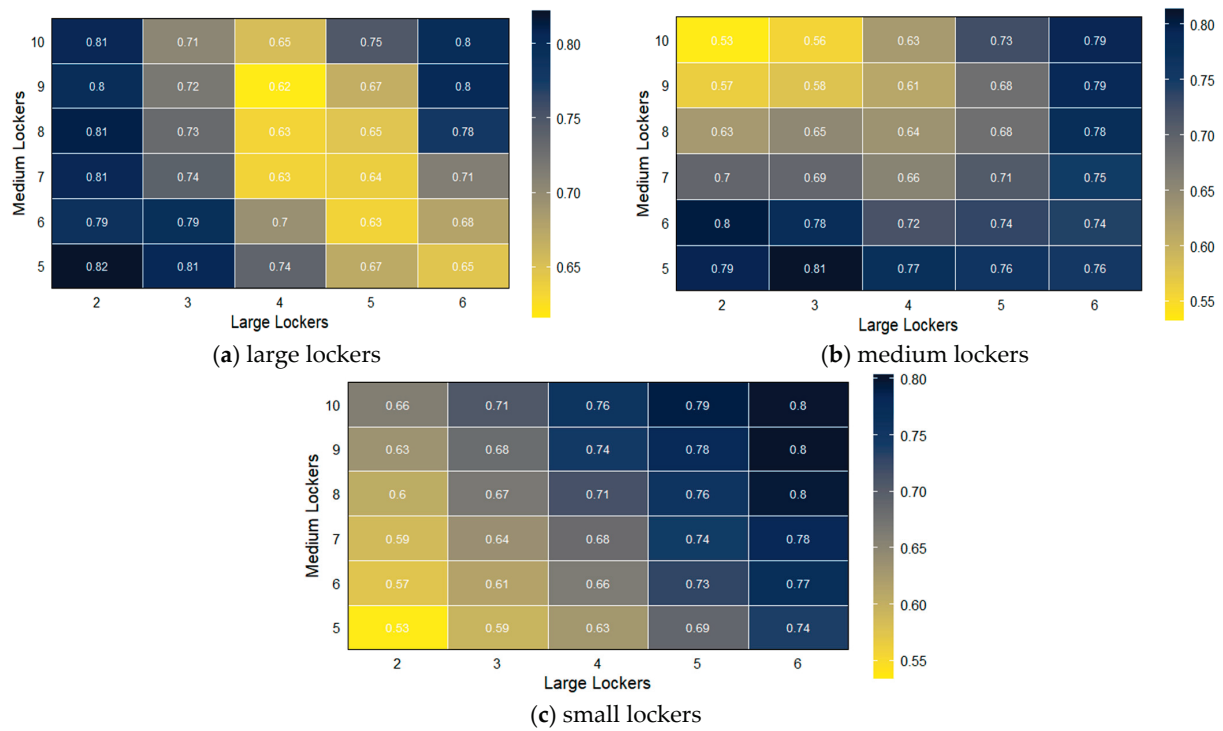


Figure 9. Locker utilization rate by locker size combination (source: authors).

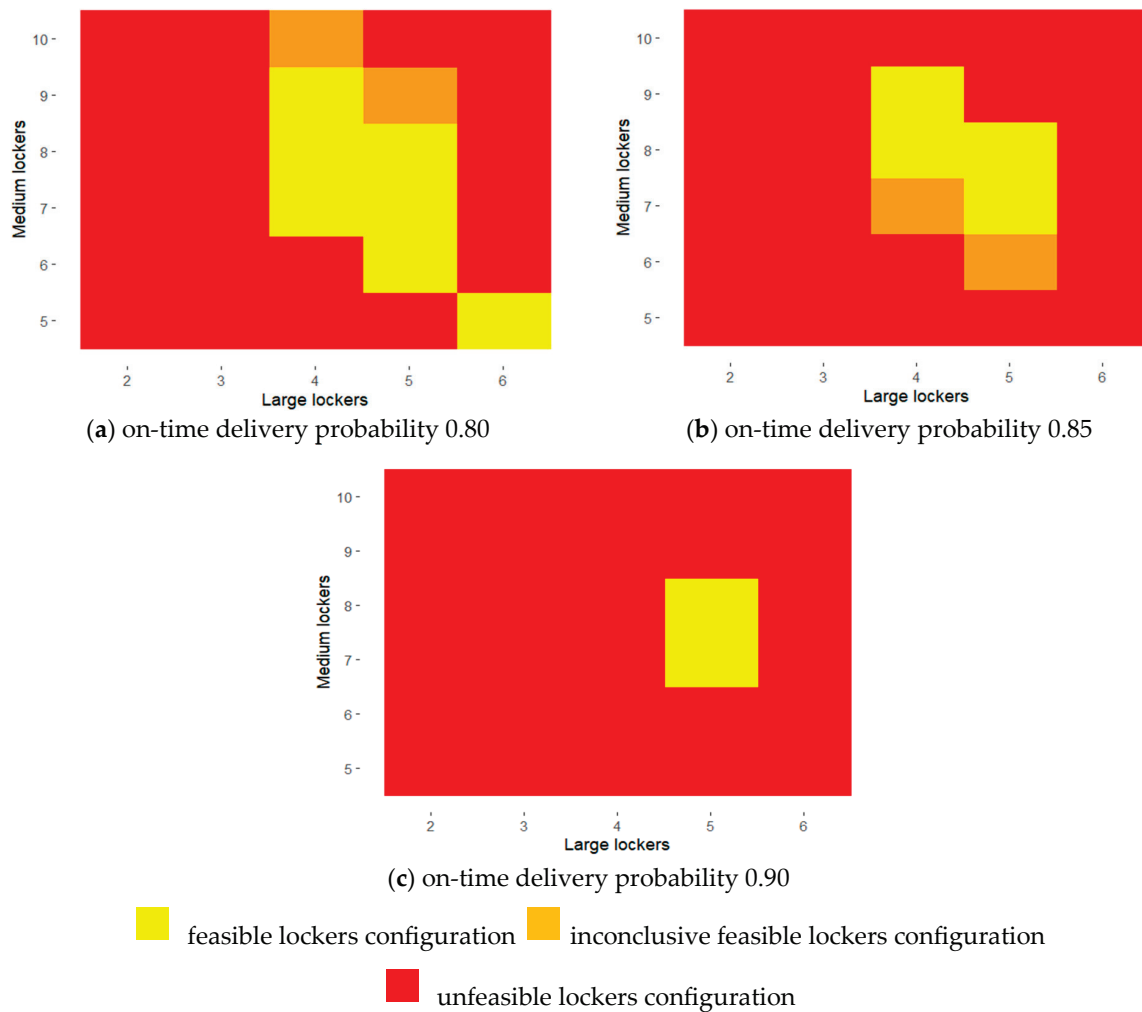


Figure 10. Automated locker configurations for successful delivery (source: authors).

Seven locker configurations satisfy the requirement of $\geq 80\%$ on-time delivery probability (Figure 10a). When the threshold increases to 85%, only four configurations remain feasible (Figure 10b). For the strictest criterion ($\geq 95\%$ on-time delivery), just two configurations meet the target $\langle L_l, L_m, L_s \rangle \in \{ \langle 5, 7, 36 \rangle, \langle 5, 8, 34 \rangle \}$ (Figure 10c). The fast decline in feasible configurations at higher thresholds highlights the trade-off between delivery reliability and system resource requirements.

Undelivered on-time parcels are temporarily stored at the central warehouse until a new delivery can be scheduled, and sufficient locker space becomes available. The end-of-day inventory of undelivered parcels serves as a key performance indicator, as excessive accumulation may lead to storage capacity constraints. To assess the time-series properties of the inventory data, stationarity was evaluated using the Dickey–Fuller Generalized Least Squares (DF-GLS) test [42–44]. The initial inventory series is detrended (Equation (21)), and the detrended series is then tested for a unit root using an augmented Dickey–Fuller-like regression (Equation (22)):

$$\tilde{I}_{j,x} = I_{j,x} - \beta_0 - \beta_1 j \quad (21)$$

$$\Delta \tilde{I}_{j,x} = \alpha \tilde{I}_{j-1,x} + \sum_{i=1}^p \phi_i \Delta \tilde{I}_{j-i,x} + \varepsilon_j \quad (22)$$

where

$I_{j,x}$ is the time series of the end-of-day undelivered x parcels (x is the size of the parcels, i.e., large, medium, small);

$\tilde{I}_{j,x}$ —the detrended time series;

$\Delta \tilde{I}_{j,x} = \tilde{I}_{j,x} - \tilde{I}_{j-1,x}$ —first difference of the detrended series;

α —the key coefficient, β —detrended series coefficients;

ϕ_i —the coefficients of the lags;

ε —the error;

p —the lag order of the autoregressive process.

The DF-GLS test's hypothesis is as follows:

- $H_0 : \alpha = 0$ —unit root exists (non-stationary time series);
- $H_a : \alpha < 0$ —the time series is stationary.

For three APL configurations that are feasible for different on-time delivery probabilities, the evolution of parcel inventories is depicted in Figure 11.

The DF-GLS test was conducted on all-time series configurations, including end-of-day inventory levels for large, medium, and small parcels ($n = 365$ daily observations). The results are presented in Table 4.

Table 4. DF-GLS test statistics.

End-of-Day Inventory	Automated Parcel Lockers Configuration		
	$\langle L_l, L_m, L_s \rangle = \langle 6, 5, 36 \rangle$	$\langle L_l, L_m, L_s \rangle = \langle 5, 6, 38 \rangle$	$\langle L_l, L_m, L_s \rangle = \langle 5, 7, 36 \rangle$
Large parcels	−8.57	−7.53	−7.98
Medium parcels	−5.90	−6.06	−5.57
Small parcels	−6.85	−6.69	−5.97

The DF-GLS is a one-sided (left) test, and the 5% critical value with $n = 365$ observations is -2.89 [43]. For all studied configurations and parcel size categories, the computed test statistics were more negative than the critical value (i.e., test statistic < -2.89). The results provide robust evidence that all inventory time series—regardless of parcel size category—are trend-stationary. This implies that fluctuations in postponed parcel inventory

levels revert to a deterministic trend, and the system does not exhibit random-walk behavior. Ensuring customer satisfaction aligns with the effective management of end-of-day inventories for feasible APLs configurations, thereby addressing the perspectives of both customers and the company.

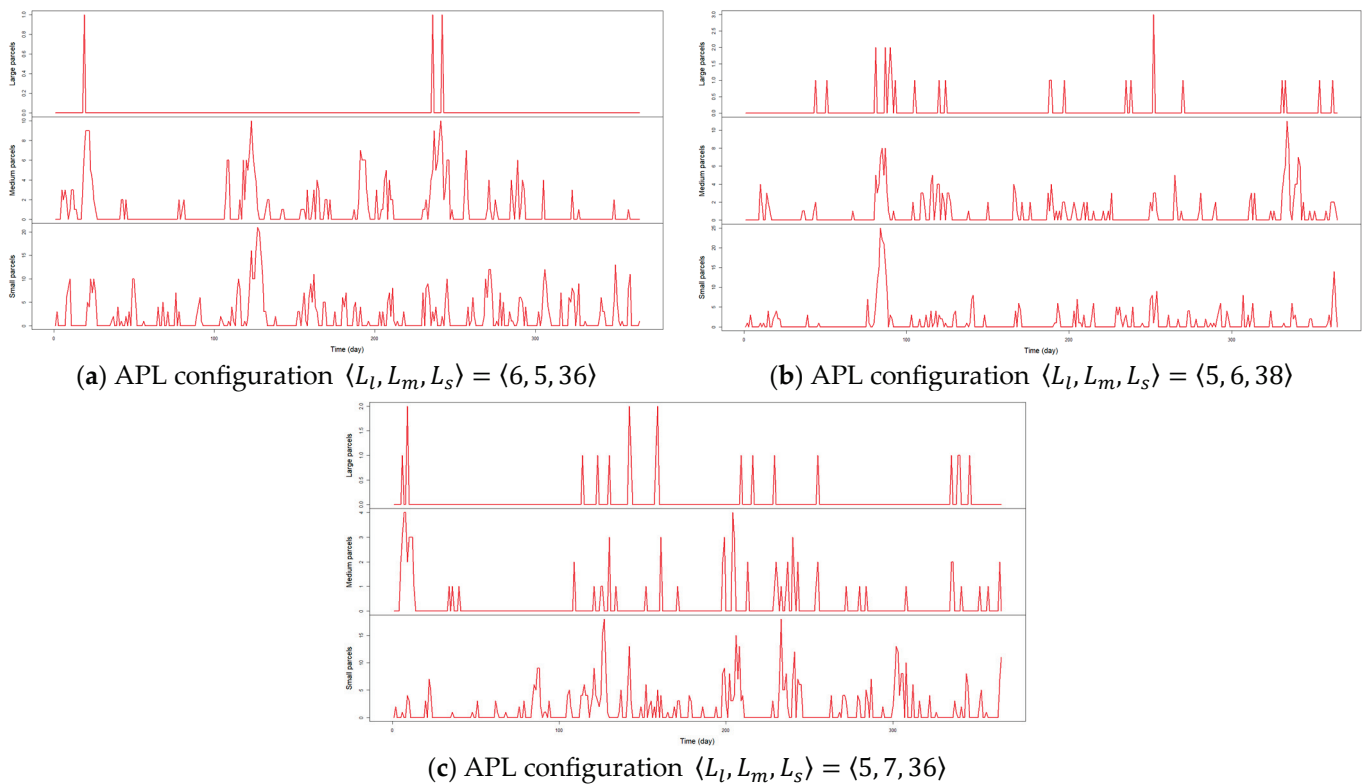


Figure 11. End-of-day undelivered parcel inventories (source: authors).

4. Discussion

The research highlights the relationship between operational performance metrics (including locker utilization rates and fluctuations in undelivered parcel inventory) and customer satisfaction, assessing how optimizing these metrics can enhance service fulfillment and user experience. The ability of larger lockers to flexibly store smaller parcels introduces an essential trade-off between flexibility and efficiency. The results demonstrate that an insufficient number of large and medium lockers can cause performance degradation, particularly for large and medium parcels, whose accommodation options are more limited compared to small parcels. Notably, when the number of large lockers is below four, the on-time delivery probability for large parcels drops below 0.5, indicating a clear service bottleneck. For 5–6 medium lockers the unsuccessful delivery rates for medium parcels are high, unless the number of large lockers increases. The on-time delivery rate for small parcels decreases as the number of large and medium lockers increases, due to the shrink in the number of small lockers (APLs space availability). System bottlenecks for small parcel delivery emerge when the configuration comprises 5–6 large lockers and 9–10 medium lockers.

Locker utilization rates serve as a critical efficiency metric for operational performance, demonstrating size-dependent spatial gradients across locker configurations. For large lockers, utilization follows a central-to-peripheral positive gradient, with minimum rates (0.62–0.65) observed for configurations of 4–5 lockers and peak rates (0.79–0.82) for 2 or 6 lockers (Figure 9a). This pattern arises because fewer large lockers restrain availability for large parcels, whereas higher numbers grant their use for medium/small parcels due

to availability. Furthermore, increasing large locker counts reduce the minimum required small lockers—a consequence of finite system capacity and the migration of smaller parcels to underutilized larger compartments. The medium-sized locker utilization rate exhibits a bidirectional (up–down/left–right) positive gradient (Figure 9b). Minimum utilization rates (0.53–0.58) occur with 9–10 medium lockers and few large lockers (2–3 units), a result of underutilized medium capacity and expanded small locker availability. Conversely, fewer medium lockers yield higher utilization rates (0.76–0.81). Configurations with more medium (9–10) and large lockers (5–6) show an increased medium locker utilization rate as small parcels are displaced to larger units under capacity constraints. The small locker utilization rate also shows a bidirectional positive gradient (bottom–up/left–right), as illustrated in Figure 9c. The lowest utilization rates (0.53–0.61) occur in configurations with 2–3 large lockers and 5–6 medium lockers, reflecting excess capacity in small compartments. As the number of large and medium lockers increases, the available capacity for small lockers decreases, leading to significantly higher utilization rates (0.78–0.80) in systems configured with 5–6 large and 9–10 medium lockers.

While this flexibility helps reduce undelivered inventory in some configurations, it may also create resource competition among parcel types. This was evident from the increased frequency of small and medium parcels being relocated to larger lockers when their designated compartments were insufficient. The balanced proportion of locker types and the frequency of such reallocations correlates with system performance and customers' on-time delivery expectancy. Only seven system configurations meet the $\geq 80\%$ on-time delivery target (Figure 10a). For the configurations $\langle L_l, L_m, L_s \rangle = \langle 4, 10, 34 \rangle$, the 95% confidence interval for large parcel on-time delivery is (0.778, 0.837), while for the configuration $\langle L_l, L_m, L_s \rangle = \langle 5, 9, 32 \rangle$, the 95% confidence interval for small parcel on-time delivery is (0.797, 0.8132). Given these lower marginal performance bounds below the 0.8 threshold, both configurations demonstrate inconclusive feasibility and should be excluded from optimal solution sets. With a $\geq 85\%$ successful delivery target, there are four feasible APL configurations (Figure 10b). For the configurations $\langle L_l, L_m, L_s \rangle = \langle 4, 7, 40 \rangle$, the 95% CI for large parcel on-time delivery is (0.848, 0.895), while for the configuration $\langle L_l, L_m, L_s \rangle = \langle 5, 6, 38 \rangle$, the 95% CI for medium parcel on-time delivery is (0.847, 0.879). These two last configurations demonstrate indeterminate feasibility within the specified performance parameters. The feasible set narrows substantially for on-time delivery thresholds above 90% (Figure 10c). Only two configurations meet the target $\langle L_l, L_m, L_s \rangle = \langle 5, 7, 36 \rangle$ and $\langle L_l, L_m, L_s \rangle = \langle 5, 8, 34 \rangle$. A comparison of Figure 9a or Figure 10a reveals that feasible configurations align with the lower utilization rate observed in large parcels. An optimal balance in the number of large lockers ensures adequate storage capacity for large parcels while also accommodating excess demand from medium and small parcels. Moreover, this configuration maintains sufficient capacity in medium and small lockers, thereby supporting overall system efficiency. The fast decline in feasible configurations at higher thresholds highlights the trade-off between service quality and resource allocation.

Furthermore, the study emphasizes the importance of inventory dynamics as a critical determinant for system stability. The Dickey–Fuller Generalized Least Squares test results confirm the stationarity of the end-of-day undelivered parcel inventories across all tested configurations and parcel sizes. This indicates that despite daily fluctuations, the system tends to revert to a long-term equilibrium, which is desirable in logistics settings. A stationary inventory dynamic ensures that the system remains resilient over time, avoiding excessive warehouse capacity.

Despite the robustness and versatility of the simulation model, several limitations should be acknowledged. The study assumes uniform distribution for daily parcel generation, which may not capture real-world temporal variations such as peak seasons or irregu-

lar delivery intervals. Real-world data could also fit different parcel demand distributions (i.e., normal, Erlang, Poisson). Additionally, customer behavior related to pick up times was modeled using fixed probabilities, although these behaviors may vary significantly across demographics or urban contexts. The model also excludes operational disruptions such as locker maintenance or failed delivery attempts due to locker access issues.

Looking ahead, future research could extend this work by incorporating dynamic delivery routing, adaptive locker allocation policies, and real-time decision-making capabilities. Integrating customer behavior models, e-commerce growth rate, including acceptance of alternative locker redirection and responsiveness to notifications, would enhance realism. Moreover, economic evaluations of locker configuration strategies, incorporating installation and maintenance costs, could complete the service-level findings and offer more comprehensive policy guidance. Finally, the role of APLs in smart-city ecosystems—particularly their synergy with public transit and autonomous delivery—permits further investigation to maximize urban logistics efficiency.

5. Conclusions

The simulation-based study presented in this paper provides valuable insights into the operational behavior of automated parcel lockers under stochastic demand conditions and constrained capacity, emphasizing the interplay between locker configurations, utilization efficiency, and service performance. The paper develops a discrete-event simulation model implemented in ARENA software to optimize the configuration of automated parcel lockers. The study examines a system with three parcel size categories and three corresponding locker types, with larger lockers capable of storing smaller parcels—a design feature that introduces trade-offs between flexibility and space efficiency. The model seeks to meet the probability of on-time parcel delivery by determining the optimal mix of locker sizes, accounting for variable dwell times (time-to-pick-up) and dynamic demand patterns. Key contributions of this research include the following:

- **Optimal locker configuration for service efficiency:** Configuration of locker types significantly influences the system's ability to maintain high customers service levels, as measured by on-time parcel delivery probabilities. The higher the customers' satisfaction target, the smaller the number of feasible configurations. A small number of large lockers (typically ≥ 4 units for testing dataset) is essential to accommodate large parcels and prevent service degradation. While additional large lockers enable flexible storage of smaller parcels, they simultaneously reduce available capacity for medium and small lockers, creating a critical system design compromise.
- **Utilization patterns and spatial efficiency:** While flexibility in locker allocation helps manage excess demand in some configurations, it may also create resource competition among parcel types. Large lockers show a central-to-peripheral utilization gradient, while medium and small lockers display bidirectional utilization gradients but in opposite trends. This nonlinear relationship underscores that optimal APLs configurations must balance absolute locker counts with their size-dependent functional interdependencies.
- **System resilience and delayed parcels inventory stationarity:** Dickey–Fuller GLS tests confirm stationarity in undelivered inventories, indicating inherent system stability despite daily fluctuations, a critical feature for long-term logistics planning.

Policymakers and logistics operators can leverage these findings to design systems that maximize delivery reliability while adapting to evolving e-commerce demands. Further research integrating real-time adaptive policies and economic evaluations will strengthen practical applicability in urban logistics networks.

Author Contributions: Conceptualization, E.R.; methodology, E.R., F.R. and S.B.; software, E.R.; formal analysis, F.C.O., A.I. and S.B.; writing—original draft preparation, E.R., F.R., F.C.O. and A.I.; writing—review and editing, E.R., F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National University of Science and Technology Politehnica Bucharest, grant number 70/11.10.2023.

Data Availability Statement: The data are presented in the paper. More information is available on request from the corresponding author.

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

References

- World Economic Forum. *The Future of the Last-Mile Ecosystem*; World Economic Forum: Geneva, Switzerland, 2021.
- Van Eck, N.J.; Waltman, L. Visualizing Bibliometric Networks. In *Measuring Scholarly Impact: Methods and Practice*; Ding, Y., Rousseau, R., Wolfram, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 285–320.
- Freitas, H.; Andrade, C.; Campos, C. Collection and delivery points in the last mile of cargo transportation: Sustainability benefits and challenges. *Ann. Environ. Sci. Toxicol.* **2023**, *7*, 061–065. [CrossRef]
- Pan, S.; Zhang, L.; Thompson, R.G.; Ghaderi, H. A parcel network flow approach for joint delivery networks using parcel lockers. *Int. J. Prod. Res.* **2020**, *59*, 2090–2115. [CrossRef]
- Orenstein, I.; Raviv, T.; Sadan, E. Flexible parcel delivery to automated parcel lockers: Models, solution methods and analysis. *EURO J. Transp. Logist.* **2019**, *8*, 683–711. [CrossRef]
- Iwan, S.; Kijewska, K.; Lemke, J. Analysis of Parcel Lockers' Efficiency as the Last Mile Delivery Solution—The Results of the Research in Poland. *Transp. Res. Procedia* **2016**, *12*, 644–655. [CrossRef]
- Savelsbergh, M.; Van Woensel, T. 50th Anniversary Invited Article—City Logistics: Challenges and Opportunities. *Transp. Sci.* **2016**, *50*, 579–590. [CrossRef]
- Lachapelle, U.; Burke, M.; Brotherton, A.; Leung, A. Parcel locker systems in a car dominant city: Location, characterisation and potential impacts on city planning and consumer travel access. *J. Transp. Geogr.* **2018**, *71*, 1–14. [CrossRef]
- Yang, X.; Wang, C.; He, X.; Zhang, H.; Xu, G. Location Optimization for Community Smart Parcel Lockers Based on Bilevel Programming. *J. Adv. Transp.* **2023**, *2023*, 1998188. [CrossRef]
- Kurowski, M.; Sobolewski, M.; Koszorek, M. Geometrical Parcel Locker Network Design with Consideration of Users' Preferences as a Solution for Sustainable Last Mile Delivery. *Sustainability* **2023**, *15*, 15114. [CrossRef]
- Grabenschweiger, J.; Doerner, K.F.; Hartl, R.F.; Savelsbergh, M. The vehicle routing problem with heterogeneous locker boxes. *Cent. Eur. J. Oper. Res.* **2021**, *29*, 113–142. [CrossRef]
- Edwards, J.B.; McKinnon, A.C.; Cullinane, S.L. Comparative Analysis of the Carbon Footprints of Conventional and Online Retailing: A “Last Mile” Perspective. *Int. J. Phys. Distrib. Logist. Manag.* **2010**, *40*, 103–123. [CrossRef]
- Deutsch, Y.; Golany, B. A Parcel Locker Network as a Solution to the Logistics Last Mile Problem. *Int. J. Prod. Res.* **2018**, *56*, 251–261. [CrossRef]
- De Oliveira, L.K.; De Oliveira, R.L.M.; De Sousa, L.T.M.; Caliari, I.D.P.; Nascimento, C.D.O.L. Analysis of Accessibility from Collection and Delivery Points: Towards the Sustainability of the E-Commerce Delivery. *Rev. Bras. Gest.* **2019**, *11*, e20190048. [CrossRef]
- Reiffer, A.S.; Kübler, J.; Kagerbauer, M.; Vortisch, P. Agent-based model of last-mile parcel deliveries and travel demand incorporating online shopping behavior. *Res. Transp. Econ.* **2023**, *102*, 101368. [CrossRef]
- Wang, X.; Yuen, K.F.; Wong, Y.D.; Teo, C.C. E-Consumer Adoption of Innovative Last-Mile Logistics Services: A Comparison of Behavioural Models. *Total Qual. Manag. Bus. Excell.* **2018**, *31*, 1381–1407. [CrossRef]
- International Transport Forum. *ITF Transport Outlook 2021*; OECD Publishing: Paris, France, 2021. [CrossRef]
- Lagorio, A.; Pinto, R.; Golini, R. Research in Urban Logistics: A Systematic Literature Review. *Int. J. Phys. Distrib. Logist. Manag.* **2016**, *46*, 908–931. [CrossRef]
- Figliozzi, M.A. Carbon Emissions Reductions in Last Mile and Grocery Deliveries Utilizing Air and Ground Autonomous Vehicles. *Transp. Res. Part D Transp. Environ.* **2020**, *85*, 102443. [CrossRef]
- McKinnon, A.C.; Ge, Y. Use of a Synchronised Vehicle Audit to Determine Opportunities for Improving Transport Efficiency in a Supply Chain. *Int. J. Logist. Res. Appl.* **2004**, *7*, 219–238. [CrossRef]
- European Environment Agency. *Electric Vehicles in Logistics: Lifecycle Assessment*; EEA Report No 15/2022; Publications Office of the European Union: Luxembourg, 2022.

22. Silva, V.; Amaral, A.; Fontes, T. Sustainable Urban Last-Mile Logistics: A Systematic Literature Review. *Sustainability* **2023**, *15*, 2285. [CrossRef]
23. Melkonyan, A.; Gruchmann, T.; Lohmar, F.; Kamath, V.; Spinler, S. Sustainability assessment of last-mile logistics and distribution strategies: The case of local food networks. *Int. J. Prod. Econ.* **2020**, *228*, 107746. [CrossRef]
24. Che, Z.-H.; Chiang, T.-A.; Luo, Y.-J. Multiobjective Optimization for Planning the Service Areas of Smart Parcel Locker Facilities in Logistics Last Mile Delivery. *Mathematics* **2022**, *10*, 422. [CrossRef]
25. Serrano-Hernandez, A.; Martinez-Abad, S.; Ballano, A.; Faulin, J.; Rabe, M.; Chicaiza-Vaca, J. A Hybrid Modeling Approach for Automated Parcel Lockers as a Last-Mile Delivery Scheme: A Case Study in Pamplona (Spain). In Proceedings of the 2021 Winter Simulation Conference (WSC), Phoenix, AZ, USA, 12–15 December 2021. [CrossRef]
26. Sawik, B.; Serrano-Hernandez, A.; Muro, A.; Faulin, J. Multi-Criteria Simulation-Optimization Analysis of Usage of Automated Parcel Lockers: A Practical Approach. *Mathematics* **2022**, *10*, 4423. [CrossRef]
27. Rabe, M.; Chicaiza-Vaca, J.; Tordecilla, R.D.; Juan, A.A. A simulation-optimization approach for locating automated parcel lockers in urban logistics operations. In Proceedings of the 2020 Winter Simulation Conference (WSC), Orlando, FL, USA, 14–18 December 2020.
28. Izco Berastegui, I.; Serrano, A.; Faulin, J.; Sawik, B. A Dynamic Forecast Demand Scenario Analysis to Design an Automated Parcel Lockers Network in Pamplona (Spain) Using a Simulation-Optimization Model. In Proceedings of the 2020 Winter Simulation Conference (WSC), San Antonio, TX, USA, 10–13 December 2023. [CrossRef]
29. Iannaccone, G.; Marcucci, E.; Gatta, V. What Young E-Consumers Want? Forecasting Parcel Lockers Choice in Rome. *Logistics* **2021**, *5*, 57. [CrossRef]
30. Sawik, B. Optimizing Last-Mile Delivery: A Multi-Criteria Approach with Automated Smart Lockers, Capillary Distribution and Crowdshipping. *Logistics* **2024**, *8*, 52. [CrossRef]
31. Lee, H.J.; Lyu, J. Personal values as determinants of intentions to use self-service technology in retailing. *Comput. Hum. Behav.* **2016**, *60*, 322–332. [CrossRef]
32. de Oliveira, L.K.; Morganti, E.; Dablanc, L.; de Oliveira, R.L.M. Analysis of the potential demand of automated delivery stations for e-commerce deliveries in Belo Horizonte, Brazil. *Res. Transp. Econ.* **2017**, *65*, 34–43. [CrossRef]
33. Cieřla, M. Perceived Importance and Quality Attributes of Automated Parcel Locker Services in Urban Areas. *Smart Cities* **2023**, *6*, 2661–2679. [CrossRef]
34. Lai, P.L.; Jang, H.; Fang, M.; Peng, K. Determinants of customer satisfaction with parcel locker services in last-mile logistics. *Asian J. Shipp. Logist.* **2022**, *38*, 25–30. [CrossRef]
35. Adnan, M.; Pereira, F.; Azevedo, C.L.; Basak, K.; Lovric, M.; Raveau, S.; Zhu, Y.; Ferreira, J.; Zegras, C.; Ben-Akiva, M. SimMobility: A Multi-Scale Integrated Agent-Based Simulation Platform. In *Transportation Research Record, Proceedings of the 95th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 10–14 January 2016*; The National Academies of Sciences, Engineering, and Medicine: Washington, DC, USA, 2016; Volume 2.
36. Nikghadam, S.; Vanga, R.; Rezaei, J.; Tavasszy, L. Cooperation between Vessel Service Providers in Ports: An Impact Analysis Using Simulation for the Port of Rotterdam. *Marit. Transp. Res.* **2023**, *4*, 100083. [CrossRef]
37. Liang, K.; Zhou, L.; Yang, J.; Liu, H.; Li, Y.; Jing, F.; Shan, M.; Yang, J. Research on a Dynamic Task Update Assignment Strategy Based on a “Parts to Picker” Picking System. *Mathematics* **2023**, *11*, 1684. [CrossRef]
38. Simchi-Levi, D.; Kaminsky, P.; Simchi-Levi, E. *Designing and Managing the Supply Chain*, 4th ed.; McGraw-Hill: New York, NY, USA, 2021.
39. Vargas-Muñoz, J.C.; Sanchez-Nitola, F.A.; Adarme Jaimes, W.; Rios, R. Enhancing Logistical Performance in a Colombian Citrus Supply Chain Through Joint Decision Making: A Simulation Study. *Logistics* **2025**, *9*, 30. [CrossRef]
40. Rosca, E.; Rusca, F.; Rosca, M.A.; Rusca, A. Performance Analysis of Automated Parcel Lockers in Urban Delivery: Combined Agent-Based–Monte Carlo Simulation Approach. *Logistics* **2024**, *8*, 61. [CrossRef]
41. Wan, J.; Wang, Y. Research on Optimization of E-Commerce Automated Storage Layout Simulation Based on Flexsim. In Proceedings of the 12th International Conference on Transportation and Traffic Engineering (ICTTE 2023), Wuhan, China, 12 February 2024. [CrossRef]
42. Dickey, D.A.; Fuller, W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431. [CrossRef] [PubMed]
43. Elliott, G.; Rothenberg, T.J.; Stock, J.H. Efficient Tests for an Autoregressive Unit Root. *Econometrica* **1996**, *64*, 813–836. [CrossRef]
44. Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer: Berlin/Heidelberg, Germany, 2005.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

UAM Vertiport Network Design Considering Connectivity

Wentao Zhang ¹ and Taesung Hwang ^{2,*}¹ Graduate School of Logistics, Inha University, Incheon 22212, Republic of Korea; wtzhang@inha.edu² Asia Pacific School of Logistics, Inha University, Incheon 22212, Republic of Korea

* Correspondence: thwang@inha.ac.kr; Tel.: +82-32-860-8230

Abstract: Urban Air Mobility (UAM) is envisioned to revolutionize urban transportation by improving traffic efficiency and mitigating surface-level congestion. One of the fundamental challenges in implementing UAM systems lies in the optimal siting of vertiports, which requires a delicate balance among infrastructure construction costs, passenger access costs to their assigned vertiports, and the operational connectivity of the resulting vertiport network. This study develops an integrated mathematical model for vertiport location decision, aiming to minimize total system cost while ensuring UAM network connectivity among the selected vertiport locations. To efficiently solve the problem and improve solution quality, a hybrid genetic algorithm is developed by incorporating a Minimum Spanning Tree (MST)-based connectivity enforcement mechanism, a fundamental concept in graph theory that connects all nodes in a given network with minimal total link cost, enhanced by a greedy initialization strategy. The effectiveness of the proposed algorithm is demonstrated through numerical experiments conducted on both synthetic datasets and the real-world transportation network of New York City. The results show that the proposed hybrid methodology not only yields high-quality solutions but also significantly reduces computational time, enabling faster convergence. Overall, this study provides practical insights for UAM infrastructure planning by emphasizing demand-oriented vertiport siting and inter-vertiport connectivity, thereby contributing to both theoretical development and large-scale implementation in complex urban environments.

Keywords: urban air mobility; vertiport location decision; connected network design; genetic algorithm; minimum spanning tree; greedy initialization

1. Introduction

UAM is set to transform urban transportation by alleviating congestion and improving mobility efficiency through the deployment of electric vertical takeoff and landing aircraft [1]. As urban populations continue to grow and conventional transportation systems become increasingly strained, UAM offers a promising alternative to enhance accessibility and reduce travel times in metropolitan areas. Shon et al. [2] demonstrated that UAM can provide economic and travel time advantages over conventional ground transportation, particularly in large cities experiencing severe congestion.

However, the successful deployment of a UAM network hinges on the optimal placement of vertiports, which serve as critical nodes for takeoff, landing, charging, and passenger transfers [3]. The vertiport location planning is inherently complex, requiring a careful balance among construction costs, passenger accessibility, and network connectivity. Herniczek and German [4] addressed this multifaceted challenge through a scalable optimization model that accounts for passenger demand, infrastructure constraints, and

service delays across major U.S. cities. Recent research has further highlighted the importance of network structure and spatial connectivity in transport planning. Wang and Chen [5] applied a multi-stage network analysis to the China Railway Express to evaluate logistical hub centrality and regional disparities, emphasizing the role of multimodal and resilient connectivity in large-scale infrastructure planning. While previous studies on facility location models have primarily focused on cost minimization or demand coverage maximization, they have largely neglected the importance of ensuring robust connectivity between critical nodes. A well-connected UAM vertiport network is essential for maintaining operational efficiency, reducing unnecessary detours, and enhancing system resilience. Lack of connectivity could lead to inefficient routing, increased operation costs, and decreased service reliability, making it imperative to adopt a connectivity-aware optimization strategy for seamless UAM integration into urban mobility networks.

Despite the increasing interest in UAM vertiport location decisions, several key challenges remain unaddressed. First, balancing cost efficiency and network-wide connectivity is a significant challenge, as existing models often fail to fully incorporate connectivity constraints in facility location problems (FLPs). Second, computational scalability poses a major hurdle, as exact optimization solvers such as CPLEX become impractical for large-scale UAM networks with over 100 candidate locations, necessitating the development of more efficient heuristic-based approaches. Third, many prior studies rely on synthetic demand models rather than real-world urban mobility data, limiting the accuracy and applicability of their findings. Finally, many existing heuristic approaches, including those based on a genetic algorithm (GA), often exhibit limited performance when applied to large-scale transportation networks, particularly under complex connectivity constraints that require feasible and well-structured network solutions. Addressing these challenges is essential for enabling a connected and operationally feasible UAM network that can be effectively deployed in real-world urban contexts.

To address the challenges of the UAM vertiport location problem, this study makes several key contributions. First, a novel connectivity-aware mathematical model is developed to find near-optimal vertiport locations while ensuring network connectivity under predefined distance constraints. Furthermore, a hybrid optimization framework is introduced, which integrates GA with an MST-based connectivity mechanism (i.e., a fundamental concept in graph theory that ensures connectivity with minimal total link cost) to generate connected and cost-effective vertiport configurations. In addition, a greedy initialization (GI) strategy is integrated into the proposed algorithm, enhancing convergence speed and solution quality. The proposed framework is validated through comprehensive numerical experiments on both synthetic fully connected networks (with 25, 50, 100, and 200 candidate vertiports) and a real-world case study in New York City (with 444 candidate locations extracted from public datasets). By incorporating connectivity constraints, scalable heuristics, and real-world case validation, this study contributes to both theoretical advancements and practical applications in large-scale UAM network planning.

The remaining part of this paper is structured as follows. Section 2 reviews the literature on FLP, focusing on connectivity-constrained models and heuristic approaches relevant to UAM. Section 3 presents the mathematical formulation of the UAM vertiport location problem, incorporating connectivity constraints to ensure that the selected network remains connected within the allowable flight range. Section 4 introduces the hybrid GA framework, which incorporates MST-based connectivity verification and a GI strategy to accelerate convergence and improve solution quality. Section 5 presents computational experiments on various synthetic networks and a real-world transportation network in New York City. The results show that the proposed algorithm consistently outperforms

traditional heuristics and CPLEX. Section 6 concludes by confirming the effectiveness and scalability of the proposed methodology and presents future research plans.

2. Literature Review

Conducting a comprehensive review of the relevant literature is essential for contextualizing the proposed connectivity-constrained vertiport location model within the broader domains of facility location theory and transportation network design. This section is structured to reflect the evolution of FLP research, beginning with foundational models and their classical formulations. It then highlights the emerging focus on connectivity-constrained facility siting in multimodal systems, underscoring the limitations of conventional approaches in capturing inter-facility network structures. Subsequently, metaheuristic optimization algorithms commonly employed for solving large-scale FLPs are examined, with particular attention to hybrid methods that balance exploration and exploitation in high-dimensional solution spaces. This review lays the groundwork for identifying a critical research gap, namely, the absence of scalable optimization frameworks that explicitly integrate network connectivity into UAM-focused FLP models.

2.1. The Facility Location Problems

The FLP is a core issue in transportation, logistics, and operations research. It involves determining optimal facility placements to minimize operational costs or maximize service coverage. Classical models such as the p -median, p -center, and hub location models focus primarily on cost minimization and demand satisfaction [6,7]. However, these models often overlook the role of network connectivity, which is essential in systems requiring coordination among multiple facilities. FLP has been applied across diverse domains, including supply chain network design, disaster response logistics, and urban service planning. Solution approaches range from exact algorithms, which offer precise optimization under structural constraints, to metaheuristic methods, including GA, which provide flexibility and scalability in solving large-scale and complex FLP variants [8,9]. Hwang et al. [10] introduced an advanced metaheuristic approach that accounts for traffic congestion and environmental considerations, suggesting that the integration of traffic congestion effects into FLP improves network efficiency and sustainability. Some research has extended facility location models to include transportation dynamics. For example, Melkote and Daskin [11] developed an integrated model that simultaneously addresses facility location and transportation network planning. Location-routing problems (LRPs), reviewed by Nagy and Salhi [12], illustrate the interplay between facility locations and vehicle routing decisions. Drexler and Schneider [13] surveyed extensions to LRP models, focusing on multi-objective formulations incorporating cost, service quality, and environmental performance. In a related domain, Bai et al. [14] investigated a biorefinery supply chain model incorporating traffic congestion, highlighting how transport conditions and route-dependent costs influence facility location decisions iteratively in urban transportation networks. Overall, recent advancements in FLP research reflect a transition from cost-centric models to integrated frameworks that incorporate transportation dynamics, environmental impacts, equity, or demographic responsiveness across diverse application scenarios.

2.2. Connectivity-Constrained Facility Location

While FLPs have been extensively studied, most traditional models do not explicitly account for inter-facility connectivity, often assuming that facilities operate independently. However, in applications such as multimodal transportation, telecommunications, and UAM, maintaining robust connectivity between facilities is essential for ensuring operational resilience and efficiency. Melkote and Daskin [11] introduced one of the earliest

integrated models that jointly optimize facility location and transportation network design, showing that restructuring network topology can reduce costs more effectively than adding new facilities. To explicitly address connectivity, researchers have proposed robust facility network design models that ensure inter-facility linkage under uncertainty. For example, Bardossy and Raghavan [15] developed a robust optimization model for the connected FLP, using a heuristic lower bound combination to efficiently generate high-quality solutions while maintaining network connectivity in uncertain environments.

Among the widely studied principles in transportation and infrastructure planning, enforcing network connectivity remains a fundamental modeling requirement for ensuring cost-effective and operationally feasible system design. For instance, Gewali et al. [16] addressed spatially constrained facility-linked networks by introducing connectivity-preserving formulations applicable in geometrically restricted environments. Similarly, Peng et al. [17] formulated a discrete group assembly problem with connectivity constraints, demonstrating that ensuring a spanning tree structure with bounded edge lengths can effectively minimize total travel distance and enhance network feasibility. These studies highlight the critical role of connectivity constraints in enabling efficient and realistic network optimization. Sun et al. [18] proposed a robust network optimization framework that incorporates network connectivity as an explicit design consideration to enhance the resilience of supply chain systems under uncertainty, demonstrating improved solution quality and reduced sensitivity to disruptions. The importance of such integrated approaches has been broadly recognized in the facility location literature, as highlighted by Saldanha-da-Gama [19]. Martens and Bley [20] proved such problems to be NP-hard and proposed approximation algorithms ensuring basic network connectivity. Srivastava and Salapaka [21] developed a joint model for facility and path optimization in both static and dynamic networks using entropy and control-based methods. Shishebori and Ghaderi [22] addressed link disruptions through a robust optimization model solved by a hybrid meta-heuristic. Güner and Sevkli [23] applied discrete Particle Swarm Optimization (PSO) to facility location, showing strong performance against classical heuristics.

Despite notable progress in facility location modeling, connectivity is still frequently overlooked as a core structural constraint. This gap is particularly pronounced in UAM applications, where reliable inter-vertiport connectivity is essential for maintaining system coordination and operational feasibility. Existing models tend to prioritize reducing cost or maximizing service coverage, yet often lack explicit mechanisms to ensure a connected network structure. Bridging this gap requires integrated modeling approaches that jointly consider vertiport siting and connectivity constraints, enabling the design of coherent and resilient UAM networks.

2.3. Optimization Algorithms for Facility Location Problems

FLPs are well-known NP-hard combinatorial optimization problems, which arise in various domains such as logistics, supply chain design, and urban infrastructure planning. Due to their computational complexity, exact methods are often feasible for small-scale instances. As a result, metaheuristic algorithms have become the dominant approaches for tackling FLPs, for example, GAs, Simulated Annealing (SA), or PSO.

PSO and GAs remain popular approaches for solving uncapacitated and capacitated FLPs. Hybrid approaches that combine PSO with GA mechanisms have shown enhanced solution robustness and faster convergence compared to standard GA methods [24]. For instance, discrete PSO with embedded local search has shown competitive results compared to evolutionary SA and GAs in solving benchmark FLP datasets [25]. SA remains a useful standalone method, especially when computational time is a critical constraint. In nuclear engineering applications, a multistart SA achieved over 98%-time savings compared to

PSO, while maintaining competitive solution quality [26]. Furthermore, recent research emphasizes hybrid metaheuristics that combine the global exploration of PSO or GAs with the local intensification of SA. For instance, Yalaoui et al. [27] developed a hybrid GA–PSO method for solving facility layout problems, demonstrating improved performance in both solution quality and computational efficiency. Similarly, Samuel and Rajan [28] applied a GA–PSO hybrid framework to large-scale scheduling problems, reporting enhanced robustness and reduced convergence time compared to standalone algorithms. These studies highlight the potential of hybrid metaheuristics in balancing global search capability with local refinement for complex facility network design tasks.

Although significant advances have been made, most existing models still fail to inherently ensure network connectivity among facilities, particularly within the context of UAM or multimodal transportation infrastructure. To address this research gap, this study proposes an MST-GA hybrid approach that explicitly integrates an MST-based connectivity framework into the GA. Furthermore, a problem-specific GI strategy is incorporated to enhance solution feasibility and guide the search process toward structurally connected, high-quality solutions from the outset. This study offers an effective and scalable methodology for complex, real-world FLPs.

2.4. Research Contribution

While facility location optimization has been extensively studied, connectivity—a critical requirement for feasible and efficient UAM network design—is often overlooked. This study introduces a connectivity-constrained vertiport location model specifically tailored for UAM applications. The proposed model simultaneously achieves cost-effective vertiport siting and robust network connectivity, both of which are essential for ensuring the operational feasibility of UAM systems. Unlike traditional models, this framework incorporates explicit connectivity to maintain inter-vertiport links, improving system resilience. To solve this model effectively, we introduce a hybrid optimization algorithm combining GI with MST-based connectivity. This approach enhances convergence stability and ensures feasible, high-quality solutions. Extensive numerical experiments and a real-world case study have been conducted to validate the performance of the proposed algorithm. The results highlight that incorporating connectivity constraints into the FLPs substantially enhances infrastructure planning effectiveness in UAM networks. By embedding connectivity into the optimization process, the proposed model facilitates the development of spatially coherent and well-integrated vertiport layouts. This work lays a solid methodological foundation for future research and large-scale implementation of connectivity-aware vertiport location strategies in UAM systems.

3. Model Formulation

UAM has emerged as a promising solution for alleviating urban congestion and enhancing transportation efficiency. The successful deployment of UAM services hinges on the strategic siting of vertiports, which serve as critical nodes for vertical takeoff and landing operations. Unlike traditional FLPs, UAM vertiport planning must explicitly incorporate connectivity considerations by requiring that each selected vertiport is directly connected to at least one other selected vertiport, thereby ensuring the formation of a connected operational network. To formulate the UAM vertiport location problem, the system is modeled as an undirected graph $G(V, E)$, where V denotes the set of nodes and E represents the set of links between these nodes. The candidate vertiport locations and the demand points are defined, respectively, as a subset $J' \subseteq V$ and $I \subseteq V$, and the demand points $i \in I$ must be assigned to one of the selected facilities $j \in J'$. The distance between

any two nodes $i \subseteq V$ and $j \subseteq V$ is denoted by d_{ij} , which represents the Euclidean distance between nodes i and j in a given network.

Due to battery range limitations and airspace regulations, the flight distance between any two selected vertiports must not exceed the maximum allowable range D_{max} , thereby ensuring operational feasibility. In real-world UAM operations, limited battery range makes inter-vertiport connectivity essential, as inadequate links can cause infeasible routes, detours, or mission failure, highlighting the practical need for range-aware network design and vertiport placement [29,30]. As a result, the selected vertiports must form a connected subgraph while minimizing total operational costs. To provide a clearer understanding of the network connectivity constraints in the proposed model, a conceptual diagram (Figure 1) adapted from Peng et al. [17] is presented to illustrate the virtual source, sink nodes, selected vertiports, and corresponding connectivity relationships.

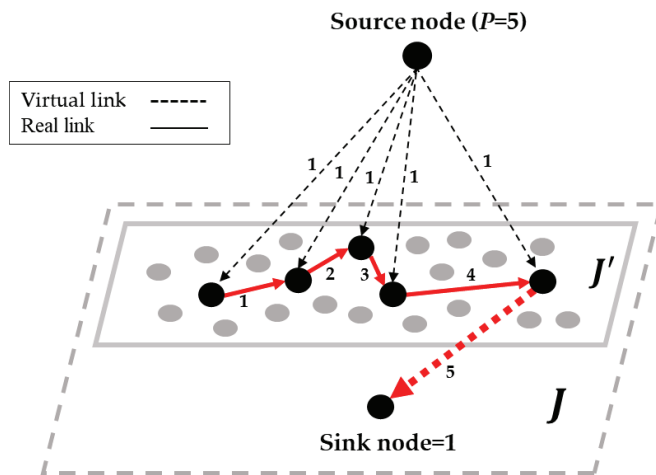


Figure 1. Conceptual illustration of network connectivity constraint.

The concept of connectivity described in Peng et al. [17] is adapted in this study, where, if the shortest distance between two nodes in a given network is less than a certain value (i.e., D_{max} in this study), they are considered to be connected to each other. In order to construct such a connected UAM vertiport network, we try to find the optimal locations of vertiports through a proposed model that maintains connectivity between the selected nodes through virtual flows as described in Peng et al. [17]. For a more detailed explanation, a virtual network is presented in Figure 1, and the candidate sites of vertiports are indicated by both black and gray circles on the plane in the figure. The black circles represent the nodes finally selected as the UAM vertiports among the facility candidate sites. In addition, there is a virtual source node at the top of the plane and a virtual sink node (represented by node number 1) at the bottom of the plane. Set J includes both the set of candidate vertiport locations J' and the virtual sink node $\{1\}$. The number of facilities to be located is indicated by P , and it is assumed to be 5 in this figure.

To determine the final locations, the source node sends a flow of positive integer 1 to each node among the candidate sites on the plane where the vertiport is to be located (i.e., total $P = 5$, and the flow will be distributed on the selected nodes, each of which receives unit flow). The node selected as the vertiport sends the flow of size 1 received from the source node and any flows received from nearby vertiport nodes within a certain distance (i.e., D_{max} in this study). Finally, the total amount of flows received from both the virtual source node and the previously connected nodes at the last selected location is sent to the virtual sink node, which is introduced to serve as an artificial aggregation point. As such, if the flow conservation is satisfied in all selected nodes and the total flow flowing into the virtual sink node is guaranteed to be the same as

the initial total out-flow from the source node, a feasible solution can be found, where the selected vertiports form a spanning tree structure. Let $L \subseteq E$ represent the subset of edges (i.e., flight routes) connecting the chosen vertiport locations. The connectivity constraints can be formulated as follows:

$$d_{jk} \leq D_{max}, \forall (j, k) \in L \quad (1)$$

Formula (1) ensures that any two connected vertiports are within the maximum allowable flight distance D_{max} , thereby preventing infeasible or excessively long inter-vertiport routes. To further guarantee network connectivity while eliminating redundant links, the following formula can be constructed:

$$|L| \leq |P| - 1 \quad (2)$$

As noted by Peng et al. [17], Formula (2) ensures that the number of edges does not exceed the number of selected nodes minus one, effectively avoiding cycle formations and supporting the construction of a sparse yet connected and cost-efficient network topology. The incorporation of these concepts ensures that the selected vertiports form a spatially cohesive and operationally viable network, thereby enhancing both the resilience and efficiency of the UAM transportation system.

Each demand node $i \in I$, representing the passenger origin point, is associated with a demand quantity h_i . Due to the absence of reliable origin–destination (OD) flow data in current UAM scenarios, demand is approximated in a node-based form. Nonetheless, incorporating OD-based demand (e.g., h_{ij}) in future studies may yield more accurate modeling outcomes. The optimization task involves selecting exactly P vertiports from a subset $J' \subseteq J$ while ensuring network connectivity and minimizing total operational costs. For each demand node $i \in I$ and vertiport $j \in J'$, the Euclidean distance is denoted by d_{ij} . Locating a vertiport at location $j \in J'$ incurs a fixed cost f_j which includes both land acquisition and infrastructure development expenses. Additionally, the transportation cost is modeled as a function of distance, where α denotes the unit transportation cost per person per kilometer for traveling from his/her demand point to the assigned vertiport. The allocation of demand to vertiports is governed by the binary decision variable Y_{ij} , where $Y_{ij} = 1$ indicates that the demand node i is assigned to vertiport j , and $Y_{ij} = 0$ otherwise. Similarly, the binary decision variable X_j determines whether a vertiport is established at location j , where $X_j = 1$ if a vertiport is constructed and $X_j = 0$ otherwise. To ensure operational feasibility, the selected vertiports must satisfy network connectivity constraints and form a connected subgraph. A connectivity set N_j is defined for each vertiport $j \in J'$, containing all directly connected candidate vertiports within the maximum allowable flight range D_{max} . The connectivity between any two selected vertiports $i, j \in J'$ is represented by the variable l_{ij} , which denotes the size of directional flow from node i to node j on the plane in Figure 1.

By integrating these components, the problem is formulated as a vertiport location decision model that minimizes the total cost, which includes passenger transportation costs incurred by traveling from demand points to assigned vertiports and vertiport construction costs at the selected nodes, and ensures connectivity among the selected UAM vertiports. This formulation provides a robust foundation for UAM infrastructure planning, ensuring both cost-effectiveness and operational feasibility in large-scale urban environments. Based on the defined parameters and decision variables, the complete model is expressed as follows:

Minimize

$$\alpha \sum_{i \in I} \sum_{j \in J'} h_i d_{ij} Y_{ij} + \sum_{j \in J'} f_j X_j \quad (3)$$

s. t.

$$\sum_{j \in J'} X_j = P \quad (4)$$

$$\sum_{j \in J'} Y_{ij} = 1, \forall i \in I \quad (5)$$

$$Y_{ij} - X_j \leq 0, \forall i \in I, j \in J' \quad (6)$$

$$X_j + \sum_{n \in N_j} l_{nj} = \sum_{m \in N_j} l_{jm}, \forall j \in J' \quad (7)$$

$$l_{ij} \geq 0, \forall i, j \in J \quad (8)$$

$$l_{ij} \leq (P - 1)X_j, \forall i, j \in J' \quad (9)$$

$$l_{1j} = 0, \forall j \in J' \quad (10)$$

$$l_{j1} \in \{0, P\}, \forall j \in J' \quad (11)$$

$$X_1 = 0 \quad (12)$$

$$J' \cup \{1\} = J \quad (13)$$

$$N_j = \{n \in J' \mid d_{jn} \leq D_{max}\}, \forall j \in J' \quad (14)$$

$$X_j \in \{0, 1\}, \forall j \in J' \quad (15)$$

$$Y_{ij} \in \{0, 1\}, \forall i \in I, j \in J' \quad (16)$$

The objective function (3) minimizes the total cost, comprising both passenger transportation costs and vertiport construction expenditures. The first term quantifies the cumulative access cost incurred by all passengers commuting from their origins to assigned vertiports, calculated as the product of unit transportation cost, demand, and travel distance. The second term represents the fixed infrastructure costs associated with constructing vertiports at the selected locations. Constraint (4) ensures that exactly P locations are selected for vertiports, thereby enforcing a predefined number of operational sites. The customer allocation constraints (5) guarantee that each demand node is assigned to exactly one vertiport, ensuring complete service coverage. Additionally, the assignment constraints (6) ensure that demand nodes can only be assigned to vertiports that have been selected, preventing allocations to non-existent infrastructure. The flow balance constraints (7) ensure that the total flow into a vertiport j equals the total flow out, maintaining flow conservation across the network. This guarantees that the connectivity between vertiports is preserved, and no flow is lost or generated unexpectedly. The network structure constraints (8) and (9) limit the number of connections between vertiports. Constraints (8) ensure non-negative virtual flow between vertiports, while constraints (9) restrict each vertiport to having at most $P - 1$ virtual inflows in Figure 1. The end-point flow constraints (10) to (12) regulate the sink node. Constraints (10) ensure no incoming flow from the sink node, while constraints (11) restrict flow to the sink node to be either 0 or the total number of vertiports to be constructed. Constraint (12) ensures the sink node is not selected as a vertiport location. The feasible neighborhood constraint (13) includes the sink node $\{1\}$ in the network but clarifies that it is excluded from a real candidate location

set J' . The neighborhood feasibility constraints (14) enforce distance limitations, ensuring that vertiports can only be connected if they are within the maximum operational distance D_{max} , reflecting maximum UAM travel distance. Finally, constraints (15) and (16) represent binary variables associated with vertiport location decisions X_j and demand allocation decisions Y_{ij} , respectively. This mathematical framework offers a rigorous and scalable approach to optimizing UAM vertiport placement, ensuring that the resulting network remains cost-efficient, connected, and thus operationally feasible. By jointly considering vertiport locations, demand allocation, and connectivity constraints, the model provides a systematic methodology for strategic UAM network planning, supporting long-term adaptability in urban mobility infrastructure development.

4. Proposed Algorithm

This section develops an algorithm to solve the proposed UAM vertiport location planning problem. The framework of the methodology in this section consists of three interlinked components: (1) application of the GA to find the near-optimal vertiport locations, (2) maintaining feasibility of the candidate solution set via a hybrid MST-GA framework, and (3) enhancing computational efficiency through a demand-aware GI strategy. This integrated approach is particularly well-suited for the UAM transportation network, where achieving connectivity and minimizing total system cost are critical within large-scale, high-complexity operational environments.

4.1. Hybrid MST-GA Framework for Connectivity-Constrained Problem

The GA, initially proposed by Holland [31], is a population-based metaheuristic inspired by the mechanisms of natural selection and evolution. Due to its capacity for global search and adaptability, the GA has proven effective for solving large-scale combinatorial optimization problems that are otherwise intractable by classical methods. In particular, GA-based approaches have been widely applied in facility location and hub network design problems within transportation systems, including air logistics, where minimizing system-wide cost and improving spatial efficiency are critical [32]. In this study, the GA is employed to find near-optimal locations of P vertiports while minimizing total system cost, which includes facility construction costs and demand-weighted transportation costs.

A feasible UAM network must satisfy two essential conditions: (i) connectivity, ensuring that all selected facilities form a connected subgraph, and (ii) flight distance constraints, restricting direct connections to a maximum allowable flight range D_{max} . To enforce these conditions, an MST is constructed over the selected facility nodes in the GA procedure; the MST-GA framework integrates GA with the MST concept to ensure all selected facilities form a connected subgraph. The workflow consists of the following steps: an initial population of candidate facility configurations is generated using GA, followed by MST construction to enforce connectivity constraints among selected facilities. An MST is computed over the selected facilities using Prim's algorithm [33], based on the pairwise Euclidean distances between them. This MST is used solely to validate the connectivity of the selected vertiports. A solution is accepted only if the resulting spanning tree satisfies the connectivity condition and all inter-vertiport links fall within the maximum allowable flight distance D_{max} as defined in Formula (1). If any edge in the constructed MST violates this constraint, the corresponding solution is deemed infeasible and excluded from the population during the GA procedure. This post-processing step ensures that the final configuration supports a connected and operationally viable UAM network without altering the evolutionary search dynamics. The detailed procedure for constructing the MST is illustrated in Algorithm 1.

Algorithm 1. Prim’s Algorithm for MST Construction among Selected Facilities.

Input: Candidate facility coordinates $\{(x_j, y_j) \mid j \in J'\}$

Output: Minimum Spanning Tree T

1: Compute pairwise distance matrix:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \forall i, j \in J', i \neq j$$

2: Construct undirected graph $G = (V, E)$ where $V = J'$, and edge weights $w_{ij} = d_{ij}$

3: Initialize MST $T = \emptyset$, select an arbitrary start node $v_0 \in V$ and let visited node set $S = \{v_0\}$

4: **while** $S \neq V$ **do**

5: Find edge $(u, v) \in E$ with minimum w_{uv} such that $u \in S, v \in V \setminus S$

6: Add (u, v) to T

7: Include v in S : $S \leftarrow S \cup \{v\}$

8: **end while**

9: **return** T

The overall procedure of the MST-GA is presented in Algorithm 2, consisting of initialization, fitness evaluation, tournament selection, crossover, mutation, and termination. This iterative process is repeated until a stopping criterion is met, such as a maximum number of generations or a plateau in fitness improvement. Through this evolutionary mechanism, the MST-GA iteratively refines feasible configurations toward an optimal set of facility locations under the predefined costs and demand conditions. To guarantee that the selected facilities form a connected network and a candidate solution is feasible, an MST-based connectivity check is applied in each generation.

Algorithm 2. Hybrid MST-GA Framework for Facility Location Decision.

Input: Candidate location set J' , demand set I , number of facilities to be built P

Output: Near-optimal facility selection vector $\mathbf{X}^*$ or “Not feasible”

1: Initialize population $\mathcal{P} = \{\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^M\}$ such that $\sum_{j \in J'} X_j^m = P$,
 $\forall m \in \{1, 2, \dots, M\}$

2: **while** stopping criterion not met **do**

3: **for all** $\mathbf{X}^m \in \mathcal{P}$ **do**

4: Evaluate fitness: $f(\mathbf{X}^m) = \sum_{i \in I} \sum_{j \in J'} \alpha h_i d_{ij} Y_{ij} + \sum_{j \in J'} f_j X_j^m$

5: **end for**

6: Select parents from $\mathcal{P}$ via tournament selection $\rightarrow \mathcal{P}_{\text{selected}}$

7: $\mathcal{P}' \leftarrow \emptyset$

8: **for all** pairs $(\mathbf{X}^{p_1}, \mathbf{X}^{p_2}) \in \mathcal{P}_{\text{selected}}$ **do**

9: $\mathbf{X}^{\text{off}} \leftarrow \text{crossover}(\mathbf{X}^{p_1}, \mathbf{X}^{p_2})$

10: $\mathbf{X}^{\text{new}} \leftarrow \text{mutation}(\mathbf{X}^{\text{off}}, p_m)$

11: **if** $\sum_{j \in J'} X_j^{\text{new}} \neq P$ **then**

12: Repair $\mathbf{X}^{\text{new}}$ to ensure $\sum X_j = P$

13: **end if**

14: $S \leftarrow \{j \in J' \mid X_j^{\text{new}} = 1\}$

15: $T \leftarrow \text{Prim}(S)$

16: **if** $|T| = P - 1$ **then**

17: Add $\mathbf{X}^{\text{new}}$ to next generation $\mathcal{P}'$

18: **end if**

19: **end for**

Algorithm 2. *Cont.*

```

20:  if  $\mathcal{P}' = \emptyset$  then
21:    return "Not feasible: No connected solution found"
22:  end if
23:  Replace population:  $\mathcal{P} \leftarrow \mathcal{P}'$ 
24: end while
25: return  $\mathbf{X}^* = \operatorname{argmin}_{\mathbf{X} \in \mathcal{P}} f(\mathbf{X})$ 

```

From a computational complexity standpoint, the MST-GA framework is composed of two principal modules: (i) fitness evaluation based on demand-weighted transportation costs and (ii) connectivity verification via MST construction. The total per-generation complexity of this framework can be formally expressed as $O(M \cdot P \cdot |I| + M \cdot P^2)$, where M denotes the population size, P is the number of facilities to be constructed, and $|I|$ represents the number of demand points. The first term, $O(M \cdot P \cdot |I|)$, corresponds to the fitness evaluation process for all individuals in the population. Each candidate solution must compute the total transportation cost by summing the weighted distances from each demand point to its assigned facility, thus quantifying the overall service efficiency. The second term, $O(M \cdot P^2)$, captures the computational burden introduced by the connectivity constraint. Specifically, for each candidate solution, an MST is constructed over the selected facility nodes using Prim's algorithm to ensure that the selected network remains connected. Since this procedure requires the evaluation of all pairwise edge weights and incremental MST construction, it incurs a time complexity of $O(P^2)$ per individual, repeated across all M individuals in each generation. Although the MST verification introduces a quadratic overhead with respect to the number of selected facilities, the overall computational complexity remains tractable for moderate-scale instances. Moreover, both the fitness evaluation and MST construction are inherently parallelizable, making the framework well-suited for modern multi-core computing architectures. Consequently, the MST-GA demonstrates excellent scalability and computational efficiency in solving large-scale UAM vertiport location problems, where both spatial feasibility and network connectivity must be simultaneously addressed. By leveraging evolutionary operators and iterative refinement, the algorithm reliably converges toward facility configurations that are cost-effective, robust, and structurally connected.

4.2. MST-GA Combined with GI

One limitation of standard GA implementations is the reliance on random initialization, which often leads to poor-quality initial solutions, slower convergence, and increased computational effort. To overcome this issue, a GI strategy is introduced to generate high-quality initial solutions based on demand and cost considerations, thereby improving the overall performance of the GA. The greedy selection rule defines S' as the set of already selected facilities and $\tilde{J}$ as the set of remaining candidate facilities. The next facility k^* to be selected is determined iteratively by minimizing the incremental cost, as defined in Formula (17):

$$k^* = \operatorname{argmin}_{k \in \tilde{J}} \sum_{i \in I} h_i d_{ik} + f_k \quad (17)$$

where h_i denotes the size of the demand at node $i \in I$, d_{ik} represents the distance from demand point $i \in I$ to the facility at node $k \in \tilde{J}$, and f_k corresponds to the facility construction cost at node $k \in \tilde{J}$. The GI algorithm operates as follows: (1) the first facility k_0 is selected based on maximum demand coverage; (2) subsequent facilities are iteratively added by choosing k^* that minimizes the cost function; and (3) the process is terminated

when P facilities have been selected. With a computational complexity of $O(P|J'||I|)$, GI significantly accelerates convergence compared to purely random initialization. By pre-selecting high-quality solutions, the search space is effectively reduced, allowing the GA to focus on refining near-optimal configurations rather than exploring unpromising regions and significantly enhancing the convergence behavior of the MST-GA framework. The detailed steps of the GI procedure are outlined in Algorithm 3, where facilities are iteratively selected to minimize incremental cost, considering both the demand-weighted distance and facility construction cost.

Algorithm 3. GI for Facility Selection.

Input: Demand set I , candidate facility set J' , number of facilities P , demand h_i , construction cost f_j , distance d_{ij}

Output: Initial facility subset $S' \subseteq J'$ with $|S'| = P$

1: Initialize $S' \leftarrow \emptyset$, $\tilde{J} \leftarrow J'$

2: Select first facility:

$$k_0 = \underset{j \in \tilde{J}}{\operatorname{argmax}} \left(\sum_{i \in I} \mathbb{I} \left[j = \underset{j' \in \tilde{J}}{\operatorname{argmin}} d_{ij'} \right] \cdot h_i \right)$$

3: Add k_0 to S' ; update $\tilde{J} \leftarrow \tilde{J} \setminus \{k_0\}$

4: **while** $|S'| < P$ **do**

5: **for all** $k \in \tilde{J}$ **do**

$$C_k = \sum_{i \in I} h_i \cdot \min(d_{ij}, d_{ik}) + f_k, \forall j \in S'$$

6: **end for**

7: $k^* = \underset{k \in \tilde{J}}{\operatorname{argmin}} C_k$

8: $S' \leftarrow S' \cup \{k^*\}$, $\tilde{J} \leftarrow \tilde{J} \setminus \{k^*\}$

9: **end while**

10: **return** S'

Building on the existing MST-GA framework, we enhance solution quality and convergence speed by incorporating a GI strategy that generates high-quality initial populations and guides the search process more effectively. The proposed methodology begins with a mixed initialization strategy, combining GI for cost-effective facility selection with conventional random initialization to maintain diversity in the solution pool. Once initialized, the GA iteratively applies selection, crossover, and mutation operators to evolve solutions toward cost minimization. A solution is considered feasible if a spanning tree can be formed among the selected vertiports using only links that do not exceed the maximum flight distance D_{max} . Any solution failing to meet these criteria is considered infeasible and discarded. This evolutionary process continues until a convergence criterion is satisfied, such as a maximum number of generations or a stabilization in fitness values. The complete procedure is summarized in Algorithm 4.

Algorithm 4. The MST-GA combined with GI Approach.

Input: Demand set I , facility set J' , distance matrix $D = [d_{ij}]$, demand vector $h = [h_i]$, construction cost vector $f = [f_j]$, population size M , number of facilities P , mutation rate p_m , maximum generation G

Output: Near-optimal facility selection vector $\mathbf{X}^*$ or “Not feasible”

1: Initialize population $\mathcal{P} = \{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(M)}\}$ where

$$\mathcal{P} = \text{GreedyInit}(M_g) \cup \text{RandomInit}(M - M_g)$$

Algorithm 4. *Cont.*

```

2:  $g \leftarrow 0$ 
3: while  $g < G$  do
4:   for all  $\mathbf{X}^m \in \mathcal{P}$  do
5:      $\mathbf{Y}^{(m)} \leftarrow \text{Assign}(\mathbf{X}^{(m)})$ 
6:     Evaluate fitness:

$$f(\mathbf{X}^{(m)}) = \sum_{i \in I} \sum_{j \in J'} \alpha h_i d_{ij} Y_{ij}^{(m)} + \sum_{j \in J'} f_j X_j^{(m)}$$

7:   end for
8:    $\mathcal{P}_{\text{sel}} \leftarrow \text{Select}(\mathcal{P})$ 
9:    $\mathcal{P}' \leftarrow \emptyset$ 
10:  for all pairs  $(\mathbf{X}^{(p_1)}, \mathbf{X}^{(p_2)}) \in \mathcal{P}_{\text{sel}}$  do
11:     $\mathbf{X}^{\text{cross}} \leftarrow \text{Crossover}(\mathbf{X}^{(p_1)}, \mathbf{X}^{(p_2)})$ 
12:     $\mathbf{X}^{\text{mut}} \leftarrow \text{Mutate}(\mathbf{X}^{\text{cross}}, p_m)$ 
13:    if  $\sum_{j \in J'} X_j^{\text{mut}} \neq P$  then
14:       $\mathbf{X}^{\text{mut}} \leftarrow \text{RepairCount}(\mathbf{X}^{\text{mut}}, P)$ 
15:    end if
16:     $S \leftarrow \{j \in J' \mid X_j^{\text{mut}} = 1\}$ 
17:     $T \leftarrow \text{Prim}(S)$ 
18:    if  $|T| = P - 1$  then
19:       $P' \leftarrow P' \cup \{\mathbf{X}^{\text{mut}}\}$ 
20:    else
21:      Continue
22:    end if
23:  end for
24:  if  $P' = \emptyset$  then
25:    return “Not feasible: No connected solution found”
26:  end if
27:   $\mathcal{P} \leftarrow \text{Update}(\mathcal{P}, P')$ 
28:   $g \leftarrow g + 1$ 
29: end while
30: return  $\mathbf{X}^* = \text{argmin}_{\mathbf{X} \in \mathcal{P}} f(\mathbf{X})$ 

```

The proposed approach combines GI-based high-quality initial populations, GA-driven iterative facility location optimization, and MST-based connectivity validation to form an integrated optimization pipeline. By jointly addressing cost efficiency, spatial feasibility, and solution quality, the model provides a robust and scalable solution for UAM vertiport placement in large-scale urban environments, ensuring connectivity. The total computational complexity of the framework can be expressed as $O(M_g \cdot P \cdot |J'| \cdot |I|) + O(G \cdot M \cdot (P \cdot |I| + P^2))$, where M_g is the number of greedy individuals, G is the number of generations, M is the population size, P is the number of selected vertiports, $|J'|$ is the number of candidate locations, and $|I|$ is the number of demand points.

5. Numerical Examples

To evaluate the solution quality and computational efficiency of the proposed algorithm, we conducted a series of numerical experiments on both synthetic and real-world datasets. The synthetic dataset consists of fully connected networks with various network sizes. These test instances were designed to assess the scalability and robustness

of the proposed algorithm under different problem sizes. In addition, a real-world case study based on New York City was conducted, comprising 444 candidate vertiport locations obtained from publicly available GIS data. In this section, the results from the proposed algorithm (i.e., MST-GA combined with GI) were compared to those computed from MST-GA and benchmark solutions obtained from CPLEX (Version 12.10), a state-of-the-art commercial optimization solver. All heuristic algorithms were coded in MATLAB R2024a and executed on a computer equipped with an Intel i5-13500H 2.60 GHz CPU and 16.0 GB of RAM.

5.1. Test Cases

Each test network consists of n fully connected nodes, where n represents the total number of candidate vertiport locations in a given network. As the network is bidirectional, each pair of nodes is linked by a single undirected edge. For illustrative purposes, Figure 2 presents small complete graphs, with $n = 3$ to $n = 5$, and $n = 10$, to conceptually demonstrate the structure of fully connected networks. The actual experiments were conducted on larger networks with $n = 25$, 50, 100, and 200 which were constructed under controlled conditions. The number of candidate nodes ranged from 25 to 200, with spatial coordinates randomly generated within square regions of increasing size: 20×20 for the 25-node network, 100×100 for both the 50-node and 100-node networks, and 200×200 for the 200-node network. In particular, the 50-node and 100-node networks were constructed based on the same spatial extent (100×100). This setting was not intended to directly analyze network density, but rather to increase the diversity of both point density and experiments across the four network configurations.

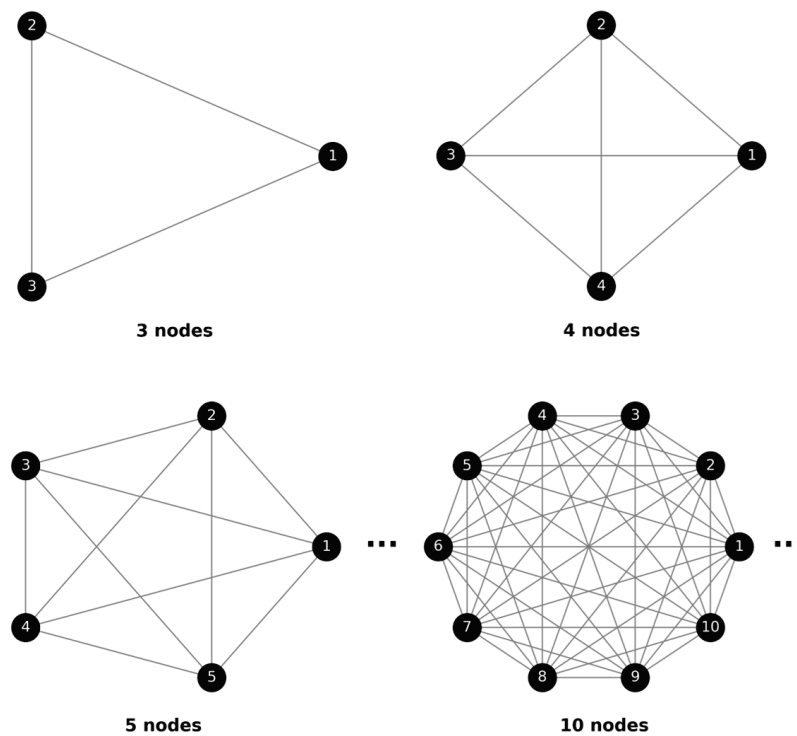


Figure 2. Illustration of fully connected networks ($n = 3$ to $n = 5$ and $n = 10$).

In each instance, the number of facilities to be selected was determined by the network size: $P = 5$ for 25 nodes, $P = 7$ for 50 nodes, $P = 9$ for 100 nodes, and $P = 10$ for 200 nodes. A maximum connection distance, denoted as D_{max} , was applied to determine the feasibility of links between candidate nodes. Serving as a core connectivity constraint, D_{max} directly shapes the feasible network topology. By systematically varying the value of D_{max} , the

study examines how different levels of connectivity influence algorithm feasibility, solution quality, and convergence behavior. Demand and facility construction cost were randomly drawn from a uniform distribution over [7, 15], introducing controlled variability while ensuring comparability across scenarios. A fixed unit cost per passenger per kilometer (α) is assumed to be 3.5, which is obtained from the real-world New York City case study in the next section. These synthetic networks do not aim to replicate specific urban settings, but function as neutral benchmarks for validating the proposed model and algorithm. Practical applicability is further demonstrated through a real-world case study presented in the subsequent section.

The experiment evaluates three solution approaches for optimizing vertiport locations while ensuring network connectivity. The first approach, MST-GA, integrates an MST-based procedure into the conventional GA to ensure selected facilities form a connected network. The second approach, MST-GA combined with GI, further enhances MST-GA by introducing a GI strategy to improve solution quality and accelerate convergence speed. The third approach utilizes CPLEX to obtain optimal solutions and serves as a benchmark for assessing heuristic methods. All approaches were tested across different network scales and various maximum connectivity distances. The results are summarized in Table 1, which reports the total cost—including transportation and facility construction costs—and the computation time for each test instance.

Table 1. Comparison of objective value and solution time.

Test Network	Objective Value (10^3)				Gap (%)	Solution Time (s)	
	$D_{\max}$	MST-GA (a)	MST-GA Combined with GI (b)	Best Solution by CPLEX Within 1 h (c)	$(b - c)/c$	MST-GA Combined with GI	CPLEX
25 nodes ($P = 5$)	10	2.13	2.10	2.10	0	1	1
	9	2.19	2.10	2.10	0	1	1
	8	2.22	2.10	2.10	0	1	1
	7	2.32	2.15	2.15	0	1	2
	6	2.53	2.33	2.33	0	1	3
	5	2.75	2.64	2.64	0	1	3
50 nodes ($P = 7$)	25	27.18	22.55	22.55	0	3	22
	23	31.19	25.98	25.98	0	3	207
	21	33.39	27.96	27.96	0	3	81
	19	36.28	32.73	32.73	0	2	186
	17	39.28	35.26	35.26	0	2	62
	15	51.31	49.68	49.68	0	2	14
100 nodes ($P = 9$)	29	48.87	43.95	43.95	0	5	308
	27	49.05	46.51	48.58	−4	5	260
	25	53.53	48.49	51.35	−6	5	801
	23	57.80	49.98	52.58	−5	6	566
	21	60.88	56.06	63.77	−12	6	808
	19	66.15	65.15	67.66	−4	7	761
	17	74.81	69.83	72.81	−4	7	1412
	15	99.58	88.18	93.53	−6	7	269
200 nodes ($P = 10$)	65	312.34	287.18	349.41	−18	5	3600
	60	357.69	302.58	392.51	−23	5	3600
	55	408.21	333.76	415.61	−20	7	3600
	50	459.61	356.94	444.86	−20	7	3600
	45	467.61	388.90	524.38	−26	7	3600
	40	514.12	429.86	538.44	−20	7	3600
	35	530.37	459.94	589.11	−22	8	3600
	30	577.99	507.56	632.29	−20	8	3600

For small-scale instances ($n = 25, 50$) in Table 1, CPLEX can obtain optimal solutions within a reasonable solution time. The proposed method, MST-GA combined with GI, achieves solution quality consistently equivalent to, and in several large-scale instances even better than, the best solutions obtained by CPLEX within an hour time limit. Specifically, in test cases with $n = 100$ and $n = 200$, MST-GA combined with GI yields objective values up to 4% to 26% lower than those of CPLEX, reflecting its ability to effectively explore the solution space in complex settings. Furthermore, compared to the basic MST-GA, the greedy-enhanced version consistently obtains lower objective values across all tested instances. While both MST-based heuristics generate feasible and connected solutions, MST-GA combined with GI significantly improves solution stability and accelerates convergence speed by integrating a GI strategy. This confirms the effectiveness of the GI in enhancing both solution quality and computational runtime efficiency. The results clearly demonstrate the advantage of MST-GA combined with GI over both CPLEX and MST-GA, particularly in large-scale scenarios, making it a practical and scalable alternative to the exact method.

The solution time comparison further confirms the scalability of the proposed approach. For small and medium networks ($n = 25, 50, 100$), CPLEX terminated within the time limit, with computation times varying by problem complexity and D_{max} . However, for all large-scale instances ($n = 200$), CPLEX consistently reached the 3600 s cap without convergence, due to the exponential complexity of the problem. In contrast, MST-GA combined with GI solved all instances, including the largest ones, in under 10 s while achieving better solutions. These results highlight the robustness, scalability, and practical effectiveness of the proposed method, confirming its suitability for solving large-scale FLPs with connectivity constraints.

To illustrate these findings visually, Figure 3 presents a comparative analysis of objective values across different problem sizes and various distance thresholds (D_{max}). It clearly shows that MST-GA combined with GI consistently outperforms both MST-GA and CPLEX, particularly in large-scale instances ($n = 100, 200$). As the problem size increases, the superiority of the proposed methodology becomes increasingly evident, as it consistently achieves lower objective values compared to the other methods. This improvement in cost efficiency highlights its strong scalability and robustness across all problem sizes. Moreover, the results indicate a general trend: as D_{max} increases, the objective value tends to decrease across all algorithms and problem sizes. This trend reflects the fact that a larger allowable connection distance provides more flexibility in facility location decisions and network connectivity, thereby enabling more cost-efficient solutions. This observation is particularly relevant for UAM network design, where D_{max} serves as a proxy for the maximum feasible inter-vertiport travel range under operational and planning constraints. In this study, D_{max} values were systematically varied from 5 km to 65 km to capture a wide range of practical scenarios, including both tightly clustered and more dispersed network configurations. While increasing D_{max} enhances network connectivity by allowing more distant inter-vertiport links, it simultaneously imposes greater demands on the energy capacity and operational coordination of UAM systems. Nevertheless, this increased flexibility can yield substantial cost savings by enabling more efficient passenger-to-vertiport assignments and reducing the total number of required facilities. Accordingly, practitioners must carefully balance technical feasibility with economic viability. This trade-off provides valuable insights for policymakers when determining whether to invest in advanced UAM technologies or to expand supporting ground infrastructure.

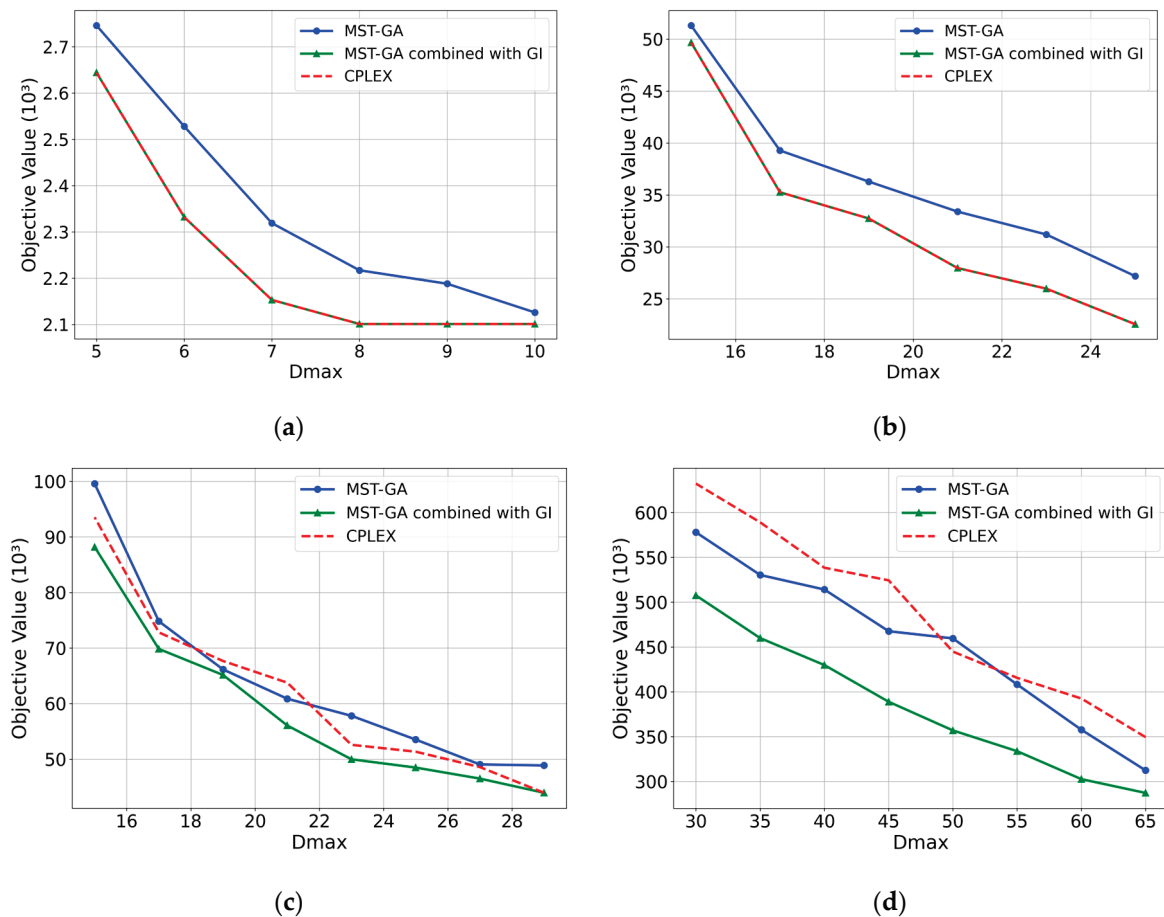


Figure 3. Comparison of the objective values for the different methods across varying problem sizes and distance thresholds: (a) 25-node network, (b) 50-node network, (c) 100-node network, (d) 200-node network.

5.2. Case Study: UAM Vertiport Locations in New York City

Following the validation of the proposed algorithm on the test networks, we extend our analysis to a real-world case study in New York City, where 444 candidate vertiport locations are considered for optimal facility placement. New York City comprises five boroughs: Manhattan, Brooklyn, Queens, the Bronx, and Staten Island. These candidate locations correspond to 423 subway stations across Manhattan, Brooklyn, Queens, and the Bronx, and 21 railway stations on Staten Island where subway service is not provided. In this study, key points within the public transportation network are chosen for candidate vertiport locations to facilitate intermodal transfer and improve overall accessibility. The proposed methodology in this study is general; candidate sites can be redefined or replaced with more context-specific locations such as tourist attractions (e.g., Times Square and the Statue of Liberty) or institutional landmarks (e.g., hospitals, museums, and universities), depending on policy goals, spatial constraints, or stakeholder priorities. Thus, the use of subway and railway stations is not an operational recommendation but rather a proxy to illustrate and evaluate the proposed algorithm under realistic demand conditions. The objective of this study is to determine a set of vertiport locations that minimizes total cost while maintaining connectivity, thereby ensuring efficient UAM operations within the urban environment. Figure 4a presents the spatial distribution of subway and railway stations represented by blue and red circles, respectively. This figure depicts coordinate projections of 444 candidate locations used in spatial analysis, and their geographic locations were manually geocoded using Google Maps to create Figure 4b. Borough boundaries

in Figure 4b are generated based on official New York City Planning datasets [34]. To ensure the practical and robust deployment of vertiports, the model incorporates a set of carefully computed parameters. Specifically, the number of UAM vertiports to be constructed, P , is assumed to be 20, in line with prior optimization studies that identify $P = 20$ as effective for demand coverage and cost-performance trade-offs [4,35]. A maximum inter-facility distance of $D_{max} = 30$ km is adopted, consistent with established practices in vertiport network design to ensure adequate service coverage and maintain intra-network efficiency [36,37].

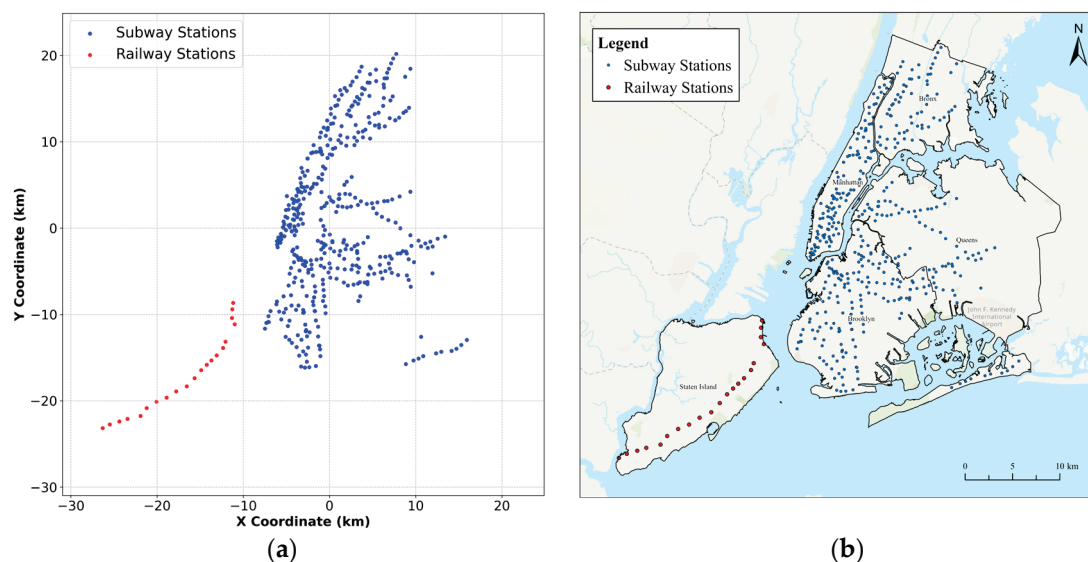


Figure 4. Spatial distribution of subway and railway stations in New York City: (a) projected coordinates for spatial analysis, (b) georeferenced station locations overlaid on administrative boundaries.

Total cost in this study includes two cost components: transportation costs incurred by users traveling to their assigned vertiports and vertiport construction costs. Transportation costs are calculated based on the amount of travel demand, travel distance from customer locations to their assigned vertiports, and unit transport cost per passenger per kilometer. In this study, all transit stations, including subway and railway stations, are treated simultaneously as demand points and candidate locations for UAM vertiport deployment. To estimate urban passenger demand for UAM in Manhattan, Brooklyn, Queens, and the Bronx, we utilized 2023 subway ridership data released by the Metropolitan Transportation Authority (MTA) [38]. The dataset includes annual ridership figures for 423 subway stations with station-level data readily available. For Staten Island, total annual railway ridership data obtained from NYC Open Data [39] was used to estimate station-level passenger demand via a spatial disaggregation approach using ArcGIS Pro 3.5 software. These values were directly used to represent demand at corresponding locations in the proposed model. Then, the estimated demand at all subway and railway stations, which serve as spatial proxies for the origins of ridership, is assigned to their nearest selected UAM vertiports based on the Euclidean distance. The resulting travel distances were used to compute the transportation costs. According to national travel behavior data, the average commuting distance per person in the United States ranges from 11 to 14 km per day, depending on region and income group [40]. Meanwhile, estimates of daily commuting costs in New York City average around USD 41 per day [41]. This implies a unit transport cost per passenger per kilometer of $\alpha = 3.5$, which is adopted in our model to reflect urban mobility conditions particularly in dense, high-cost areas like New York City.

The construction cost f_j for each candidate location $j \in J'$ is given by Formula (18):

$$f_j = C_{\text{fixed}} + P_{\text{land}} \times A \quad (18)$$

where $C_{\text{fixed}} = 2,000,000$ USD represents the fixed infrastructure cost, P_{land} denotes the unit land price (USD/m²), and $A = 6600$ m² is the required area (m²) for a standard vertiport. This cost structure reflects mid-range construction benchmarks reported in empirical design studies [42,43]. The assumed footprint is consistent with regulatory guidelines for small- to medium-scale vertiports, including safety buffers and operational zones, as consolidated from FAA and EASA standards. Since the proposed model considers candidate locations across all five boroughs of New York City distinctly, the unit land price P_{land} is computed separately for each borough. Due to the absence of an official dataset detailing borough-level standard land values in New York City, the unit land price is derived from a combination of average land price estimates, residential and commercial property values, and general market trends, based on data from publicly accessible data platforms [44–46]. The construction cost model here simplifies costs to fixed infrastructure and land acquisition. While borough-level land price variations are incorporated, finer-grained differences such as proximity to the city center or spatial concentration effects (e.g., via the Herfindahl–Hirschman Index) are not explicitly modeled. Other influential factors, including site-specific grid connection costs, complex airspace regulations, and community acceptance issues (e.g., noise or land use conflicts), may also significantly affect actual costs. These simplifications were adopted to ensure computational tractability, given the primary focus of this study on model and algorithm development. Nonetheless, future research should aim to incorporate more detailed spatial and economic cost factors to enhance planning realism.

To solve the problem, two algorithms, including (i) MST-GA and (ii) MST-GA combined with GI, are coded in this case study, and the results are compared. The first experiment applied the MST-GA to identify the near-optimal subset of 20 vertiport locations across New York City. As shown in Figure 5a, the selected sites represented by green squares form a connected network that satisfies the maximum flight distance constraint. The algorithm achieved a total cost of USD 8.48×10^9 and completed in 21.1 s. The cost convergence curve in Figure 6a demonstrates a gradual decrease in total cost over 1000 iterations. The red marker in Figure 6a indicates the result obtained at the 1000th iteration. For reference, the dashed line in the two subfigures in Figure 6 denotes the total cost achieved by MST-GA at this final iteration. To enhance the performance of the algorithm, MST-GA combined with GI was evaluated under identical settings. As illustrated in Figure 5b, this approach produced a similarly connected vertiport network while achieving a lower total cost of USD 7.85×10^9 and a significantly reduced computation time of 5.9 s. The convergence behavior shown in Figure 6b confirms that the algorithm stabilizes within 500 iterations, and the result is marked with a red square. Compared to the MST-GA, this enhanced version, incorporating a greedy algorithm, exhibits faster convergence and better solution quality while maintaining network connectivity. Figure 5c illustrates the placement of 20 UAM vertiports across New York City obtained from the MST-GA combined with GI. Each green square marks a selected vertiport location, positioned to ensure network connectivity and integration with subway and railway infrastructure. Note that urban passenger demand is represented by a heatmap derived from the public transit ridership data, which shows the effectiveness of the selected vertiport locations to cover the given UAM demand. By mapping the spatial layout of the proposed UAM network onto the actual urban landscape, the figure demonstrates the practical relevance of the proposed approach for large-scale vertiport planning in complex metropolitan settings.

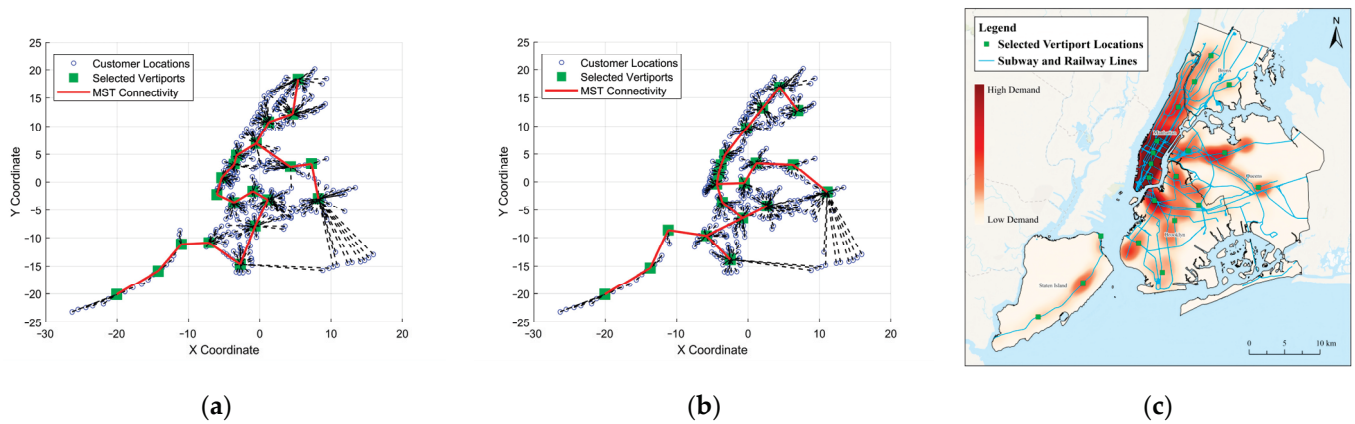


Figure 5. UAM vertiport layout solutions in New York City ($P = 20$): (a) layout generated by MST-GA, (b) layout generated by MST-GA combined with GI, (c) geographic distribution of 20 vertiport locations generated by MST-GA combined with GI overlaid on demand heatmap.

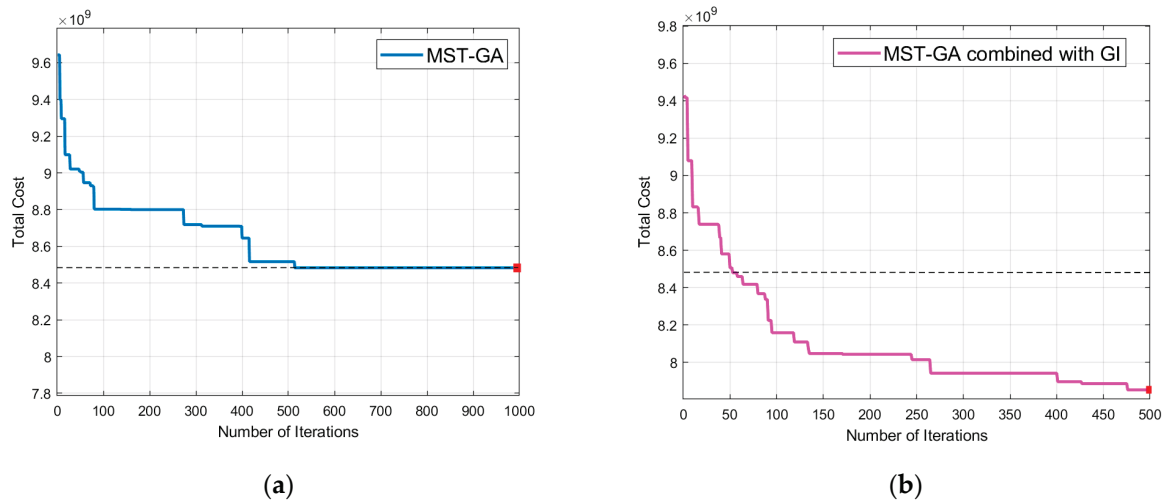


Figure 6. Total cost convergence comparison in the New York City case study: (a) MST-GA, (b) MST-GA combined with GI.

A comparative summary of the two algorithms is presented in Table 2, highlighting their respective performance in terms of total cost and computational time. Both methods successfully identified connected vertiport networks under the specified maximum flight distance constraint. However, the MST-GA combined with GI achieved a significantly lower total cost of 7.85×10^9 , compared to 8.48×10^9 for the standard MST-GA, representing a reduction of approximately 7.4%. In terms of computational efficiency, the time required to reach the corresponding solution was 5.9 s for the MST-GA combined with GI, whereas the standard MST-GA required 21.1 s; this indicates a 72% reduction in computation time while generating better solution quality. These results confirm that the proposed MST-GA framework, especially when combined with a greedy algorithm, offers an effective and scalable solution for large-scale vertiport location planning. By explicitly incorporating inter-vertiport connectivity constraints, the model addresses a key operational risk: infeasible routing, excessive detours, or network fragmentation. In practice, limited battery range and the absence of reachable alternatives can disrupt service when adjacent vertiports are not within range. Ensuring sufficient connectivity is thus essential for feasible routing, flexible detours, and overall system reliability. Its ability to integrate real-world passenger demand data, enforce spatial and connectivity constraints, and deliver high-

quality solutions with reduced computation time demonstrates its potential as a practical decision-support tool for UAM infrastructure design in complex urban environments.

Table 2. Comparison of the results obtained from MST-GA and MST-GA combined with GI.

Comparison Metrics	MST-GA	MST-GA Combined with GI
Total Cost (USD)	8.48×10^9	7.85×10^9
Solution Time (s)	21.1	5.9

6. Conclusions and Future Research

This study presents a hybrid mathematical framework for the UAM vertiport network planning, aiming to minimize total system cost while ensuring that the selected vertiport locations form a connected network. Unlike previous studies, this work explicitly incorporates connectivity constraints into the UAM vertiport location model, thereby avoiding network configurations that contain isolated nodes. The model balances passenger access cost, infrastructure investment, and network connectivity, which are essential for scalable and resilient UAM network design. To solve this problem efficiently, we propose a hybrid metaheuristic algorithm, referred to as MST-GA combined with GI, which integrates an MST-based connectivity mechanism into a GA. The MST component ensures structural feasibility by maintaining inter-vertiport links within a predefined distance, while the GI strategy generates high-quality initial solutions based on demand and cost. This combination enables efficient search space exploration and improves convergence toward feasible, high-performance solutions.

Extensive numerical experiments on synthetic networks (ranging from 25 to 200 candidate nodes) and a real-world case study in New York City demonstrate the superior performance of MST-GA combined with GI. Compared to the baseline MST-GA and CPLEX benchmarks, the proposed method consistently achieves a lower total cost, stronger feasibility, and significantly reduced runtime. From a computational perspective, MST-GA combined with GI demonstrates strong scalability and robustness. The suggested methodology ensures fast and stable convergence, even in large-scale or complex settings where traditional methods struggle. This study contributes methodologically by showing that integrating MST-based connectivity verification with a GA and a problem-specific GI strategy can substantially improve the efficiency and reliability of metaheuristics for solving large-scale, connectivity-constrained FLPs. Incorporating real-world subway and railway ridership data in the New York City case study further enhances the model's practical relevance in guiding UAM infrastructure deployment strategies by capturing spatial demand heterogeneity and supporting demand-oriented planning.

While the proposed approach proves to be effective for vertiport planning in the various case studies, future research can extend this framework to enhance its applicability and robustness in complex urban environments. The proposed framework can be extended to address multi-period planning horizons, integration with various public transit systems, uncertainties in network availability, and precise demand estimation methods. First, integrating multi-period planning and network expansion mechanisms would enable phased vertiport deployment strategies aligned with long-term urban development goals. Second, integration with multimodal transportation systems and smart mobility platforms could ensure selected locations align with existing urban mobility patterns and further increase their relevance to holistic urban transport planning. Moreover, exploring alternative connectivity formulations, such as resilience-oriented design or stochastic inter-vertiport link reliability, may further strengthen the robustness of UAM network configurations under uncertainties. Lastly, given that not all public transit passengers are potential UAM users, due to factors

such as fare sensitivity, travel purposes, and service expectations, future research should consider more representative demand estimation approaches, including adoption behavior models, stated preference surveys, or market-based forecasts. Integrating tourism mobility data, mobile device location traces, or smart card data in passenger flows could improve the model's adaptability to evolving urban mobility patterns.

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

Funding: This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant RS-2022-NR070854.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

- Long, Q.; Ma, J.; Jiang, F.; Webster, C.J. Demand analysis in urban air Mobility: A literature review. *J. Air Transp. Manag.* **2023**, *113*, 102436. [CrossRef]
- Shon, H.; Kim, S.; Lee, J. Optimal planning of urban air mobility systems accounting for ground access trips. *Int. J. Sustain. Transp.* **2024**, *18*, 356–378. [CrossRef]
- Macias, J.J.E.; Khalife, C.; Slim, J.; Angeloudis, P. An integrated vertiport placement model considering vehicle sizing and queuing. *J. Air Transp. Manag.* **2023**, *113*, 102486. [CrossRef]
- Herniczek, M.T.K.; German, B.J. Scalable combinatorial vertiport placement method for an urban air mobility commuting service. *Transp. Res. Rec.* **2024**, *2678*, 369–382. [CrossRef]
- Wang, G.; Chen, M. Performance evaluation and strategic analysis of logistics development for China Railway Express: A spatial connectivity perspective. *Systems* **2025**, *13*, 166. [CrossRef]
- Farahani, R.Z.; SteadieSeifi, M.; Asgari, N. Multiple criteria facility location problems: A survey. *Appl. Math. Model.* **2010**, *34*, 1689–1709. [CrossRef]
- Daskin, M.S. *Network and Discrete Location: Models, Algorithms, and Applications*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2011; ISBN 978-1-118-53701-5. [CrossRef]
- Ahmed, S.; Tawarmalani, M.; Sahinidis, N.V. A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Math. Program.* **2004**, *100*, 355–377. [CrossRef]
- Gen, M.; Ida, K. Advances in multiobjective hybrid genetic algorithms for intelligent manufacturing and logistics systems. In Proceedings of the International Conference on Active Media Technology, Maebashi, Japan, 29–31 October 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 379–389. [CrossRef]
- Hwang, T.; Lee, M.; Lee, C.; Kang, S. Meta-heuristic approach for high-demand facility locations considering traffic congestion and greenhouse gas emission. *J. Environ. Eng. Landsc. Manag.* **2016**, *24*, 233–244. [CrossRef]
- Melkote, S.; Daskin, M.S. An integrated model of facility location and transportation network design. *Transp. Res. Part A Policy Pract.* **2001**, *35*, 515–538. [CrossRef]
- Nagy, G.; Salhi, S. Location-routing: Issues, models and methods. *Eur. J. Oper. Res.* **2007**, *177*, 649–672. [CrossRef]
- Drexel, M.; Schneider, M. A survey of variants and extensions of the location-routing problem. *Eur. J. Oper. Res.* **2015**, *241*, 283–308. [CrossRef]
- Bai, Y.; Hwang, T.; Kang, S.; Ouyang, Y. Biofuel refinery location and supply chain planning under traffic congestion. *Transp. Res. Part B Methodol.* **2011**, *45*, 162–175. [CrossRef]
- Bardossy, M.; Raghavan, S. Robust optimization for the connected facility location problem. *Electron. Notes Discret. Math.* **2013**, *44*, 149–154. [CrossRef]
- Gewali, L.; Bhadury, J.; Chandrasekaran, R. Network design subject to facility location. In Proceedings of the 16th International Conference on Computer Applications in Industry and Engineering, Las Vegas, NV, USA, 5–7 November 2003; pp. 128–131. Available online: <https://www.researchgate.net/publication/220991317> (accessed on 31 March 2025).
- Peng, F.; Hwang, T.; Ouyang, Y. Formulation and algorithms for discrete minimum-cost group assembly problem. *Transp. Res. Rec.* **2013**, *2333*, 9–15. [CrossRef]

18. Sun, J.; Chen, Z.; Chen, Z.; Li, X. Robust optimization of a closed-loop supply chain network based on an improved genetic algorithm in an uncertain environment. *Comput. Ind. Eng.* **2024**, *189*, 109997. [CrossRef]
19. Saldanha-da-Gama, F. Facility location in logistics and transportation: An enduring relationship. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *162*, 102903. [CrossRef]
20. Martens, M.; Bley, A. Securely connected facility location in metric graphs. In *Algorithms—ESA 2010*; de Berg, M., Meyer, U., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 289–300. [CrossRef]
21. Srivastava, A.; Salapaka, M.V. Simultaneous facility location and path optimization in static and dynamic networks. *IEEE Trans. Control Netw. Syst.* **2020**, *7*, 1700–1711. [CrossRef]
22. Shishebori, D.; Ghaderi, A. An integrated approach for reliable facility location/network design problem with link disruption. *Int. J. Supply Oper. Manag.* **2015**, *2*, 640–661. [CrossRef]
23. Güner, A.R.; Sevkli, M. A discrete particle swarm optimization algorithm for uncapacitated facility Location Problem. *J. Artif. Evol. Appl.* **2008**, *2008*, 861512. [CrossRef]
24. Pradeepmon, T.G.; Paul, B. A hybrid algorithm for uncapacitated facility location problems. *Int. J. Serv. Econ. Manag.* **2011**, *3*, 197–206. [CrossRef]
25. Sevkli, M.; Guner, A.R. A continuous particle swarm optimization algorithm for uncapacitated facility location problem. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4150, pp. 316–323. [CrossRef]
26. Rodrigues, L.; Alvim, L.G.M.; Alvim, A. Multistart simulated annealing applied to a reduced scale thermo-hydraulic loop of a pressurized water reactor core. *Prog. Nucl. Energy* **2014**, *70*, 91–97. [CrossRef]
27. Yalaoui, N.; Amodeo, L.; Mahdi, H.; Yalaoui, F. Hybrid method to solve a facility layout problem: Genetic algorithm—Particle swarm optimization. *IFAC Proc. Vol.* **2009**, *42*, 1251–1255. [CrossRef]
28. Samuel, G.; Rajan, C.A. Hybrid particle swarm optimization—genetic algorithm and particle swarm optimization—shuffled frog leaping algorithm for long-term generator maintenance scheduling. *Int. J. Electr. Power Energy Syst.* **2015**, *65*, 432–442. [CrossRef]
29. Kitthamkesorn, S.; Chen, A. Maximum capture problem for urban air mobility network design. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *187*, 103569. [CrossRef]
30. Jiang, Y.; Li, Z.; Wang, Y.; Xue, Q. Vertiport location for eVTOL considering multidimensional demand of urban air mobility: An application in Beijing. *Transp. Res. Part A Policy Pract.* **2025**, *192*, 104353. [CrossRef]
31. Holland, J.H. *Adaptation in Natural and Artificial Systems*; The MIT Press: Ann Arbor, MI, USA, 1975; ISBN 978-0-262-27555-2. [CrossRef]
32. Topcuoglu, H.; Ergin, F.C.; Ermis, M.; Yilmaz, G. Solving the uncapacitated hub location problem using genetic algorithms. *Comput. Oper. Res.* **2005**, *32*, 967–984. [CrossRef]
33. Prim, R.C. Shortest connection networks and some generalizations. *Bell Syst. Tech. J.* **1957**, *36*, 1389–1401. [CrossRef]
34. NYC Department of City Planning. Borough Boundaries Dataset. NYC Open Data. n.d. Available online: <https://www.nyc.gov/content/planning/pages/resources/datasets/borough-boundaries> (accessed on 25 March 2025).
35. Zhao, Y.; Feng, T. Strategic integration of vertiport planning in multimodal transportation for urban air mobility: A case study in Beijing, China. *J. Clean. Prod.* **2024**, *447*, 142988. [CrossRef]
36. Chae, M.; Kim, S.H.; Kim, M.; Park, H.T. Vertiport locations optimization in discrete search space. In Proceedings of the 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC), Barcelona, Spain, 1–5 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [CrossRef]
37. Hess, T.; Karatas, M.; Erişkin, L. A maximal covering approach for vertiport location problem. In Proceedings of the 2024 International Conference on Decision Aid Sciences and Applications (DASA), Manama, Bahrain, 11–12 December 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [CrossRef]
38. Metropolitan Transportation Authority (MTA). Subway and Bus Ridership Statistics—2023. 2023. Available online: <https://www.mta.info/agency/new-york-city-transit/subway-bus-ridership-2023> (accessed on 31 March 2025).
39. Metropolitan Transportation Authority (MTA). LIRR Ridership—2023. *MTA Performance Metrics*. Available online: <https://metrics.mta.info/?lirr/ridership> (accessed on 31 March 2025).
40. Abbiasov, T.; Heine, C.; Glaeser, E.; Ratti, C.; Sabouri, S.; Salazar-Miranda, A.; Santi, P. *The 15-minute City Quantified Using Mobility Data*; NBER Working Paper No. 30752; National Bureau of Economic Research: Cambridge, MA, USA, 2022. [CrossRef]
41. Urban Transport Magazine. Cost of Commuting in the US—A Comparison. *Urban Transp. Mag.* Available online: <https://www.urban-transport-magazine.com/en/cost-of-commuting-in-the-us-a-comparison> (accessed on 31 March 2025).
42. Taylor, M.; Saldanli, A.; Park, A. Design of a vertiport design tool. In Proceedings of the 39th Digital Avionics Systems Conference (DASC), Herndon, VA, USA, 8–10 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–10. [CrossRef]
43. Schweiger, K.; Preis, L. Urban air mobility: Systematic review of scientific publications and regulations for vertiport design and operations. *Drones* **2022**, *6*, 179. [CrossRef]
44. PropertyShark. NYC Real Estate Market Trends. n.d. Available online: <https://www.propertyshark.com/mason/market-trends/residential/nyc> (accessed on 31 March 2025).

45. LandSearch. New York County, NY Land for Sale. n.d. Available online: <https://www.landsearch.com/properties/new-york-county-ny> (accessed on 31 March 2025).
46. Statista. Average Price of Residential Properties per Square Foot in New York City, by Property Type. n.d. Available online: <https://www.statista.com/statistics/1235728> (accessed on 31 March 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Development of a Networked Multi-Participant Driving Simulator with Synchronized EEG and Telemetry for Traffic Research

Poorendra Ramlall, Ethan Jones and Subhradeep Roy *

Department of Mechanical Engineering, Embry-Riddle Aeronautical University, 1 Aerospace Blvd, Daytona Beach, FL 32114, USA; ramlallp@my.erau.edu (P.R.); jones55@my.erau.edu (E.J.)

* Correspondence: roys5@erau.edu

Abstract: This paper presents a multi-participant driving simulation framework designed to support traffic experiments involving the simultaneous collection of vehicle telemetry and cognitive data. The system integrates motion-enabled driving cockpits, high-fidelity steering and pedal systems, immersive visual displays (monitor or virtual reality), and the Assetto Corsa simulation engine. To capture cognitive states, dry-electrode EEG headsets are used alongside a custom-built software tool that synchronizes EEG signals with vehicle telemetry across multiple drivers. The primary contribution of this work is the development of a modular, scalable, and customizable experimental platform with robust data synchronization, enabling the coordinated collection of neural and telemetry data in multi-driver scenarios. The synchronization software developed through this study is freely available to the research community. This architecture supports the study of human–human interactions by linking driver actions with corresponding neural activity across a range of driving contexts. It provides researchers with a powerful tool to investigate perception, decision-making, and coordination in dynamic, multi-participant traffic environments.

Keywords: driver behavior modeling; EEG synchronization; human-in-the-loop simulation; hyperscanning; multi-participant driving simulation; networked driving simulator; telemetry data; virtual reality

1. Introduction

Driving simulators have emerged as a powerful alternative to real-world experiments for studying driver behavior, vehicle dynamics, and traffic interactions. Their advantages are twofold: they ensure participant safety, and they offer unparalleled control over experimental conditions [1]. In contrast to on-road testing, simulators enable the reproduction of rare or dangerous scenarios—such as sudden merging, near-miss events, or collisions—without risk to human life [2]. Furthermore, in a simulator-based setup, researchers can precisely manipulate variables like the number and behavior of surrounding vehicles, road geometry, weather conditions, and even the time of day, thereby eliminating confounding effects that often complicate field studies [2–5]. However, a critical challenge remains in determining how closely participant behavior mimics real-world driving [6]. Drivers may behave differently if the simulation environment lacks realistic sensory cues. To address this, modern simulators rely on immersive technologies such as head-mounted VR headsets, motion platforms that simulate inertial feedback, and visually convincing physics engines that govern vehicle behavior [7–9]. The convergence of gaming hardware and high-fidelity simulation software has significantly reduced the cost barrier. For example, commercial

racing platforms like Assetto Corsa now offer realistic vehicle physics, advanced rendering, and customization options suitable for academic use [10].

While these developments support naturalistic driving, a major limitation persists: most driving simulators are single-participant systems. In such setups, a human driver operates in a virtual world surrounded by simulated vehicles whose behaviors follow predefined, non-adaptive scripts. This restricts interaction to a unidirectional dynamic, where the human driver reacts to static or rule-based traffic, but the surrounding vehicles do not respond to the driver's behavior in any meaningful or realistic way [11–14]. As a result, these systems fail to capture the rich, bidirectional human–human interactions observed in real-world traffic—such as mutual negotiation during lane changes, merging behavior, or implicit signaling between drivers. The literature identifies three dominant approaches to studying driver interactions: simulation models that assume driver behavior through mathematical or rule-based formulations [15,16]; conventional simulators that include only one human driver surrounded by scripted traffic [11,17,18]; and naturalistic studies conducted on real roads with human drivers [19–21]. While the first two approaches lack authenticity in capturing human–human interactions, the third introduces safety risks and uncontrollable environmental variables [11,21–23]. To overcome these limitations, networked driving simulators (Figure 1) have emerged, connecting multiple human drivers to a shared virtual environment. These systems enable dynamic, reciprocal behavior between drivers, creating opportunities to study how complex traffic interactions unfold when all participants are real people [4,11,12,24,25]. Importantly, drivers tend to behave more attentively and naturally when surrounded by other humans compared to simulated traffic, and without this human presence, it is not scientifically sound to draw conclusions about interactive driving behavior. Despite their growing relevance, very few studies have used networked simulators to collect data to build realistic models of multi-human driving interaction.

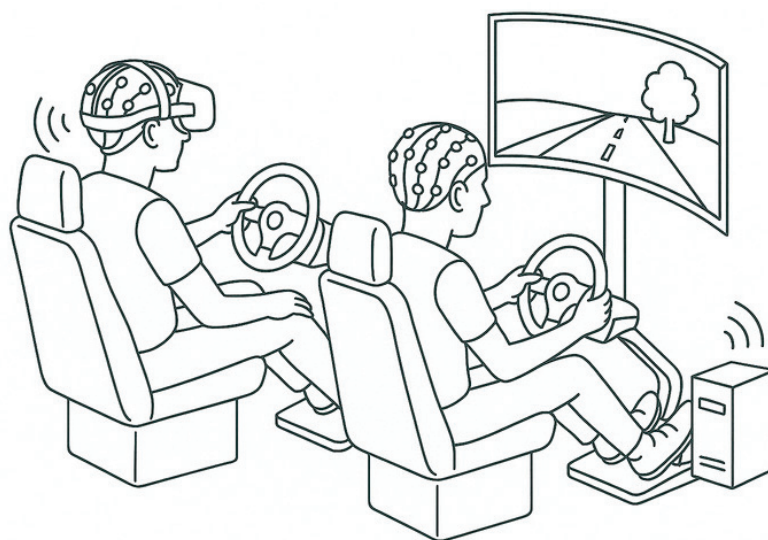


Figure 1. Illustrative schematic of the multi-participant driving simulator framework, showing two driving stations equipped with either VR headsets or curved screens. Each simulator represents an individual driver node within a shared virtual environment.

Driving simulators not only allow for the observation of behavioral responses but also provide a rich stream of vehicle telemetry, such as speed, acceleration, steering angle, brake pressure, and throttle input. These data can be supplemented with physiological measures to gain insight into drivers' internal states. One of the most powerful tools in this context is electroencephalography (EEG), which captures brain activity with high temporal

resolution and has been widely used to study attention, cognitive workload, and stress during dynamic tasks [26,27]. Traditional EEG systems, designed for clinical use, are often bulky and impractical for simulator environments, particularly those incorporating motion platforms. For such applications, dry scalp EEG systems—which require no conductive gel and can be worn comfortably—are preferable, despite their reduced signal fidelity [28–30]. Integrating EEG into a multi-participant simulator, however, introduces significant technical challenges related to data synchronization. Each participant’s EEG and driving data are captured using separate software platforms, such as Assetto Corsa for driving telemetry and a dedicated system for EEG signals. These must be precisely synchronized to a common time base to allow for their valid interpretation. This complexity is compounded when scaling to multiple participants, where synchronization must be ensured not just within each participant’s data but also across all participants involved in the experiment. Such alignment is crucial when studying collaborative or competitive driving tasks—like highway merging or overtaking—where researchers may wish to compare the timing of brake pressure and neural firing across individuals. In this paper, we present, for the first time, a comprehensive framework for a multi-participant networked driving simulator integrated with dry EEG, with the full synchronization of behavioral and neural data streams. We introduce a custom-built software that synchronizes data acquisition across multiple systems and participants, making it significantly easier for other researchers to adopt this cost-effective and scalable approach.

2. System Architecture

The experimental system consists of a set of immersive, networked driving simulators designed to study human driving behavior and cognitive activity in a shared virtual environment. Each simulator functions as an independent node within a multi-participant traffic simulation, allowing for multiple participants to drive simultaneously while their driving telemetry and EEG data are recorded in real time. This section describes the core components of the system, including the hardware and software used in each simulator and the EEG apparatus.

2.1. Driving Simulator Configuration

The driving simulator used in this study is designed to mimic the physical and perceptual cues of real-world driving as closely as possible, while remaining modular and cost-effective. Each simulator (Figure 2) includes the following: a physical driving cockpit (seat and frame), a motion platform to provide inertial feedback, a steering wheel and pedal set, a visual interface (either a high-resolution curved monitor or a VR headset), the Assetto Corsa simulation software, and a dry-electrode EEG headset paired with a trigger hub. The simulators are connected via a local server running Assetto Corsa, allowing participants to interact with one another in a shared virtual scenario. While the configuration supports flexibility in hardware selection, such as different wheelbases or display types, compatibility with Assetto Corsa must be ensured. Together, these elements create an immersive and interactive driving environment where naturalistic driver behavior can be observed and recorded. The system is built with flexibility and reproducibility in mind, allowing for individual components to be modified or upgraded based on experimental needs and participant preferences.



Figure 2. Experimental setup showing a participant operating the multi-participant driving simulator. Key components are labeled as follows: (1) driving cockpit, (2) motion platform, (3) steering wheel and pedals, (4) curved display screen, (5) VR headset, (6) EEG headset, and (7) wireless trigger hub.

2.1.1. Driving Cockpit

The driving cockpit serves as the physical foundation of the simulator. It consists of a rigid frame and an adjustable racing seat, both sourced from Next Level Racing. The frame is built from aluminum profiles and steel reinforcements, offering the structural integrity necessary to support a motion platform and high-torque steering systems. It also includes mounting interfaces for the pedal box, steering column, and motion platform, ensuring ergonomic alignment and minimal vibration interference.

The cockpit's adaptable design is crucial for participant comfort and realism. Seat position, tilt, and distance from the pedals can all be adjusted to accommodate drivers of different heights and postures, allowing them to assume a natural driving position. This is important not only for realism, but also for minimizing muscular strain during extended trials and reducing motion artifacts in EEG recordings. The cockpit also includes brackets and cable channels for securing wiring and peripheral devices, reducing clutter and preventing mechanical interference during operation.

2.1.2. Motion Platform

A key feature of the simulator is the integration of a motion platform between the seat and the base of the cockpit. The motion platform used in this setup is the Motion Platform V3 by Next Level Racing, which is designed for racing and flight simulators. It provides two degrees of freedom—pitch and roll—simulating longitudinal and lateral vehicle dynamics such as braking, acceleration, and cornering forces.

The motion feedback adds kinesthetic realism to the visual experience, enabling participants to feel the forces associated with driving maneuvers. This haptic feedback plays a critical role in immersing the participant and eliciting naturalistic responses to stimuli, such as adjusting steering angle or braking force in response to perceived deceleration. Calibration of the platform was performed using the manufacturer's Platform Manager software, which is iteratively tuned to provide realistic yet non-fatiguing movement profiles.

The gain and scaling of the motion response were adjusted to match the physical limits of the driving scenario while accounting for participant safety and comfort.

Importantly, the motion platform must operate silently and with minimal mechanical delay to avoid introducing latency between the visual and haptic feedback, as this can result in motion sickness or reduced immersion. The careful integration and isolation of vibrations were also implemented to ensure they did not introduce noise into the EEG signal during recording. To reduce mechanical noise that could interfere with EEG recordings, vibration isolation was implemented by mounting and tuning steel springs between the base of the motion platform and the cockpit frame. These springs absorb residual vibrations and decouple platform motion from the EEG headset, contributing to a more stable neural signal acquisition.

2.1.3. Steering Wheel and Pedals

The steering and pedal system forms the primary input interface for the driver. In our setup, the steering wheel system includes a direct-drive servo motor base and an interchangeable wheel rim, both compatible with Assetto Corsa. Direct-drive steering systems are preferred over gear- or belt-driven alternatives due to their high torque resolution, instantaneous force feedback, and absence of mechanical lag. This ensures the driver feels realistic resistance and feedback during cornering, collisions, or road texture changes, enhancing behavioral fidelity.

The pedal set includes independent modules for the clutch, brake, and accelerator, each with customizable tension and travel distance. Load-cell-based braking is used to measure actual pedal force rather than displacement, providing a more accurate representation of real-world braking behavior. This level of granularity is essential in experiments focused on fine motor control and reaction time measurements. Both the steering and pedal systems are mounted securely to the cockpit frame to eliminate flex and shifting, preserving consistent input-response characteristics across trials.

2.1.4. Visual Interface

Visual immersion is achieved through either a high-resolution curved monitor (Samsung R55 series) or a VR headset (HTC Vive Pro 2), depending on participant preference and experimental requirements. The curved monitor offers a panoramic field of view that helps approximate the peripheral vision available in a real vehicle. With a wide aspect ratio and high refresh rate, it ensures smooth rendering of the simulation environment, minimizing aliasing and screen tearing.

For a more immersive experience, the VR headset is used to place the participant directly within the 3D driving environment. The HTC Vive Pro 2 features dual OLED displays with a combined resolution of 4896 x 2448 and a refresh rate of 120 Hz, providing crisp visuals and minimal latency. Head tracking allows the simulation to dynamically adjust the driver's point of view based on head orientation, reinforcing spatial awareness and realism.

However, VR is not suitable for all participants. Some may experience discomfort, disorientation, or motion sickness due to the sensory mismatch between the visual and vestibular systems. For this reason, the system offers the option to use either visual interface, allowing researchers to select the most appropriate display mode based on participant tolerance and experimental goals. Regardless of their choice, the simulator delivers a consistent visual experience by maintaining identical field-of-view angles, screen distances, and rendering parameters across modalities.

2.1.5. Simulation Software: Assetto Corsa

The simulation environment is powered by Assetto Corsa, a commercially available racing simulator known for its realistic vehicle physics, detailed graphics, and extensive modding capabilities. Although designed primarily for entertainment, Assetto Corsa's open-source asset architecture makes it highly suitable for academic research. Many commercial driving simulator platforms are expensive and do not support multi-driver scenarios [10]. Frutos et al. [10] evaluated multiple racing simulators for their suitability in transportation research and found that Assetto Corsa scored highly in both graphics quality and physical realism.

Its physics engine models tire friction, suspension behavior, drivetrain dynamics, and aerodynamic forces with high fidelity, creating a convincing driving experience. This realism is crucial when studying naturalistic behavior, as drivers are more likely to engage meaningfully with the environment when it behaves as expected. Additionally, Assetto Corsa supports multiple input types, customizable force feedback profiles, and telemetry output through shared memory, allowing researchers to collect real-time data on speed, throttle, braking, steering angle, gear selection, and more.

One of Assetto Corsa's most valuable features for this project is its ability to host local multiplayer sessions via a dedicated server. This enables multiple simulators to be networked into a shared driving scenario, allowing human drivers to interact with one another in real time. Because all vehicles operate in the same virtual environment, it is possible to observe genuine driver–driver interactions, such as merging, overtaking, and cooperative maneuvers. The system supports the importing of custom-designed tracks and vehicles, allowing experiments to be tailored to specific road geometries, traffic densities, or behavioral hypotheses.

The combination of Assetto Corsa's physics accuracy, real-time data access, and multiplayer networking makes it a powerful platform for cognitive and behavioral driving research. When paired with physiological data collection methods like EEG, it provides a rich multimodal dataset for investigating human performance in complex traffic scenarios.

2.2. EEG Apparatus and Configuration

To capture participants' brain activity, the EEG setup includes the DSI-24 dry electrode headset and a trigger hub (Figure 3), both manufactured by Wearable Sensing [31,32]. The DSI-24 collects EEG signals using 21 dry electrodes arranged according to the international 10–20 system [33,34]. This electrode placement method ensures standardized coverage of key cortical regions while allowing for rapid deployment and minimal participant discomfort.



(a) DSI-24



(b) Trigger Hub

Figure 3. (a) DSI-24 dry-electrode EEG headset and (b) wireless trigger hub used for EEG synchronization, both manufactured by Wearable Sensing [31,32].

Each electrode records small variations in scalp potential due to underlying neural activity. These signals are sampled and streamed using the DSI Streamer software, which also supports synchronized data capture across multiple headsets. The trigger hub provides an external pulse to all EEG headsets at the start of each experiment, enabling the precise temporal alignment of EEG recordings across participants.

Dry-electrode EEG systems present a compelling solution for traffic research involving driving simulators, particularly in studies requiring rapid participant turnover, naturalistic behavior, and integration with immersive technologies like VR. Compared to traditional wet systems, dry EEG offers significant practical advantages, such as a minimal setup time, enhanced participant comfort, and better compatibility with head-mounted displays, making them ideal for non-clinical, real-world experimental settings [35–37]. These strengths are especially valuable in dynamic experiments where participants are exposed to motion feedback and frequent headset reuse is needed. While dry EEG systems typically exhibit higher electrode impedance and may be more susceptible to motion artifacts and environmental noise, these challenges can be mitigated with proper headset fit and calibration, particularly for participants with dense or curly hair. The lower electrode count does reduce spatial resolution, but for applications focused on cognitive state estimation rather than fine-grained source localization, this tradeoff is acceptable. Overall, the balance of usability and operational efficiency makes dry EEG a practical and effective choice for large-scale, motion-enabled traffic simulation studies [35–37].

3. Methodology

3.1. Data Collection

The system supports the simultaneous recording of two parallel data streams for each participant: EEG signals and driving telemetry. EEG data can be sampled at 300 Hz using the DSI-24 headset and managed through the DSI Streamer software. In parallel, the system can capture detailed driving telemetry, including speed, steering angle, brake pressure, and throttle input, via a custom Python 3.13 script that interfaces with the Assetto Corsa's shared memory API [38]. While Assetto Corsa's server provides updates at approximately 8 Hz, the script is designed to sample key driving variables at up to 300 Hz, aligning with the EEG sampling rate to facilitate high-resolution synchronization across modalities.

To ensure synchronized EEG recording across multiple simulators, the system uses a trigger hub that sends a simultaneous pulse to all EEG headsets at the start of each experiment. This mechanism enables the consistent alignment of EEG data across participants, independent of individual system clocks. Driving telemetry logging, in contrast, is initiated manually on each simulator PC, which means timestamps across telemetry files are not inherently synchronized. The system architecture is designed to generate two parallel raw data streams per participant, one for EEG signals and one for driving telemetry, supporting flexible post-processing and alignment strategies.

The EEG files include time-indexed data from 21 electrodes, metadata fields, and trigger markers. The driving data files include time-series values for vehicle parameters, a local PC clock variable (`pc_time`), and a clock (`session_time_remaining`) that represents the countdown of the simulation session.

3.2. Data Processing and Synchronization

To enable meaningful cross-participant comparisons, the synchronization of these datasets is essential. The `session_time_remaining` variable is first converted to a forward-counting value, `time_elapsed`, and all time variables are reformatted into consistent units (seconds). Coordinate strings are parsed and converted into numerical values where necessary.

Since all EEG headsets are triggered simultaneously using the trigger hub, the EEG datasets are inherently synchronized. Similarly, driving telemetry files from different simulators can be aligned using the shared `session_time_remaining` clock. However, to avoid introducing artifacts from EEG interpolation, synchronization is anchored to the EEG `Time` variable. All driving telemetry data are interpolated to match this reference timeline.

The `pc_time` values are synchronized by anchoring them to the EEG `Time`, which is calibrated across participants using a hardware trigger provided by the EEG trigger hub. Driving data streams from each simulator are synchronized using the `time_elapsed` variable derived from the Assetto Corsa server clock `session_time_remaining`, which itself is synchronized via NTP across machines. Additionally, each recording captures a reference `pc_time` at the start of both EEG and simulator data collection. Using the initial EEG `pc_time` and incremental `Time` values, a synthetic `pc_time` series is constructed. The simulator's `pc_time` is then locally aligned with this synthetic EEG clock. Global synchronization is achieved by referencing all data streams to the unified EEG `Time`.

The final result is a fully synchronized dataset in which each time sample corresponds to one EEG time point, containing EEG data and aligned driving telemetry from all participants. This unified dataset allows researchers to explore high-resolution patterns in driver behavior and brain activity under both solo and interactive driving conditions. The process of data synchronization is illustrated in Figure 4.

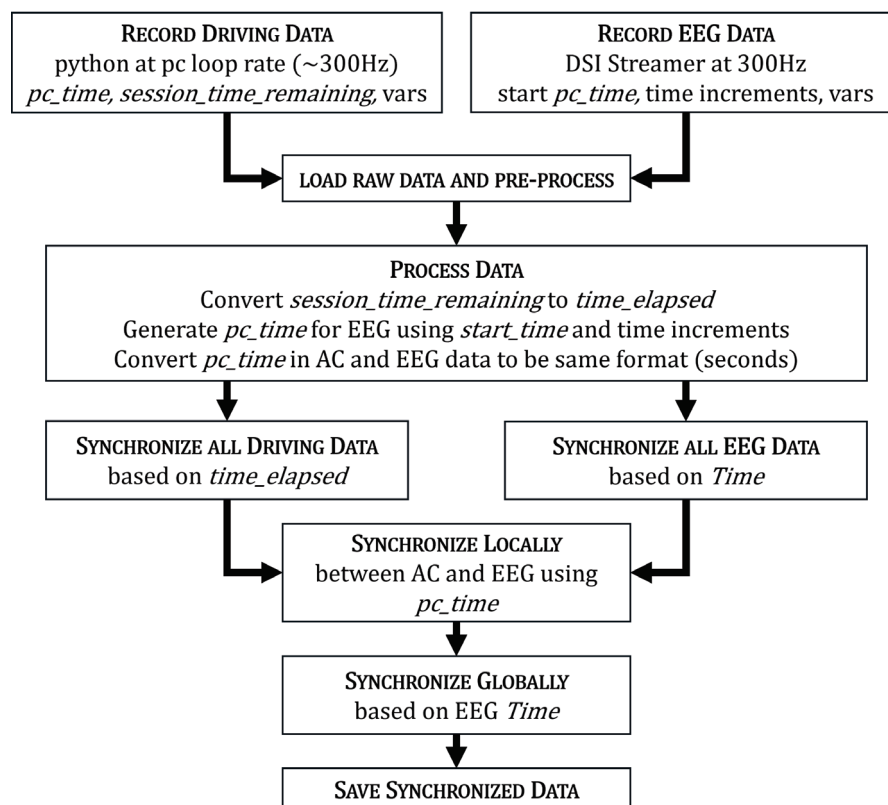


Figure 4. Flowchart illustrating the data synchronization process used to align EEG signals and driving telemetry across multiple participants.

Although individual differences exist in EEG signals and driving behavior, synchronization in this framework is based on a shared temporal reference rather than behavioral or cognitive similarity. All data streams are aligned using a unified time base derived from the EEG trigger pulse and the simulator's session clock. This approach ensures consistent cross-participant alignment while preserving the uniqueness of each participant's response dynamics.

This synchronization ability is the central contribution of the presented work. While each EEG signal is inherently individual, the unified temporal alignment allows for future exploration of between-participant neural relationships, such as inter-brain coupling in different driving scenarios—for example, collaborative (merging) or competitive (overtaking) tasks.

Graphical User Interface for Data Synchronization

To support the synchronization approach described above and enhance overall usability, a custom graphical user interface (GUI) called *syncApp* was developed using MATLAB R2024a and JavaScript, and is publicly accessible at <https://www.subhradeeprooy.com/software> (accessed on 31 June 2025).

A snapshot of this GUI is shown in Figure 5. The GUI enables users to load raw EEG and driving data, execute synchronization routines, and visualize time-aligned data streams. Researchers can inspect variable plots, validate data quality, and export synchronized datasets for further analysis. The interface is designed to lower the technical barrier for researchers interested in replicating or extending this experimental setup in their own laboratories.

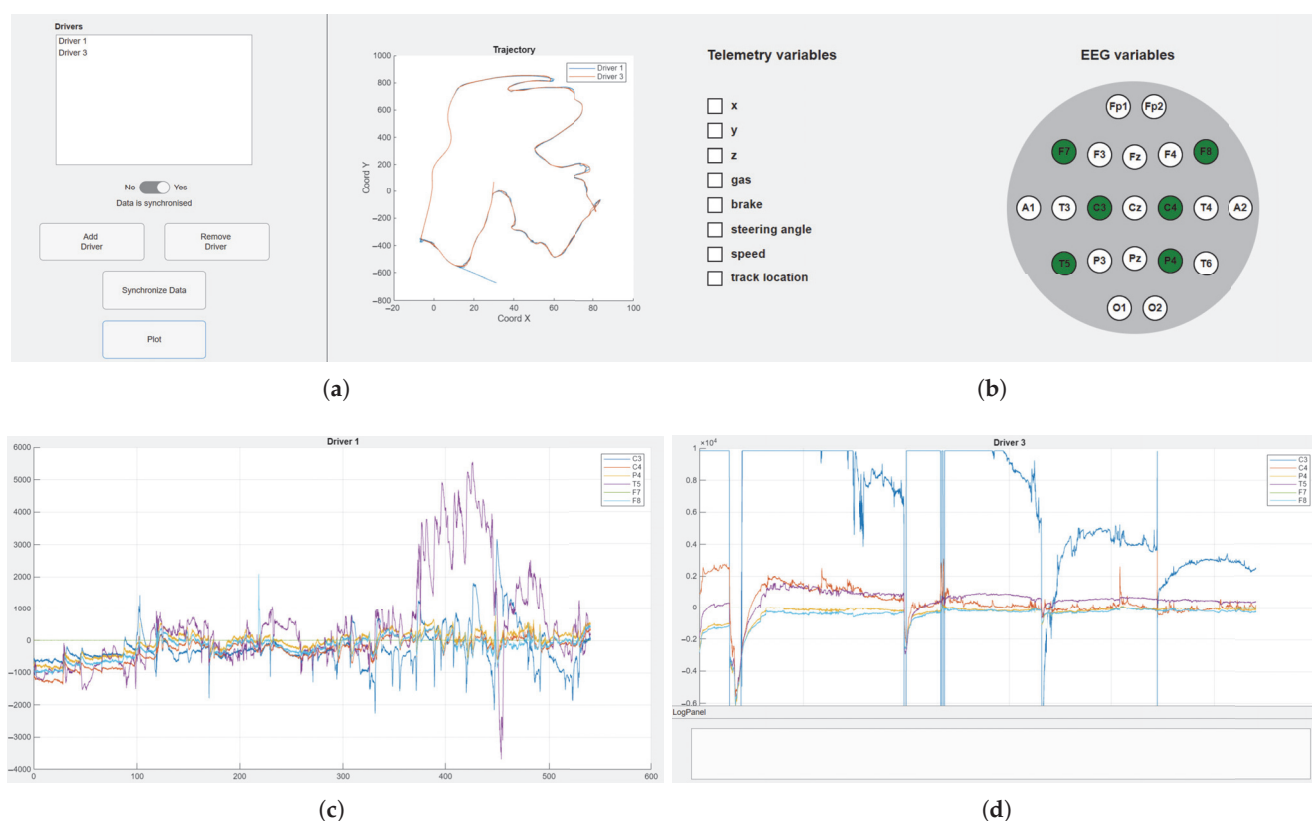


Figure 5. Snapshots of GUI interface showing data loading and sample EEG channel visualization for two drivers. (a) data loading and synchronization controls; (b) telemetry and EEG Channel selection; (c) data visualization for the first driver; (d) data visualization for the second driver.

While the GUI does not conduct EEG signal analysis itself, it implements the synchronization algorithm shown in Figure 4 to generate precisely time-aligned EEG and driving telemetry datasets. These synchronized outputs are compatible with widely used EEG analysis platforms such as EEGLAB [39], MNE-Python [40], and QStates [41]. By formatting the data for direct use, the GUI enables streamlined signal-level analyses—including spectral decomposition, phase-locking value computation (as demonstrated in our prior work [42]), and other cognitive state assessments. More importantly, this synchronization capability

extends beyond conventional single-participant analysis. It creates new opportunities for advanced multi-participant paradigms such as hyperscanning, allowing researchers to explore inter-brain neural coupling and collective cognitive dynamics during complex driving interactions, an area rarely addressed in prior traffic simulation studies. This opens a path toward studying cognition not only at the individual level but also at the networked group level in realistic, dynamic environments.

The GUI consists of two key features: data loading and synchronization (shown in Figure 5a), and the visualization of various synchronized variables for the drivers (illustrated in Figure 5b–d). The GUI is designed to handle both asynchronous data and data that has already been synchronized. During the data loading process, a toggle button is provided to indicate whether the data is raw or pre-synchronized. If pre-synchronized data is loaded, no further synchronization is required, and the data can be visualized immediately.

4. System Validation and Usability Testing

This set-up is designed to be easily replicated. To achieve the best results, minimize latency, and maximize immersion, the minimum performance characteristics should include an Intel i9 processor (or similar), an NVIDIA 3060Ti GPU (or similar), and at least 16 GB of RAM. A major consideration when designing this system is its overall usability, reliability, and resilience. Consequently, tests are conducted to validate the system design.

Since the set-up uses commercially available and well-tested software (Assetto Corsa), there is higher confidence in the reliability of this environment. When incorporating the motion feedback, the nature of the simulation results in the movement of the subject within their seat. Additionally, as subjects reposition and turn their heads, several motion artifacts appear in the EEG data. These motion artifacts can be identified by their characteristic spread across multiple EEG channels, which distinguishes them from localized neural activity, and can then be attenuated using standard preprocessing techniques. The EEG system ensures data reliability by transmitting recordings wirelessly to a local computer using Bluetooth communication, while also maintaining a redundant copy through onboard storage. After various trials, the signal quality of the EEG is verified to be excellent, with no apparent discontinuities or breaks in the transmission.

Pilot tests are conducted during and after the development of the synchronization algorithm to verify that the data are being synchronized properly, by means of computing the lag times. After interpolation of the driving data and complete synchronization with the higher sampled EEG data, the data points are exactly aligned. In addition, a general problem of data collection from real systems is the challenge of maintaining precise clock sampling due to hardware imperfections and transmission times, among other issues [43]. Before synchronization, the driving data and EEG data contain a clock drift of approximately 3.312 ms and 3.333 ms, respectively, after about 30 min of continual data collection. After synchronization, the latency across the two datasets is measured to be under 2 ms from repeated trials, with a mode of around 1.66 ms.

To evaluate system usability, 15 participants tested the simulator and provided structured feedback on comfort, immersion, and control through a post-session survey, provided in Appendix A. Their input was instrumental in refining several aspects of the setup, including the tuning of motion feedback amplitude and pedal stiffness, and determination of optimal experiment durations to minimize discomfort.

Three core usability dimensions were assessed: comfort, realism, and handling. Comfort was based on participant experiences with the VR headset, EEG headset, motion feedback, overall driving feel, and any VR-related disorientation (survey questions 1 through 6). Realism reflected user perceptions of the virtual environment, immersive quality, and the physical realism of motion and haptic feedback (questions 7 through 10). Handling

captured impressions of the responsiveness of the brake, accelerator, and steering wheel (questions 11 through 13). The full list of survey questions is provided in the Appendix A. The average ratings across these categories are reported in Table 1.

Table 1. Average participant ratings for key simulator characteristics including comfort, realism, and handling, collected during the system tuning and usability testing phase.

Characteristic	Average Score [1–10]	Survey Questions
Comfort	7.5	1–6
Realism	8.0	7–10
Handling	8.1	11–13
Overall Rating	7.8	

On average, participants rated comfort at 7.5/10, realism at 8.0/10, and handling at 8.1/10, resulting in an overall usability rating of 7.8/10 for the system. The most common sources of discomfort were motion sickness during prolonged VR use and physical strain from the EEG headset when worn for extended periods (over one hour). These findings support the inclusion of a mixed interface option, allowing participants to choose between VR and monitor displays, and highlight the importance of limiting session durations to enhance user comfort and data quality.

5. Demonstration of Synchronized Data and Sample EEG Analysis

The developed framework enables precise temporal synchronization between EEG and driving telemetry data collected from multiple participants. Figure 6 presents an example of successfully synchronized data, showing time-aligned gas and brake pedal inputs alongside O1 electrode EEG activity for two drivers participating in a shared driving scenario. This level of synchronization allows researchers to explore how neural and behavioral responses unfold in parallel across interacting participants, supporting rich multi-modal analysis of driving behavior.

While detailed EEG analysis is not the primary focus of this study, the synchronized EEG and driving telemetry datasets produced by the system are fully compatible with standard cognitive state assessment tools. For example, Figure 7 illustrates a sample application using QStates software, where concentration levels during a sustained attention task are estimated based on α -band EEG activity. Both a linear estimator and a multivariate normal probability density function (MVNPDF) are applied to classify cognitive states, with lower α -activity being associated with higher concentration and elevated activity corresponding to relaxation. Similar analyses can be extended to other frequency bands (e.g., beta, delta, theta, gamma) to support broader assessments of driver attention, workload, and fatigue during various driving tasks.

Beyond individual-level analysis, the framework also enables more advanced applications, such as *hyperscanning*, which examines inter-brain synchrony and coordinated neural activity across multiple participants. While hyperscanning has seen widespread use in social neuroscience, it remains largely unexplored in driving research due to the absence of systems capable of collecting synchronized multi-user EEG and behavioral data. By bridging this gap, the present system lays the foundation for future investigations into shared decision-making, neural coordination, and real-time interaction in dynamic traffic environments.

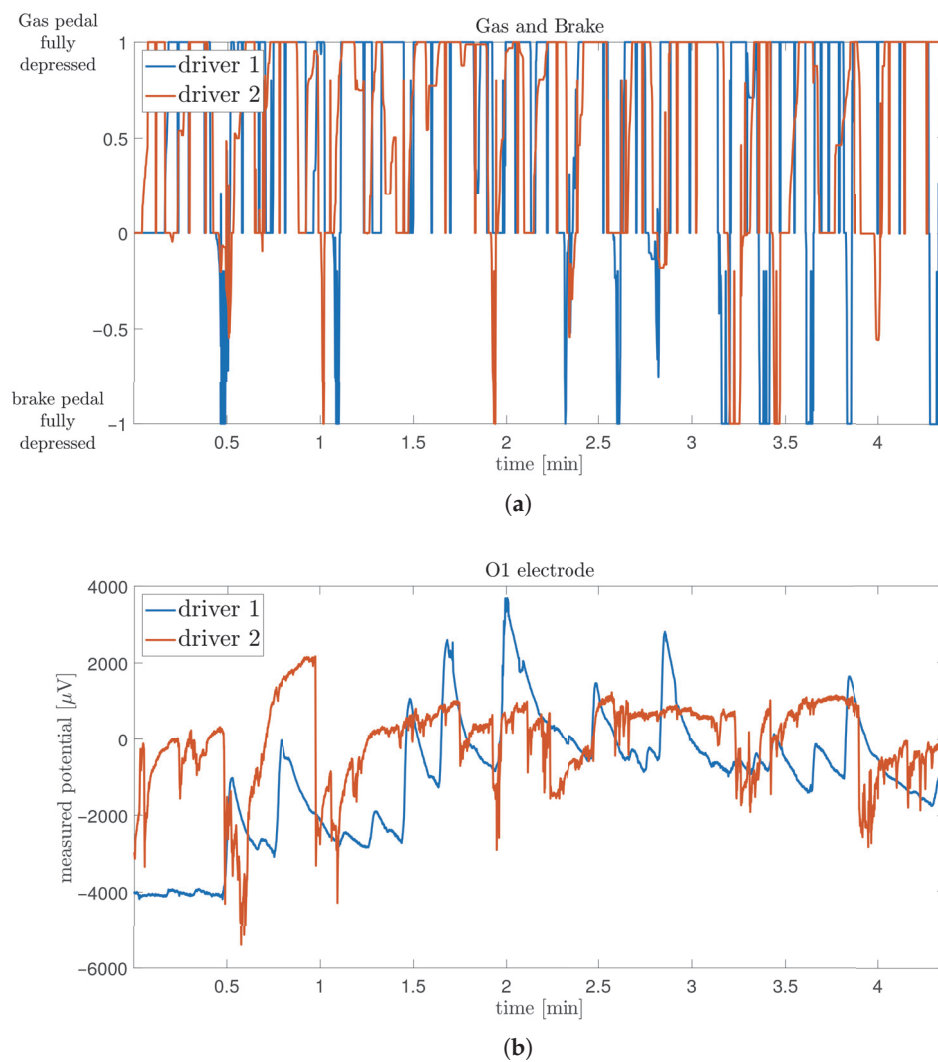


Figure 6. Example of synchronized data showing gas and brake pedal inputs along with O1 electrode EEG signals recorded from two drivers during a shared driving session. (a) Gas and brake values. (b) O1 electrode values.

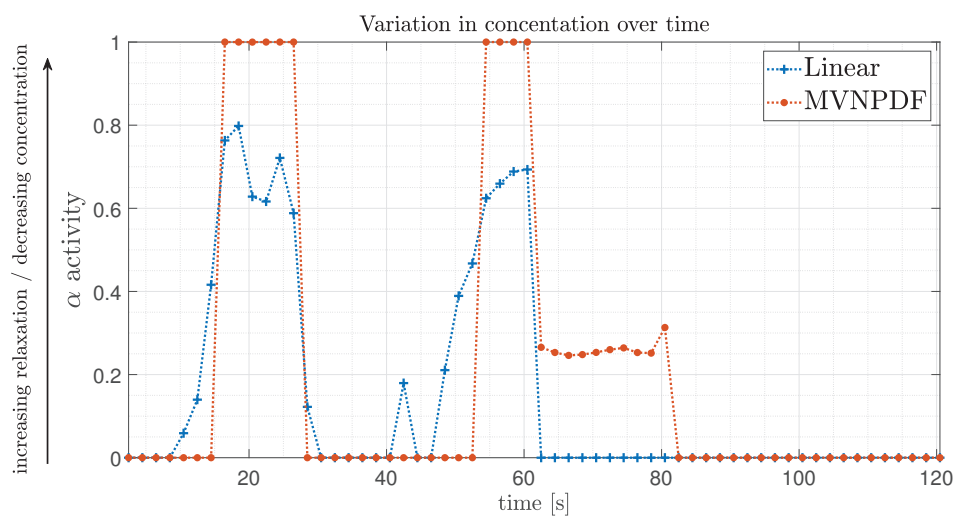


Figure 7. Variation in concentration levels during a sustained attention task, classified based on alpha band EEG activity.

6. Comparison to Other Networked Simulators

To contextualize the capabilities of the developed framework, we compare it against existing networked driving simulators based on six key criteria: scalability, support for VR, EEG integration, immersion, availability, and cost. The immersion level is qualitatively assessed based on factors such as motion and haptic feedback, environmental realism (e.g., weather and lighting), physics fidelity, and visual rendering.

Table 2 summarizes this comparison. As shown, the present system is one of the few that combines high immersion, multi-participant scalability, and synchronized EEG capability, while remaining accessible and cost-effective for broader research adoption.

Table 2. Comparison between the proposed simulation framework and existing networked driving simulators across key evaluation criteria.

Study	Scalable	VR	EEG	Immersion	Availability	Cost
CDSL (present study)	yes	yes	yes	very high	commercial	mid
DLR [24]	no	no	no	high	research	high
NII [24]	yes	no	no	very low	commercial	low
OSU [24]	no	no	no	high	research	high
Goedicke et al. [44]	no	yes	no	mid	commercial	low
Miller et al. [45]	yes	yes	no	very low	commercial	low
Yang et al. [46]	yes	no	no	low	research	mid
Xu and Lin [47]	yes	no	no	low	commercial	low

Unlike prior work that either involves single drivers or lacks neural measurement, our system uniquely enables synchronized cognitive and behavioral data collection from multiple interacting participants. This lays the foundation for future studies on human–human driving interactions, shared decision-making, and inter-brain neural dynamics in traffic settings.

7. Driving Scenarios and Environmental Conditions

The framework supports the simulation of a wide range of driving scenarios under dynamic conditions, such as varying weather conditions and different times of day. Vehicle handling and responsiveness were adapted using realistic physics models—for instance, reduced traction on wet roads and limited visibility during fog or nighttime conditions. These features allow researchers to examine how drivers perceive and respond to naturalistic environmental challenges.

To support targeted investigations of driver behavior under diverse conditions, we designed a flexible and modular virtual environment. A custom virtual route was developed (Figure 8) and divided into distinct zones featuring diverse traffic and road conditions. These segments were designed to support the study of a range of driving behaviors, including single-lane car-following, lane-changing maneuvers involving lateral interactions between vehicles, and responses to varying road surface quality, from rough dirt segments to smooth paved sections.

Two illustrative examples are shown in Figures 9 and 10. Figure 9 illustrates a single-lane car-following scenario under three distinct environmental conditions—clear daytime, foggy daytime, and clear nighttime—captured at the same location along the route to highlight differences in visibility and ambient context. Figure 10 illustrates a two-vehicle merging scenario, where a red car is merging into the lane occupied by a white car. From the white car’s perspective, the red vehicle is visible directly ahead, while the red car’s driver could see the white car through the side mirror.

This setup demonstrates how drivers experience the same scenario from different viewpoints and interact in real-time, enabling the system to support real-time behavioral coupling between participants. When paired with EEG data, such scenarios can support future investigations into the cognitive and behavioral aspects of human–human driving interactions.



Figure 8. Custom-designed virtual route developed to support a variety of driving scenarios for experimental testing.

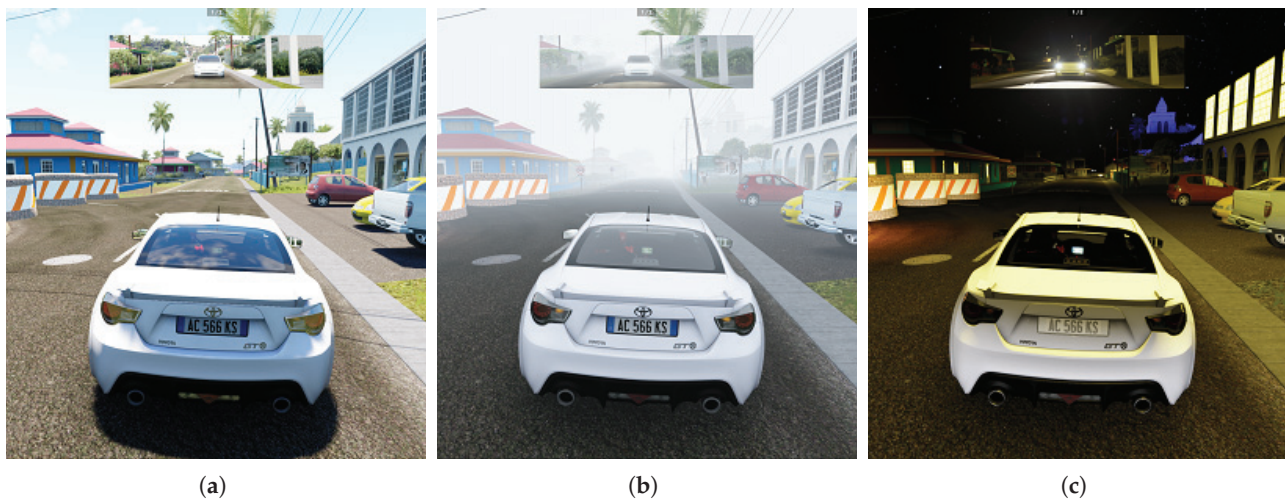


Figure 9. Example simulation conditions, including variations in weather and time of day, for a single-lane car-following scenario on the custom-designed virtual route. (a) clear weather conditions; (b) foggy weather conditions; (c) clear weather conditions, nighttime.



Figure 10. Merging scenario on the custom-designed virtual route, showing the different perspectives of the involved drivers. (a) merging scenario: third-person view; (b) merging scenario: perspective of white car; (c) merging scenario: perspective of red car.

8. Limitations

While the system presented in this paper offers a robust platform for synchronized multi-participant traffic experiments, several limitations remain. First, although dry-electrode EEG systems provide ease of use and faster setup times, they can be susceptible to motion artifacts and signal noise. This may impact the fidelity of neural data, especially in motion-rich scenarios involving aggressive driving maneuvers or platform-induced vibrations.

To minimize motion-related artifacts during EEG acquisition, it is recommended that headsets are fitted snugly to ensure stable electrode contact. Environmental conditions should be optimized, such as maintaining a cool room temperature, to reduce perspiration, and participants should be advised to avoid physical exertion and spicy food prior to recording. A 1–50 Hz bandpass filter, applied in the DSI Streamer software, can attenuate high-frequency noise, including 60 Hz power line interference. Although this study does not focus on EEG analysis, the recorded EEG data are fully compatible with standard preprocessing pipelines, such as EEGLAB [39] and QStates [41], allowing future users of the framework to perform artifact rejection and advanced signal analysis as needed [39,48,49].

Second, the current system architecture does not incorporate additional physiological sensors such as eye-trackers, galvanic skin response monitors, or heart rate sensors, which could offer complementary insights into the driver's cognitive and emotional state. While integration is technically feasible, it requires further development and synchronization infrastructure. Third, although Assetto Corsa offers excellent realism and modding capabilities, future work may benefit from transitioning to fully open-source simulation platforms for greater experimental control and transparency. Finally, scalability is currently constrained by hardware requirements and local networking limitations. While the system can support multiple participants, expanding to large-scale simulations with a dozen or more drivers may require significant infrastructure upgrades and performance optimization. Addressing these limitations will be a key focus of future work aimed at extending the platform's capabilities and ensuring broader usability across research domains.

9. Conclusions

This paper presents a comprehensive, modular, and cost-effective framework for conducting traffic experiments using networked, multi-participant driving simulators integrated with synchronized EEG data collection. Built upon commercially available hardware and open simulation platforms, the system replicates key physical and perceptual cues of real-world driving, enabling the observation of naturalistic human behavior in a controlled lab environment. Through the use of Assetto Corsa, motion-enabled cockpits, flexible visual interfaces, and dry-electrode EEG systems, we demonstrate a setup that is both scalable and adaptable to various experimental needs.

A key contribution of this work lies in the synchronization methodology, which enables the fine-grained temporal alignment of behavioral (driving) and neural (EEG) data both within and across participants. The system captures rich datasets that can inform

future studies on driver cognition, cooperative and competitive behavior, and distributed decision-making in traffic scenarios. We also release the synchronization software, along with sample datasets, to facilitate broader adoption and collaborative development.

By demonstrating system performance, data quality, synchronization accuracy, and compatibility with EEG analysis tools, this work establishes the necessary foundation for future research that integrates behavioral and neural data in multi-participant traffic scenarios.

10. Future Work

Beyond enabling synchronized data acquisition, this platform opens the door to a wide range of novel research questions that have previously been difficult to study due to the lack of an affordable setup capable of supporting multi-driver experiments with synchronized data streams. For example, researchers can investigate how drivers cognitively and behaviorally respond to complex traffic scenarios such as near-miss incidents, merging conflicts, or cooperative maneuvers. The system also supports multi-participant synchronization, facilitating hyperscanning studies that examine inter-brain neural coupling during real-time human–human driving interactions—an area largely unexplored in traffic research. Its flexibility allows for the implementation of diverse driving tasks under varying conditions such as different times of day, different road geometries, and varying weather conditions, all while maintaining experimental control. The modular setup further enables comparisons between hardware configurations (e.g., VR headset vs. monitor) to examine how different components affect cognitive engagement. Together, these capabilities address key gaps in the literature, where most of the existing work relies on single-driver setups or lacks physiological insight. By providing a robust framework for synchronized EEG and telemetry data collection, this work lays the foundation for future cognitive and behavioral analyses in interactive driving contexts.

Author Contributions: Within the scope of this study, the authors P.R. and S.R. contributed equally in most aspects. The author E.J. was responsible for the creation of the GUI under the guidance of S.R. and P.R. The following briefly outlines the contributions based on each category. Conceptualization, S.R. and P.R.; methodology, S.R.; software, P.R.; validation, P.R. and S.R.; formal analysis, P.R., E.J. and S.R.; resources, S.R.; data curation, P.R.; writing—original draft preparation, P.R.; writing—review and editing, P.R. and S.R.; visualization, P.R.; supervision, S.R.; project administration, S.R.; funding acquisition, S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by National Science Foundation CAREER award (CMMI- 2238359).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article. The GUI developed for this study is dubbed *syncApp* and is available at <https://www.subhradeeprooy.com/software> (accessed on 31 June 2025).

Acknowledgments: During the preparation of this manuscript, the author(s) used ChatGPT 4o for the purposes of generating the illustrative schematic of the driving simulator. The authors thank the Mechanical Engineering Department and the Dean of the College of Engineering at Embry-Riddle Aeronautical University for providing the laboratory space that made this research possible.

Conflicts of Interest: The authors declare no conflicts of interest. The funding agency 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.

Appendix A

Survey Questions

Please rate each statement from **0 (Strongly Disagree/Not at all)** to **10 (Strongly Agree/Extremely)** by checking one box per row.

#	Statement:	0	1	2	3	4	5	6	7	8	9	10
Section 1: Comfort												
1	I was comfortable using the Virtual Reality environment	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
2	I was not disoriented by the VR environment	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
3	I did not encounter any motion sickness	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
4	The EEG cap was comfortable to wear	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
5	The feedback from the motion platform (seat) was tuned to a comfortable level	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
6	The driving simulator allowed for a comfortable driving experience	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Section 2: Realism												
7	I have extensive real-world driving experience	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
8	The VR environment improved immersiveness	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	The simulation environment is adequately realistic	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
10	The feedback from the motion platforms improved immersiveness	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Section 3: Handling												
11	The brake pedal was comfortable to use	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
12	The gas pedal was comfortable to use	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
13	The steering on the simulator rig was comfortable to use	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

References

1. Krasniuk, S.; Toxopeus, R.; Knott, M.; McKeown, M.; Crizzle, A.M. The effectiveness of driving simulator training on driving skills and safety in young novice drivers: A systematic review of interventions. *J. Saf. Res.* **2024**, *91*, 20–37. [CrossRef] [PubMed]
2. Alonso, F.; Faus, M.; Riera, J.V.; Fernandez-Marin, M.; Useche, S.A. Effectiveness of Driving Simulators for Drivers' Training: A Systematic Review. *Appl. Sci.* **2023**, *13*, 5266. [CrossRef]
3. de Winter, J.; Leeuwen, P.M.; Happee, R. Advantages and Disadvantages of Driving Simulators: A Discussion. In Proceedings of the Measuring Behavior Conference, Utrecht, The Netherlands, 28–31 August 2012.

4. Abdelgawad, K.; Gausemeier, J.; Grafe, M.; Berssenbrügge, J. Interest Manager for Networked Driving Simulation Based on High-Level Architecture. *Designs* **2017**, *1*, 3. [CrossRef]
5. Chen, F.; Terken, J.; Chen, F.; Terken, J. Driving Simulator Applications. In *Automotive Interaction Design: From Theory to Practice*; Springer Nature: Singapore, 2023; pp. 239–256.
6. Chen, L.; Fang, J.; Li, J.; Xie, J. Research on the Effectiveness of Driving Simulation Systems in Risky Traffic Environments. *Systems* **2025**, *13*, 329. [CrossRef]
7. Tiu, J.; Harmon, A.C.; Stowe, J.D.; Zwa, A.; Kinnear, M.; Dimitrov, L.; Nolte, T.; Carr, D.B. Feasibility and Validity of a Low-Cost Racing Simulator in Driving Assessment after Stroke. *Geriatrics* **2020**, *5*, 35. [CrossRef]
8. Zhang, S.; Zhao, C.; Zhang, Z.; Lv, Y. Driving simulator validation studies: A systematic review. *Simul. Model. Pract. Theory* **2025**, *138*, 103020. [CrossRef]
9. Feliciani, C.; Crociani, L.; Gorrini, A.; Nagahama, A.; Nishinari, K.; Bandini, S. Experiments and Usability Tests of a VR-Based Driving Simulator to Evaluate Driving Behavior in the Presence of Crossing Pedestrians. In *Proceedings of the Traffic and Granular Flow 2019*, Pamplona, Spain, 2–5 July 2019; Springer International Publishing: Cham, Switzerland, 2020; pp. 471–477.
10. de Frutos, S.H.; Castro, M. Assessing sim racing software for low-cost driving simulator to road geometric research. *Transp. Res. Procedia* **2021**, *58*, 575–582. [CrossRef]
11. Paz, A.; Veeramisti, N.; Khaddar, R.; de la Fuente-Mella, H.; Modorcea, L. Traffic and Driving Simulator Based on Architecture of Interactive Motion. *TheScientificWorldJournal* **2015**, *2015*, 340576. [CrossRef]
12. Muehlbacher, D. The Multi-Driver Simulation: A Tool to Investigate Social Interactions Between Several Drivers. In *UR:BAN Human Factors in Traffic: Approaches for Safe, Efficient and Stress-free Urban Traffic*; Bengler, K., Drüke, J., Hoffmann, S., Manstetten, D., Neukum, A., Eds.; Springer Fachmedien: Wiesbaden, Germany, 2018; pp. 379–391.
13. Lane, D.; Roy, S. Validating a data-driven framework for vehicular traffic modeling. *J. Phys. Complex.* **2024**, *5*, 025008. [CrossRef]
14. Lane, D.; Roy, S. Using information theory to detect model structure with application in vehicular traffic systems. *IFAC PapersOnLine* **2023**, *56*, 367–372. [CrossRef]
15. Łach, L.; Svyetlichnyy, D. Comprehensive Review of Traffic Modeling: Towards Autonomous Vehicles. *Appl. Sci.* **2024**, *14*, 8456.
16. Calvert, S.; van Arem, B. A generic multi-level framework for microscopic traffic simulation with automated vehicles in mixed traffic. *Transp. Res. Part C Emerg. Technol.* **2020**, *110*, 291–311. [CrossRef]
17. Prevati, G.; Mastinu, G. SUMO Roundabout Simulation with Human in the Loop. *SUMO Conf. Proc.* **2023**, *4*, 29–40. [CrossRef]
18. Godley, S.T.; Triggs, T.J.; Fildes, B.N. Driving simulator validation for speed research. *Accid. Anal. Prev.* **2002**, *34*, 589–600. [CrossRef]
19. Ashley, G.; Osman, O.A.; Ishak, S.; Codjoe, J. Investigating Effect of Driver-, Vehicle-, and Road-Related Factors on Location-Specific Crashes with Naturalistic Driving Data. *Transp. Res. Rec.* **2019**, *2673*, 46–56. [CrossRef]
20. Transportation Research Board and National Academies of Sciences, Engineering, and Medicine. *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk*; The National Academies Press: Washington, DC, USA, 2014. [CrossRef]
21. Ehsani, J.P.; Harbluk, J.L.; Bärghman, J.; Williamson, A.; Michael, J.P.; Grzebieta, R.; Olivier, J.; Eusebio, J.; Charlton, J.; Koppel, S.; et al. Naturalistic Driving Studies: An Overview and International Perspective. In *International Encyclopedia of Transportation*; Vickerman, R., Ed.; Elsevier: Oxford, UK, 2021; pp. 20–38.
22. Fitch, G.; Socolich, S.; Guo, F.; McClafferty, J.; Olson, R.; Pérez-Toledano, M.; Hanowski, R.; Hankey, J.; Dingus, T. *The Impact of Hand-Held and Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk*; Technical Report; Virginia Polytechnic Institute and State University: Blacksburg, VA, USA, 2013.
23. Carsten, O.; Kircher, K.; Jamson, S. Vehicle-based studies of driving in the real world: The hard truth? *Accid. Anal. Prev.* **2013**, *58*, 162–174. [CrossRef]
24. Abdelgawad, K.; Gausemeier, J.; Dumitrescu, R.; Grafe, M.; Stöcklein, J.; Berssenbrügge, J. Networked Driving Simulation: Applications, State of the Art, and Design Considerations. *Designs* **2017**, *1*, 4. [CrossRef]
25. Roy, S. Quantifying interactions among car drivers using information theory. *Chaos: Interdiscip. J. Nonlinear Sci.* **2020**, *30*, 113125. [CrossRef]
26. Casson, A.J. Wearable EEG and beyond. *Biomed. Eng. Lett.* **2019**, *9*, 53–71. [CrossRef]
27. Cao, Z.; Chuang, C.H.; King, J.K.; Lin, C.T. Multi-channel EEG recordings during a sustained-attention driving task. *Sci. Data* **2019**, *6*, 19. [CrossRef]
28. Usman, S.M.; Khalid, S.; Akhtar, R.; Bortolotto, Z.; Bashir, Z.; Qiu, H. Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: Review of available methodologies. *Seizure* **2019**, *71*, 258–269. [CrossRef] [PubMed]
29. Li, Z.; Fields, M.; Panov, F.; Ghatan, S.; Yener, B.; Marcuse, L. Deep Learning of Simultaneous Intracranial and Scalp EEG for Prediction, Detection, and Lateralization of Mesial Temporal Lobe Seizures. *Front. Neurol.* **2021**, *12*, 705119. [CrossRef] [PubMed]
30. Sultana, M.; Jain, O.; Halder, S.; Matran-Fernandez, A.; Nawaz, R.; Scherer, R.; Chavarriaga, R.; del R. Millán, J.; Perdikis, S. Evaluating Dry EEG Technology Out of the Lab. In *Proceedings of the 2024 IEEE International Conference on Metrology*

- for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE), St Albans, UK, 21–23 October 2024; pp. 752–757.
31. Wearable Sensing. DSI-24. Available online: <https://wearablesensing.com/dsi-24> (accessed on 10 April 2025).
32. Wearable Sensing. Wireless Trigger Hub. Available online: <https://wearablesensing.com/wireless-trigger-hub> (accessed on 10 April 2025).
33. Acharya, J.N.; Hani, A.; Cheek, J.; Thirumala, P.; Tsuchida, T.N. American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.* **2016**, *33*, 308–311.
34. EMOTIV. Understanding the 10-20 System of EEG Electrode Placement. Available online: <https://www.emotiv.com/blogs/how-to/understanding-the-10-20-system-of-eeeg-electrode-placement> (accessed on 10 April 2025).
35. Bang, J.S.; Won, D.O.; Kam, T.E.; Lee, S.W. Motion Sickness Prediction Based on Dry EEG in Real Driving Environment. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 5442–5455. [CrossRef]
36. Wang, F.; Chen, D.; Lu, B.; Wang, H.; Fu, R. A Novel Semi-Dry Electrode Suitable for Collecting EEG Data for Detecting Driving Fatigue in Long-Period Driving Case. *IEEE Sens. J.* **2023**, *23*, 17891–17900. [CrossRef]
37. Zander, T.O.; Andreessen, L.M.; Berg, A.; Bleuel, M.; Pawlitzki, J.; Zawallich, L.; Krol, L.R.; Gramann, K. Evaluation of a Dry EEG System for Application of Passive Brain-Computer Interfaces in Autonomous Driving. *Front. Hum. Neurosci.* **2017**, *11*, 78. [CrossRef]
38. Romagnoli, G. ACSharedMemoryDocumentation. 2015. Available online: <https://assetocorsamods.net/threads/doc-shared-memory-reference-acc.3061> (accessed on 17 November 2024).
39. Delorme, A.; Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef]
40. Gramfort, A.; Luessi, M.; Larson, E.; Engemann, D.A.; Strohmeier, D.; Brodbeck, C.; Goj, R.; Jas, M.; Brooks, T.; Parkkonen, L.; et al. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **2013**, *7*, 267. [CrossRef]
41. McDonald, N.J.; Soussou, W. QUASAR’s QStates cognitive gauge performance in the cognitive state assessment competition 2011. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 6542–6546.
42. Beauchene, C.; Roy, S.; Moran, R.; Leonessa, A.; Abaid, N. Comparing brain connectivity metrics: A didactic tutorial with a toy model and experimental data. *J. Neural Eng.* **2018**, *15*, 056031. [CrossRef]
43. Salimnejad, M.; Pappas, N.; Kountouris, M. So timely, Yet so stale: The impact of clock drift in real-time systems. *Comput. Res. Repos.* **2025**, *2501*, 549. [CrossRef]
44. Goedicke, D.; Zolkov, C.; Friedman, N.; Wise, T.; Parush, A.; Ju, W. Strangers in a Strange Land: New Experimental System for Understanding Driving Culture Using VR. *IEEE Trans. Veh. Technol.* **2022**, *71*, 3399–3413. [CrossRef]
45. Miller, J.; Kalivarapu, V.; Holm, M.; Finseth, T.; Williams, J.; Winer, E. A Flexible Multi-Modal Multi-User Traffic Simulation for Studying Complex Road Design. In Proceedings of the 40th Computers and Information in Engineering Conference (CIE), International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Virtual, Online, 17–19 August 2020; Volume 9, p. V009T09A065.
46. Yang, M.; Bao, Q.; Shen, Y.; Qu, Q.; Zhang, R.; Han, T.; Zhang, H.; Gao, M. Strategic crossing decisions in vehicle interactions at uncontrolled intersections: A networked driving simulator study. *Accid. Anal. Prev.* **2025**, *215*, 107990. [CrossRef] [PubMed]
47. Xu, J.; Lin, Y. Impact of Distracted Drivers on Individual and Group Behavior of Following Vehicles: A Networked Multi-Driver Simulator Study. *Transp. Res. Rec.* **2018**, *2672*, 45–54. [CrossRef]
48. Delorme, A.; Sejnowski, T.; Makeig, S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage* **2007**, *34*, 1443–1449. [CrossRef]
49. Michel, C.M.; Brunet, D. EEG Source Imaging: A Practical Review of the Analysis Steps. *Front. Neurol.* **2019**, *10*, 325. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Length Requirements for Urban Expressway Work Zones' Warning and Transition Areas Based on Driving Safety and Comfort

Aixiu Hu ^{1,2}, Ruiyun Huang ^{2,3}, Yanqun Yang ^{2,3,*}, Ibrahim El-Dimeery ⁴ and Said M. Easa ^{2,5}

¹ Jinshan College of Fujian Agriculture and Forestry University, Fuzhou 350002, China; huaixiu@fafu.edu.cn

² Joint International Research Laboratory on Traffic Psychology & Behaviors, Fuzhou University, Fuzhou 350116, China; seasa@torontomu.ca (S.M.E.)

³ College of Civil Engineering, Fuzhou University, Fuzhou 350116, China

⁴ Faculty of Postgraduate Studies and Scientific Research, German University in Cairo, Cairo 11835, Egypt

⁵ Department of Civil Engineering, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada

* Correspondence: yangyanqun@fzu.edu.cn

Abstract: As aging urban expressways become more pronounced, maintenance and construction work on these roadways is increasingly necessary. Some lanes may need to be closed during maintenance and construction, decreasing driving safety and comfort in the work zone. This situation often leads to traffic congestion and a higher risk of traffic accidents. Notably, 80% of work zone traffic accidents occur in the warning and upstream transition areas (or simply warning and transition areas). Therefore, it is crucial to appropriately determine the lengths of these areas to enhance both safety and comfort for drivers. In this study, we examined three different warning lengths (1800 m, 2000 m, and 2200 m) and three transition lengths (120 m, 140 m, and 160 m) using the entropy weighting method to create nine simulation scenarios on a two-way, six-lane urban expressway. We selected various metrics for driving safety and comfort, including drivers' eye movement, electroencephalogram, and driving behavior indicators. A total of 45 participants (mean age = 23.9 years, standard deviation = 1.8) were recruited for the driving simulation experiment, and each participant completed all 9 simulation scenarios. After eliminating 5 invalid datasets, we obtained valid data from 40 participants. We employed a combination of the analytic network process and entropy weighting method to calculate the comprehensive weights of the eight evaluation indicators. Additionally, we introduced the fuzzy theory, utilizing a trapezoidal membership function to evaluate the membership matrix values of the indicators and the comprehensive evaluation grade eigenvalues. The ranking of the experimental scenarios was determined using these eigenvalues. The results indicated that more extended warning lengths correlated with increased safety and comfort. Specifically, the best driver safety and comfort levels were observed in Scenario I, which featured a 2200 m warning length × 160 m transition length. However, the difference in safety and comfort across different transition lengths diminished as the warning length increased. Therefore, when road space is limited, a thoughtful combination of reasonable lengths can still provide high driving safety and comfort.

Keywords: urban expressway; work zone safety; warning length; transition length; driving simulation; entropy weighting method; fuzzy evaluation

1. Introduction

Urban expressways are integral to urban roads. They are provided with a central separation belt and have four or more lanes, which are either fully or partially controlled by grade separation, allowing vehicles to drive at a higher speed. They serve short commutes within the city, enhancing urban traffic efficiency and shortening travel time. Their flow is concentrated during peak hours. After decades of construction, China's expressway system has undergone gradual improvement, and urban expressway maintenance and construction work will become the norm. The driving environment becomes more complicated, and the road traffic capacity is reduced because some lanes need to be closed during maintenance and construction of urban expressways, which decreases driving safety and comfort in work zones. The work zone becomes the bottleneck section on urban expressways and the accident-prone point of urban roads.

According to the Standardization Administration of the People's Republic of China [1], the work zone is divided into six elements: advance warning area, upstream transition area, buffer area, work area, downstream transition area, and termination area, as shown in Figure 1 [1]. For simplicity, the first two areas are referred to here as warning and transition areas. The warning area allows drivers to adjust their speeds and change lanes by looking for gaps, making its length crucial for traffic diversion. The transition area helps prevent sudden changes in traffic flow during lane changes, facilitating a smooth transition of vehicles; therefore, its length is significant. Relevant research shows that traffic accidents in the warning and transition areas account for 80% of the total traffic accidents in the work zone [2,3]. Therefore, reasonably setting the warning and transition lengths is essential for improving driving safety and comfort in the work zone.

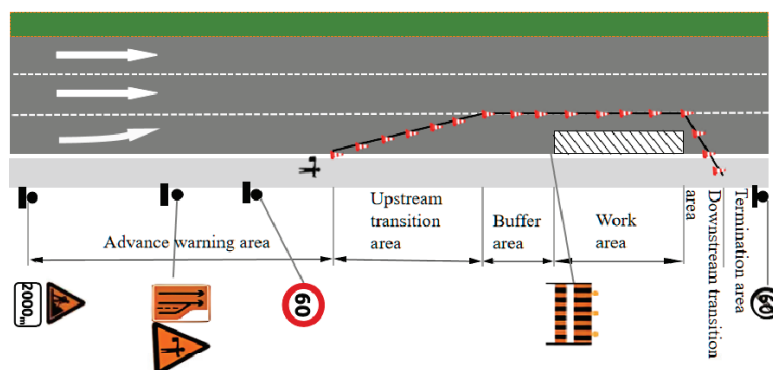


Figure 1. Elements of the work zone.

1.1. Studies on Warning and Transition Areas

There are few studies available in the literature on urban expressway work zones. However, studies on highway work zones have achieved specific results that can serve as a reference for urban expressways. Li et al. [4] recommended considering the total length required for lane changing without deliberate deceleration as the length of the warning area based on the vehicle lane change model. Ge et al. [5] proposed a method for determining the length of the highway warning area using the VISSIM simulation model and the surrogate safety assessment model for agents (SSAM). Liu et al. [6] established a model for setting the length of the transition area under various construction schemes based on the theory of acceptable gap. Based on the lane change conflict model, Liu [7] evaluated the length of warning and transition areas from dynamic and driver psychological perspectives. Wang et al. [8] selected three spatial safety evaluation indicators to analyze the spatial safety of the transition area under different operating speeds and mixed traffic flow conditions. Zhang et al. [9] proposed optimal work zone lengths for different

traffic volumes using optimization. Singh et al. [10] set up a work area model using a cellular automaton model and improved traffic congestion by optimizing the work area length. Han et al. [11] analyzed the variation in road network evaluation parameters for the work zone length under different traffic volumes and organization modes. They proposed reasonable recommended values for the work zone length.

There are quite a few norms and standards for work zones in China, such as the Road Traffic Signs and Markings Part 4: Work Zones [1], and Safe Operation Regulations published by the Ministry of Transport [12], as well as the management measures for construction sections formulated by the various provinces and the subordinate local departments. Although the standards stipulate the layout of the work zone, only the minimum value is determined for the length of various construction sections, which makes it difficult for highway management and construction units to accurately determine the length during implementation. At the same time, a series of safety accidents have exposed problems: the existing norms have low standards, broader provisions, and a mismatch with the current traffic situation.

In summary, existing research primarily focuses on highway work zones, and the length of the warning and transition areas is based solely on road traffic characteristics, without considering drivers' driving factors or the influence of different combinations of section lengths. Although urban expressways are similar to highways, they differ in traffic functions and characteristics, making it challenging to apply existing results directly. Therefore, analyzing the length of warning and transition areas on urban expressways is necessary, considering traffic and driving characteristics.

1.2. Studies on Driving Safety and Comfort

There are few comprehensive evaluations of driving safety and comfort because most studies analyze an indicator or specific indicators regarding the evaluation mode of construction areas. Omkar et al. [13] analyzed traffic safety in terms of the conflict probability at selected roads with and without work zone sections along the same road. Tang et al. [14] evaluated vehicle safety in highway work zones by analyzing collision risk using vehicle trajectory data collected from work zones. Garber et al. [15] analyzed traffic accident data in work zones and summarized the leading causes and characteristics of accidents. Zhang et al. [16,17] used various analytical models to evaluate the impact of different traffic flow conditions on risk in highway work zones. Ishak et al. [18], Umar et al. [19], and Li et al. [20,21] analyzed various construction sites. They found that the frequency of traffic accidents was higher in the warning and transition areas and the work area. Li et al. [22] examined the effect of exceptional weather, different driving experiences, and gender on traffic safety in work zones. Heaslip et al. [23] used algorithms to assess the impact of drivers' familiarity with the road and driving behavior on traffic safety. Several studies have found that drivers' behavior and cognitive abilities also impact road capacity [24,25]. Khan [26] divided subjective comfort into 5 levels and studied the subjective comfort of 108 drivers when the length of the transition area was decreased by 10% to 40% (10% interval). The study did not consider the safety indicators of vehicle operation in the work zone.

With the advancement of science and technology, measurement methods for assessing people's perceptions and emotions are becoming increasingly accurate, providing strong support for establishing the relationship between the characteristics of the road complex environment and the mental and physiological indicators of the driver [27,28]. Based on visual indicators, Belyusar et al. [29] found that dynamic billboards significantly affect drivers' driving, increasing driving risk. Costa et al. [30] found a linear relationship between visual recognition distance and duration, which can affect the acquisition of key

information on the sign. Li et al. [31] concluded that tunnel brightness has a significant impact on mental and physiological indicators of the driver, including gaze duration, saccade amplitude, and blink frequency, etc. Zhou et al. [32] conducted vehicle tests in four kinds of tunnels, using EEG power spectral density, pupil diameter, and other indicators to investigate the driver's visual load and tension. Wang [33] investigated the relationship between traffic safety and biomass index at the entrance of highway tunnels based on pupil diameter and EEG power ratio $(\alpha + \theta)/\beta$, as well as other indicators.

A driver's vehicle handling behavior can also reflect driving safety and comfort to some extent. This behavior includes controlling vehicle speed, controlling the driving route, and handling emergencies. It indicates that the driver's driving safety and comfort are poor if they frequently change lanes, speed, or brake suddenly [34,35]. Therefore, vehicle handling behavior can also characterize driving safety and comfort during the driving process in the work zone.

1.3. Study Objectives

This paper focuses on enhancing driving safety and comfort in work zones along urban expressways. The study creates a simulated environment of a work zone on a two-way, six-lane urban expressway to carry out indoor driving simulation experiments. The objectives of the study are threefold: (1) to evaluate the performance of various combinations of warning and transition area lengths, (2) to implement comprehensive performance indicators, including visual, cognitive, and driving behavior effects, on driver's safety and comfort, including drivers' eye movements and electroencephalogram (EEG), and (3) to develop a systematic evaluation methodology that combines the analytic hierarchy process (AHP) and the entropy weighting method (EWM) to address the limitations of both subjective and objective weighting approaches. The results of this research can serve as a reference for determining the appropriate lengths of work zones on urban expressways in the future. This paper breaks through the limitations of traditional single evaluation indices and develops a multidimensional evaluation system encompassing physiology, psychology, and behavior that addresses the comprehensive evaluation of driver safety and comfort. This study focuses on the impact of the combination of the warning area and transition area lengths, rather than the length of a single section. It examines the interaction of different combinations on driving safety and comfort, providing a basis for refining the design of the work area.

2. Experimental Design

2.1. Experimental Scenarios

Since there is no stipulation for the length of each section of the work zone on the urban expressway in the existing norms, this paper selected the length of each section according to the standards of the first-class highway [12]. The length of the warning area was gradually increased by 200 m based on meeting the minimum length of 1600 m in the norms, and a total of six lengths was established. The length of the transition area gradually increased by 20 m, meeting the minimum length of 100 m in the norms, and a total of 6 lengths was set up. Additionally, an 80 m buffer area, a 1000 m work area, a 30 m downstream transition area, and a 30 m termination area were established according to the regulations [12].

Peng et al. [36], Ren [37], and Yang [38] used VISSIM simulation. They selected the delay rate, conflict rate, and average travel time as evaluation indicators to determine the optimal ranges of the lengths of warning and transition areas for different types of work zones. Therefore, in this study, six VISSIM simulation models of the warning length and six VISSIM simulation models of the transition length were constructed us-

ing the control variable method. The VISSIM simulation experiment output the results of five evaluation indicators: average queue length, number of traffic conflicts, average travel delay, average travel time, and parking times. According to the results of the five evaluation indicators, the lengths of the warning and transition areas were comprehensively evaluated, respectively, based on the entropy weight method [36–40]. The results showed that the warning area's lengths were ranked from best to worst, as follows: 1800 m > 2000 m > 2200 m > 2400 m > 1600 m > 2600 m. Therefore, the warning lengths of 1800 m, 2000 m, and 2200 m were selected as the control variables for driving simulation experiments. The transition area's lengths were ranked from best to worst, as follows: 140 m > 120 m > 160 m > 100 m > 180 m > 200 m. Therefore, the transition lengths of 120 m, 140 m, and 160 m were selected as the control variables for driving simulation experiments.

Apart from the warning and transition lengths, all other elements in the experimental scenarios remained consistent. The experimental road was a two-way, six-lane urban expressway with a total length of approximately 8 km and a single-lane width of 3.5 m. The design speed in the basic section was 120 km/h, and the speed limit was 60 km/h. The work zone was arranged following SAC [1], and the work conditions were set for closed outer lane operation (Figure 2).

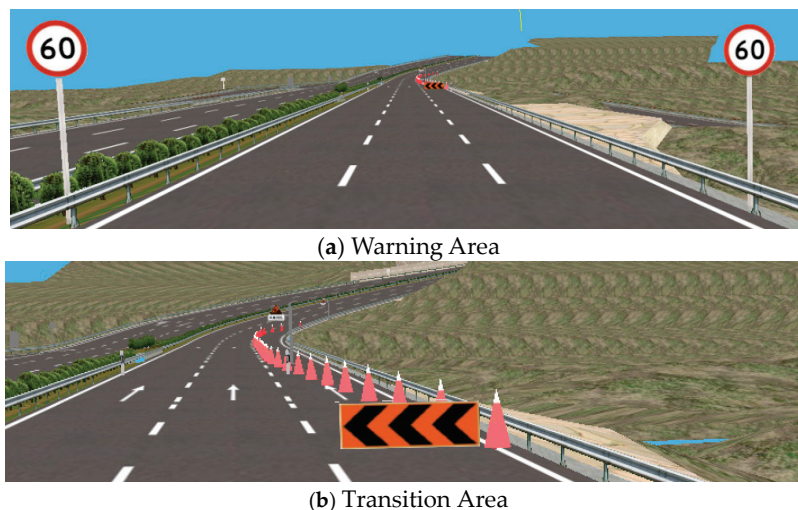


Figure 2. Simulation scenes of the work zone on a two-way, six-lane urban expressway.

Considering the influence of traffic flow, a steady flow of 2000 pcu/h was selected as the traffic volume loaded in this experiment. Therefore, this study used the warning and transition lengths as the control variables. The experimental design involved three warning lengths (1800 m, 2000 m, and 2200 m) and three transition lengths (120 m, 140 m, and 160 m). The differences in the nine experimental scenarios are shown in Table 1.

Table 1. The experimental scenarios.

Scenario	Warning Length	Transition Length
A	1800 m	120 m
B	1800 m	140 m
C	1800 m	160 m
D	2000 m	120 m
E	2000 m	140 m
F	2000 m	160 m
J	2200 m	120 m
H	2200 m	140 m
I	2200 m	160 m

2.2. Participants

The minimum sample size required was determined based on the expected variance, target confidence, and error magnitude, as follows:

$$N = \frac{Z^2 \sigma^2}{E^2}, \quad (1)$$

where N = experimental sample size, Z = standard normal distribution statistic, σ = standard deviation (SD; 0.25 to 0.5), and E is the maximum error. This experiment used the following variables: $Z = 1.96$ (95% confidence level), $\sigma = 0.3$, and $E = 10\%$ [41]. The minimum sample was calculated to be 35.

In this experiment, 45 participants (25 males and 20 females) with valid driving licenses were recruited to ensure a sufficient sample size after excluding invalid data. Their ages ranged from 20 to 28 years ($M = 23.9$, $SD = 1.8$). The driving experience of participants ranged from 1 to 6 years ($M = 1.6$, $SD = 1.9$). The participants were self-reported to be in good physical condition with normal visual function and normal or corrected visual acuity of 1.0 or above. All participants had driving experience on urban expressways and were unfamiliar with the driving simulation scenarios of this experiment. All participants were aware of this experiment's procedures and had signed an informed consent form. All participants were guaranteed to have a regular diet and rest for 24 h before the experiment, without drinking alcohol or taking drugs or strenuous exercise.

2.3. Performance Indicators

Relevant studies have shown that the driver's mental and physiological characteristics and driving behavior during the driving process are the most critical factors affecting driving safety. The data on the driver's eye movement, EEG, and driving behavior could reflect the driver's safety and comfort levels during the driving process. This experiment collected three types of indicators: visual effect (eye movement) indicators (U_1), cognitive effect (EEG (U_2)) indicators, and driving effect (manipulation behavior) indicators (U_3).

2.3.1. Visual Effect Indicators

Pupil area (U_{11} , pixel): The pupil area reflected the driver's comfort level with vision. The pupils will dilate when the visual effect is poor, and the pupils will constrict when the visual effect is good. Meanwhile, the pupil area will also change during the operation, and the magnitude of the change is closely related to the difficulty of the operation [42].

Saccade frequency (U_{12} , times/s): Saccade is a quick movement of the eyes from one point to another with a viewpoint dwell time of less than 120 ms. Saccade frequency is calculated as the number of saccades occurring per unit time. The saccade frequency is lower when the drivers are familiar with the driving environment, or the road environment is simple. The saccade frequency is higher when the driver is unfamiliar with the driving environment or the road environment is complex, which indicates the challenge of searching for information [43].

2.3.2. Cognitive Effect Indicators

Absolute power of α wave (U_{21} , μV^2): This wave will increase when the drivers are relaxed and their attention is focused. At this time, the driver's cognitive load is moderate, allowing them to deal with road emergencies more calmly, which is more conducive to driving safety [44].

Absolute power of β wave (U_{22} , μV^2): The beta waves generally appear when people are stimulated by external events, with intense alertness and high cognitive load. The continuous increase in the absolute power of β waves often indicates that people gradually

change from alertness to tension, causing drivers' fatigue and reducing driving safety and comfort [45].

θ/β (U_{23}): Since a single EEG energy value can only represent the drivers' physiological changes in a single frequency band, scholars often combine the EEG energy ratios (α/β and θ/β , etc.) to objectively and comprehensively analyze the driver's driving status [44–46]. To avoid repeated characterization of EEG signals, we selected low-frequency θ waves and high-frequency β waves with low correlation for the ratio calculation. When the theta wave increases, the brain is relaxed, indicating that a person's mental stress has been relieved. Therefore, the θ/β value can characterize the change in cognitive level. The larger the θ/β value is, the more relaxed the driver is. At this time, the driver's cognitive load is moderate, allowing them to handle road emergencies more calmly, which is more conducive to driving safety [44,47,48].

2.3.3. Driving Effect Indicators

Longitudinal acceleration (U_{31} , m/s^2): It refers to the change in speed per unit time during the braking process of the vehicle, which can intuitively reflect the operational stability of the drivers during speed changes. Rapid acceleration or deceleration behavior is considered relatively unsafe driving behavior [49]. The greater the fluctuation in acceleration is, the greater the risk of driving.

Instantaneous speed of entering the work area (U_{32} , km/h): This refers to the vehicle speed in the normal direction when it enters the work zone. To some extent, it reflects the rationality of the warning and transition areas' lengths and the speed control effect of the signs and markings [50].

Lane change duration (U_{33} , s): It refers to the difference between the time the vehicle travels to the centerline of the target lane and the time when the wheels first press the lane line. The duration of a lane change can reflect its difficulty to some extent [51].

2.4. Equipment

This study collected data on eye movement, EEG, and driving behavior during simulation driving, using corresponding data acquisition systems. The experiment was conducted indoors, controlling the effects of climate, light, and noise. Three types of equipment were used in this study.

2.4.1. Driving Simulator

The driving simulation experiment was conducted using the DSR-1000TS2.0 automobile driving simulation system (Kunming University of Science and Technology Science & Technology Industrial Management Company, Kunming, China), which included the cockpit, 120° ring-view display system, and the independent console (Figure 3). The cockpit included individual car seats and vehicle-mounted devices (steering wheel, wing mirrors, and pedals). The 120° ring-view display system could display driving scenarios for the participants. The independent console could record and export the experiment's driving behavior and vehicle dynamics parameters, such as speed, longitudinal acceleration, and lane change duration. The system featured an auxiliary sound system that realizes the sound of the experimental environment and the vehicle.

2.4.2. Eye Tracker and D-Lab Software

The Dikablis eye-tracking system (Ergoneers GmbH, Geretsried, Germany) was used to track and measure the drivers' eye movement features during the simulation experiment, working with the D-Lab analysis software (version 3.55, Ergoneers GmbH, Geretsried, Germany). The device could track the drivers' sights, and the opening and closing states of their eyes in real time, recording the changes in eye movement parameters, such as pupil

area, gaze duration, and saccade frequency. The eye-tracking device collected samples at a frequency of 60 Hz with an accuracy range of 0.1° to 0.3° .



Figure 3. DSR-1000TS2.0 driving simulation system.

The D-Lab software uses the Dikablis eye tracker to track and measure the driver's eye movement characteristics. The D-Lab platform creates a time segment based on the scenario video playback that needs to be analyzed, called a "task segment", and selects the eye movement data items required in the "task segment".

2.4.2.1. 32-Channel NE Wireless EEG System

The 32-channel NE wireless EEG (Neuroscan, Charlotte, NC, USA) collected the drivers' EEG data during the simulation experiment. The equipment had 32 experimental channels, each with a bandwidth of 0 to 250 Hz. The sampling rate was 500 SPS, the resolution was 24 bits, and the noise was less than 1 μ VRMS (0~250 Hz). It achieved EEG data transmission through wireless technology and obtained EEG signals, such as α , β , and θ waves.

2.5. Procedure

The participants were first asked to sign an experimental consent form and complete a personal information questionnaire, which included name, gender, age, number of years of driving, and mental state on that day. Before the formal experiment, the participants read the experimental instructions and understood the experimental tasks. Then, the participants travelled for a few minutes on a non-experimental scenario to familiarize themselves with the operation of the driving simulation cabin. After the participant was familiar with the operation, the experimenter assisted them in wearing EEG caps and Dikablis Glasses and calibrated them. If the participants felt uncomfortable, they could leave the experiment at any time.

The experimenter created the experiment number information in the D-Lab and started the formal experiment. To avoid learning effects, the experimenter randomly presented the experimental scenarios. According to the experimental requirements, the participants entered the section at a speed of 100 km/h. The participants changed their driving behavior and controlled their speed according to the sign prompts. The experimenters needed to confirm whether the participants' EEG signals and eye movement images were normal during the experiment. After each scenario, the participants had a 5 min break before completing the nine-scenario experiments.

The above experimental process was repeated until all 45 participants had completed the experiment. At the end of the experiment, each participant was paid 50 yen as a reward. If a participant had an accident in the experimental scenario, such as crashing into another vehicle or going off the road, they were asked to stop and rerun the experiment.

2.6. Data Pre-Processing

The raw eye movement data were imported into D-Lab software. Then, the eye movement indicator data, such as pupil area and saccade frequency, were extracted based on the determined area of interest. Finally, the driving behavior data were output from the vehicle dynamics module, and the collected EEG data were exported to MATLAB software (<https://www.mathworks.cn/products/matlab.html>) for pre-analysis.

3. Methodology

The fuzzy method is a highly effective multi-factor tool for comprehensively evaluating things affected by multiple factors. It quantifies some factors that are not easily measurable and systematically reflects the strengths and weaknesses of the evaluation object from various perspectives and levels, transforming the problems into a quantitative evaluation. Calculating the weights is the key step of the fuzzy comprehensive evaluation. A combination of the analytic network process and entropy weighting method (ANP-EWM) was used to calculate comprehensive weights to overcome the shortcomings of subjective and objective weighting methods. This method not only refers to experience for weighting but also avoids excessive interference from subjective factors and overcomes the lack of objectivity in the weight determination of traditional fuzzy evaluation methods.

3.1. Combined Subjective and Objective Weights

The ANP and EWM were combined to determine the combined subjective and objective weights. The procedure involved the following steps:

Step 1: Determine the subjective weights using ANP

The ANP is a commonly used subjective weighting method in which experts measure the relative importance between various evaluation indicators and score them on a 1–9 scale. Then, the experimenter obtained judgment matrices and conducted consistency tests. Finally, the experimenter obtained the subjective weights (w_{cj}) of the indicators and the subjective weight vector $Wc = [w_{c1}, w_{c2}, \dots, w_{cn}]$ [52].

Step 2: Determine the objective weights using EWM [53,54]

Using the entropy weight method to determine the driver's response evaluation index system's weight can objectively and realistically reflect the evaluation effect. It is necessary to normalize the original data to eliminate inconsistencies in units and dimensions between indicator data. Therefore, the min–max normalization method was used to normalize the data and obtain the normalized decision matrix B , as follows:

$$B_{m \times n} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix} = b_{ij} (0 \leq i \leq m, 0 \leq j \leq n), \quad (2)$$

where m = the number of evaluation objects, n = the number of evaluation indicators, and b_{ij} = the normalized value of the j -th evaluation indicator for the i -th evaluation object.

According to the normalized decision matrix, the entropy value of the j -th indicator in the evaluation index system is given by:

$$e_j = -\frac{1}{\ln(m)} \sum_{i=1}^m f_{ij} \ln(f_{ij}), \quad (3)$$

$$f_{ij} = \frac{A + b_{ij}}{\sum_{i=1}^m (A + b_{ij})}, \quad (4)$$

where A is the translation amount (0.00001). It is necessary to add a translation amount to the value of 0 after standardization to obtain a small positive value, ensuring the completeness and validity of the data, and making the logarithm in Equation (3) meaningful.

The objective weights for indicators are given by:

$$w_{pj} = \frac{1 - e_j}{n - \sum_{j=1}^n e_j}. \quad (5)$$

Step 3: Combination weights method [55]

The subjective and objective weights were combined and assigned as follows:

$$w_j = \frac{w_{cj} \cdot w_{pj}}{\sum_{j=1}^n (w_{cj} \cdot w_{pj})}, \quad (6)$$

where w_j = the combined weight for the j -th indicator, w_{cj} = the subjective weight for the j -th indicator, and w_{pj} = the objective weight for the j -th indicator.

3.2. Ranking Using Fuzzy Evaluation Method [56,57]

The specific steps of the improved fuzzy comprehensive evaluation method based on the combined weights were as follows:

(a) Determine the factor set U and evaluation set V : The factor set is a set of elements composed of various factors affecting the evaluation object, represented by $U = (u_1, u_2, \dots, u_m)$. The evaluation set is a set composed of various results that the evaluator may make on the evaluation object, represented by $V = (v_1, v_2, \dots, v_n)$. Generally, the evaluation set was divided into five levels, and each was regarded as a fuzzy subset.

(b) Determine the membership degree: The membership degree of each factor was determined using a trapezoidal membership degree function, divided into three segments: partial-small, intermediate, and partial-large (Figure 4). For the partial-small segment (a to b), the membership degree increased from 0 at a to 1 at b. For the intermediate segment (b to c), the membership degree was 1. For the partial-large segment (c to d), the membership degree decreased from 1 at c to 0 at d. These segments represent the range where the indicator value was considered small, intermediate, and large.

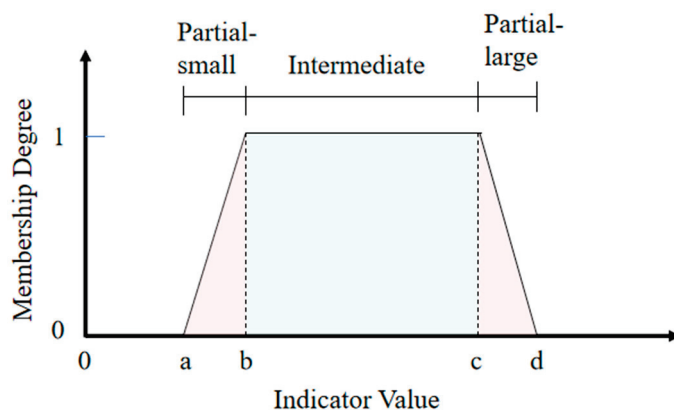


Figure 4. Illustration of the trapezoidal membership degree function.

(c) Construct the membership matrix R : The evaluation indicators were processed and normalized to the $[0, 1]$ interval using the min–max normalization method to eliminate the inconsistency of units and dimensions between indicators. Then, the membership degrees

of the factor set were determined at each evaluation level, and the membership degree matrices, $R_{m \times n}$, were constructed as follows:

$$R_{m \times n} = \begin{pmatrix} r_{11} & \dots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \dots & r_{mn} \end{pmatrix} = r_{ij} (0 \leq i \leq m, 0 \leq j \leq n), \quad (7)$$

where r_{ij} is the membership degree of the i -th element in the factor set U at the j -th level in the evaluation set V . The ANP-EWM method was applied to determine the weight vector $W = (w_1, w_2, \dots, w_m)$.

(d) Obtain the fuzzy comprehensive evaluation result vector S : The weight vector W and the membership matrix R were synthesized to obtain each evaluation object's fuzzy comprehensive evaluation vector S , as follows:

$$S = W \circ R_{m \times n} = (w_1, w_2, \dots, w_m) \begin{pmatrix} r_{11} & \dots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \dots & r_{mn} \end{pmatrix} = (s_1, s_2, \dots, s_n), \quad (8)$$

where s_i is the membership degree of the evaluation object at the j -th level in the evaluation set V .

(e) Analyze the evaluation results: The weighted average method was used to process the membership degree to calculate the comprehensive evaluation grade eigenvalue G , as follows [56]:

$$G = \frac{\sum_{i=1}^m j \times s_i^k}{\sum_{i=1}^m s_i^k} \quad (9)$$

where k is the undetermined coefficient, taken equal to 2 in this study [57].

Unlike the ineffective maximum membership degree of the traditional fuzzy comprehensive evaluation, the presented (improved) comprehensive evaluation index is more efficient and more intuitively reflects the specific results of the evaluation object.

4. Analysis and Results

4.1. Statistical Analysis

The participants' eye movement, EEG, and driving behavior in each scenario were obtained through driving simulation experiments. After excluding invalid data (such as Dikablis Glasses not calibrated, missing data, and incomplete driving tasks), the data of 40 subjects were retained. A two-way repeated measures analysis of variance (ANOVA) was used to analyze the influence of three warning lengths (1800 m, 2000 m, and 2200 m) and three transition lengths (120 m, 140 m, and 160 m) and their interactions on the participants' visual effects, cognitive effects, and driving effects at a significance level (α) of 0.05. Before the ANOVA was performed, the normality of the data was checked using the Shapiro–Wilk method, which indicated that the data in each group passed the normal distribution test and had no outliers [53–58].

In this experiment, the IBM SPSS 27.0 Statistics software was used for the two-way repeated measures ANOVA. If the data did not meet Mauchly's sphericity test ($p < 0.05$), they were corrected using the Greenhouse–Geisser method. The average and test results of participants' visual, cognitive, and driving behavior effects are shown in Tables 2–4. When the main effect existed, post hoc multiple comparisons were performed. When the interactive effect existed, simple effects analyses were performed.

Table 2. Results of two-way repeated measures ANOVA of participants' visual effects.

Evaluation Indicator	Work Zone Element	Length	Mean	SE	95% CI of the Difference	F	Sig.	Partial η^2
Pupil area	Warning length	1800 m	733.363	28.089	(676.547, 790.178)	39.050	<0.001	0.500
		2000 m	590.179	22.716	(544.233, 636.126)			
		2200 m	540.127	21.933	(495.764, 584.490)			
	Transition length	120 m	679.476	27.022	(624.820, 734.133)	13.601	<0.001	0.259
		140 m	621.293	24.094	(572.558, 670.028)			
		160 m	562.900	21.443	(519.527, 606.273)			
	Warning length*Transition length					1.850	0.136	0.045
Saccade frequency	Warning length	1800 m	1.029	0.052	(0.923, 1.135)	23.722	<0.001	0.378
		2000 m	0.763	0.036	(0.690, 0.836)			
		2200 m	0.708	0.037	(0.633, 0.783)			
	Transition length	120 m	1.091	0.047	(0.996, 1.186)	40.813	<0.001	0.511
		140 m	0.784	0.039	(0.705, 0.864)			
		160 m	0.625	0.044	(0.536, 0.715)			
	Warning length*Transition length					0.928	0.429	0.023

Table 3. Results of two-way repeated measures ANOVA of participants' cognitive effects.

Evaluation Indicator	Work Zone Element	Length	Mean	SE	95% CI of the Difference	F	Sig.	Partial η^2
Absolute power of α	Warning length	1800 m	718.732	45.998	(625.693, 811.771)	83.756	<0.001	0.682
		2000 m	514.580	30.440	(453.008, 576.152)			
		2200 m	385.020	24.433	(335.599, 434.441)			
	Transition length	120 m	586.494	37.016	(511.622, 661.366)	13.034	<0.001	0.250
		140 m	544.993	33.278	(477.683, 612.303)			
		160 m	486.846	29.503	(427.171, 546.520)			
	Warning length*Transition length					15.863	<0.001	0.289
Absolute power of β	Warning length	1800 m	362.532	12.936	(336.366, 388.697)	49.972	<0.001	0.562
		2000 m	457.846	19.254	(418.901, 496.791)			
		2200 m	606.741	21.826	(562.594, 650.888)			
	UT length	120 m	495.299	19.626	(455.601, 534.997)	2.149	0.123	0.052
		140 m	447.988	16.894	(413.816, 482.160)			
		160 m	483.831	17.448	(448.539, 519.124)			
	Warning length*Transition length					6.104	<0.001	0.135
θ/β	Warning length	1800 m	1.216	0.017	(1.181, 1.251)	281.991	<0.001	0.879
		2000 m	0.856	0.013	(0.830, 0.882)			
		2200 m	0.748	0.010	(0.727, 0.769)			
	Transition length	120 m	1.053	0.013	(1.027, 1.079)	99.684	<0.001	0.719
		140 m	0.923	0.011	(0.900, 0.946)			
		160 m	0.844	0.008	(0.828, 0.861)			
	Warning length*Transition length					33.057	<0.001	0.459

Table 4. Results of two-way repeated measures ANOVA of participants' driving effects.

Evaluation Indicators	Work Zone Element	Length	Mean	SE	95% CI of the Difference	F	Sig.	Partial η^2
Longitudinal acceleration	Warning length	1800 m	−2.124	0.021	(−2.167, −2.080)	1162.637	<0.001	0.968
		2000 m	−1.381	0.019	(−1.420, −1.342)			
		2200 m	−0.833	0.017	(−0.867, −0.799)			
	Transition length	120 m	−1.760	0.020	(−1.801, −1.719)	302.734	<0.001	0.886
		140 m	−1.471	0.019	(−1.510, −1.433)			
		160 m	−1.106	0.018	(−1.144, −1.069)			

Table 4. Cont.

Evaluation Indicators	Work Zone Element	Length	Mean	SE	95% CI of the Difference	F	Sig.	Partial η^2
Warning length*Transition length						27.145	<0.001	0.410
Instantaneous speed of entering the work area	Warning length	1800 m	47.025	0.346	(46.326, 47.725)	110.153	<0.001	0.739
		2000 m	43.620	0.325	(42.963, 44.277)			
		2200 m	40.806	0.233	(40.334, 41.278)			
	Transition length	120 m	46.189	0.326	(45.529, 46.850)	63.379	<0.001	0.619
		140 m	44.050	0.320	(43.402, 44.697)			
		160 m	41.212	0.302	(40.601, 41.823)			
Warning length*Transition length						2.673	0.034	0.064
Lane change duration	Warning length	1800 m	9.587	0.199	(9.184, 9.990)	20.195	<0.001	0.341
		2000 m	8.653	0.149	(8.353, 8.954)			
		2200 m	8.013	0.153	(7.703, 8.323)			
	Transition length	120 m	9.547	0.192	(9.158, 9.936)	20.086	<0.001	0.340
		140 m	8.827	0.153	(8.518, 9.135)			
		160 m	7.880	0.181	(7.515, 8.245)			
Warning length*Transition length						0.056	0.994	0.001

4.2. Visual Effect Results

4.2.1. Pupil Area (U_{11} , Pixel)

The pupil areas of participants were collected. The distribution of participants' pupil areas in different experimental scenarios is shown in Figure 5a. As the warning length increased, the participants' pupil areas gradually decreased and showed a convergence trend. As the transition length increased, the participants' pupil areas gradually decreased and showed a convergence trend. This showed that adequate warning and transition lengths can improve drivers' visual comfort.

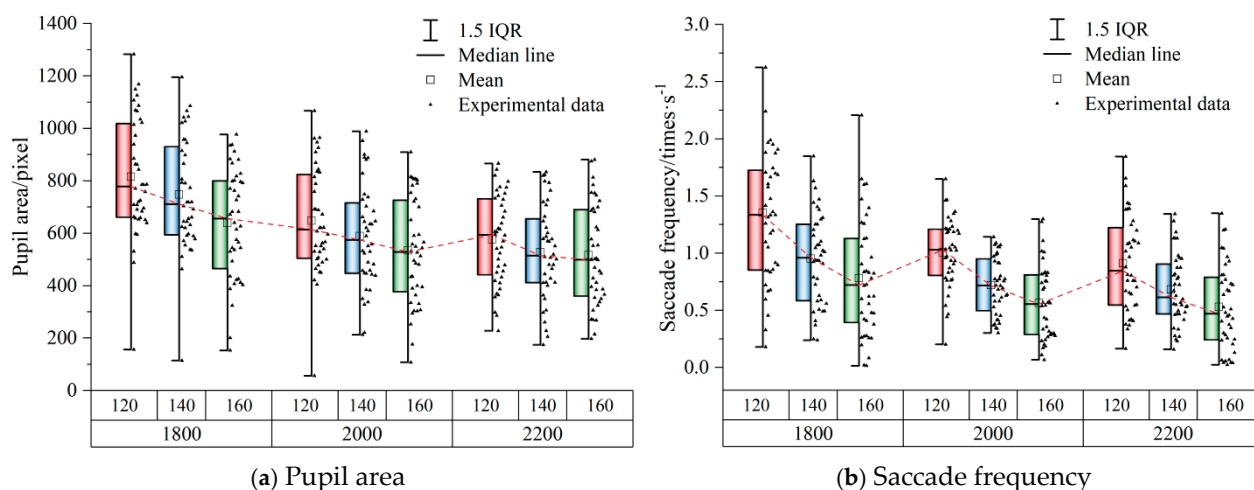


Figure 5. Comparison of eye movement indicators under different conditions.

Table 2 shows that the main effects of warning and transition lengths on pupil areas were statistically significant. Still, the interaction between warning and transition lengths in pupil areas was not statistically significant. Therefore, post hoc multiple comparisons were performed for within-group effects where significant differences existed, as shown in Table 5.

Table 5. Post hoc multiple comparisons tests of participants' pupil areas (pixel).

Warning or Transition Length	Warning or Transition Length	Mean Difference (I–J)	SE	Sig.	95% CI of the Difference
Warning length (I)	Warning length (J)				
1800 m	2000 m	143.184 *	23.453	<0.001	(84.511, 201.856)
	2200 m	193.236 *	26.709	<0.001	(126.420, 260.052)
2000 m	2200 m	50.052 *	16.791	0.015	(8.047, 92.058)
Transition length (I)	Transition length (J)				
	140 m	58.183 *	15.337	0.002	(19.816, 96.550)
120 m	160 m	116.577 *	26.286	<0.001	(50.818, 182.335)
140 m	160 m	58.393	23.931	0.058	(−1.473, 118.259)

* The significance level of the mean value difference was 0.05.

Table 5 shows that the pupil areas between different warning lengths all showed significant differences, and the pupil area was most prominent when the warning length was 1800 m, indicating that the driver's visual comfort was the worst. When the transition length was 120 m, the pupil areas differed significantly from the other transition lengths. The pupil areas showed no significant difference between the 140 m and 160 m transition lengths, indicating that participants' visual comfort was equivalent in these two transition lengths.

4.2.2. Saccade Frequency Results

The saccade frequencies of participants were collected. The distribution of participants' saccade frequencies in different experimental scenarios is shown in Figure 5b. As the warning length increased, the participants' saccade frequency decreased, and the changing trend of different transition lengths was similar. As the transition length increased, the participants' saccade frequency decreased, and the changing trend of different warning lengths was similar. When the transition length was 120 m, the participants showed frequent saccade behavior, and the saccade frequency was highest when the warning length was 1800 m.

As shown in Table 2, the main effects of both warning and transition lengths on saccade frequency were statistically significant. Still, the interaction between warning and transition lengths on saccade frequency was not statistically significant. Therefore, post hoc multiple comparisons were performed for within-group effects where significant differences existed, as shown in Table 6.

Table 6. Post hoc multiple comparisons tests of participants' saccade frequency (times/s).

Warning or Transition Length	Warning or Transition Length	Mean Difference (I–J)	SE	Sig.	95% CI for Difference
Warning length (I)	Warning length (J)				
1800 m	2000 m	0.266 *	0.057	<0.001	(0.122, 0.410)
	2200 m	0.321 *	0.052	<0.001	(0.190, 0.452)
2000 m	2200 m	0.055	0.038	0.459	(−0.040, 0.150)
Transition length (I)	Transition length (J)				
	140 m	0.306 *	0.049	<0.001	(0.183, 0.429)
120 m	160 m	0.465 *	0.053	<0.001	(0.332, 0.598)
140 m	160 m	0.159 *	0.055	0.018	(0.023, 0.295)

* The significance level of the mean value difference was 0.05.

Table 6 shows that the saccade frequency was the highest when the warning length was 1800 m, indicating that drivers had a high search demand for the surrounding road environment and poorer visual stability. The saccade frequency showed no significant difference between 2000 m and 2200 m warning lengths. The saccade frequencies between different transition lengths all showed significant differences, and the saccade frequency was most prominent when the transition length was 120 m, indicating that the driver's visual stability was the worst.

4.3. Cognitive Effect Results

4.3.1. Absolute Power of α Wave

The absolute power of the α wave of participants was measured. The distribution of this indicator across different experimental scenarios is shown in Figure 6a. As the warning length increased, this indicator decreased gradually. This indicated that the drivers' cognitive load and driving pressure increased gradually as the warning length increased, which was not conducive to driving safety. However, the difference in the absolute power of the α wave under different transition lengths gradually decreased as the warning length increased.

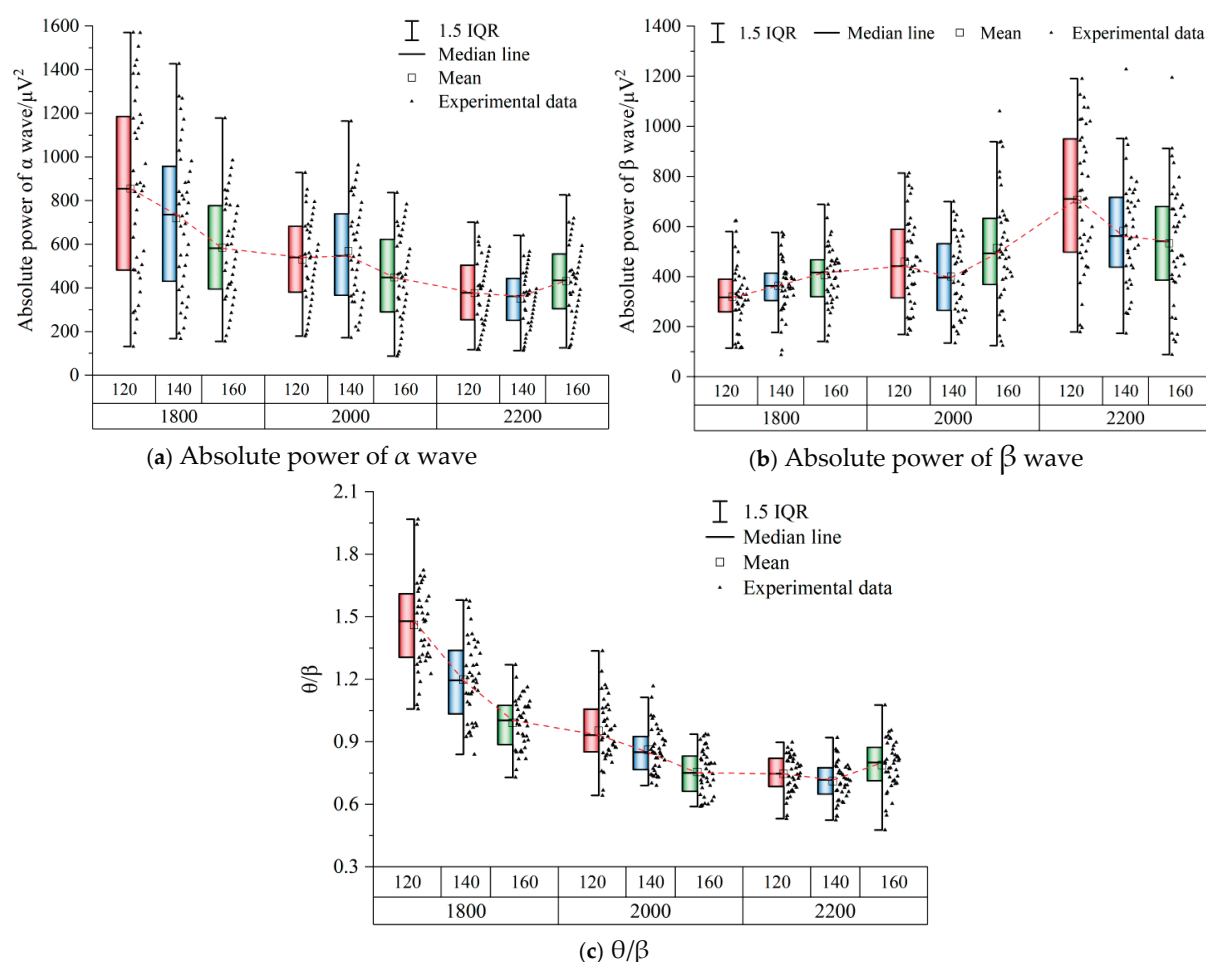


Figure 6. Comparison of EEG indicators under different conditions.

As shown in Table 3, the main effects of the warning and transition lengths and the interaction effect on the absolute power of the α wave were statistically significant. Therefore, simple effects comparisons were performed for the warning and transition lengths separately.

The pairwise comparison revealed that when the warning length was 2000 m, there was no significant difference in the absolute power of the α wave between the transition lengths of 120 m and 140 m ($M(120-140) = -38.947$, $SE = 31.974$, $p = 0.692$). In contrast, the pairwise comparisons showed significant differences between other transition lengths. When the transition length was 160 m, there was no significant difference in the absolute power of the α wave between the warning lengths of 2000 m and 2200 m ($M(2000-2200) = 16.397$, $SE = 9.642$, $p = 0.291$), while the pairwise comparisons showed significant differences between other warning lengths. When the warning length was 1800 m, the absolute power of the driver's α wave was the highest, so the drivers' cognitive load was the lowest.

4.3.2. Absolute Power of β Wave

The absolute power of the β wave of participants was collected. The distribution of participants' absolute power of the β waves in different experimental scenarios is shown in Figure 6b. As the warning length increased, the participants' absolute power of the β wave showed an increasing trend. This showed that the drivers' cognitive load, driving tension, and pressure gradually increased as the warning length increased, which is not conducive to driving safety. When the warning length was 2200 m, the transition length was 120 m, and the participants' absolute power value of the β wave was the highest.

As shown in Table 3, the main effects of the warning and transition lengths and the interaction effect on the absolute power of the β wave were statistically significant. Therefore, simple effects comparisons were performed for the warning and transition lengths separately.

The pairwise comparison revealed that when the warning lengths were 1800 m and 2200 m, there were significant differences in the absolute power of the β wave between the transition lengths of 120 m and 160 m ($M(120-160) = -85.444$, $SE = 28.048$, $p = 0.012$; $M(120-160) = 174.706$, $SE = 59.965$, $p = 0.018$). When the warning length was 2000 m, there were significant differences in the absolute power of the β wave between the transition lengths of 140 m and 160 m ($M(140-160) = -114.461$, $SE = 41.052$, $p = 0.024$). When the transition length was 140 m, there was no significant difference in the absolute power of the β wave between the warning lengths of 1800 m and 2000 m ($M(1800-2000) = -37.154$, $SE = 25.897$, $p = 0.478$). When the transition length was 160 m, there was no significant difference in the absolute power of the β wave between the lengths of 2000 m and 2200 m ($M(2000-2200) = -17.738$, $SE = 50.765$, $p = 1.000$), while the pairwise comparisons showed significant differences between other warning lengths.

4.3.2.1. θ/β Ratio

The θ/β values of participants were calculated. The distribution of the participants' θ/β values across different experimental scenarios is shown in Figure 6c. As the warning length increased, the participants' θ/β values gradually decreased. This showed that the drivers' cognitive load and driving tension gradually increased as the warning length increased, which is not conducive to driving safety. However, the difference in the θ/β values under different transition lengths gradually decreased as the warning length increased. When the warning length was 2200 m, the θ/β values under different transition lengths were roughly the same, indicating that the drivers' cognitive load and tension were comparable.

As shown in Table 3, the main effects of the warning and transition lengths and the interaction effect on the θ/β values were statistically significant. Therefore, simple effects comparisons were performed for the warning and transition lengths separately.

The pairwise comparison revealed that when the warning length was 2200 m, there were no significant differences in the θ/β values between the transition length of 120 m and the other transition lengths ($M(120-140) = 0.035$, $SE = 0.018$, $p = 0.160$; $M(120-160) = -0.042$, $SE = 0.022$, $p = 0.186$), while the pairwise comparisons showed significant differences between other transition lengths. When the transition length was 160 m, there was no significant difference in θ/β values between the warning lengths of 2000 m and 2200 m ($M(2000-2200) = -0.034$, $SE = 0.029$, $p = 0.737$), while the pairwise comparisons showed significant differences between other warning lengths. When the warning length was 1800 m, the drivers' θ/β values were the highest, so the drivers' cognitive load was the lowest.

4.4. Driving Effect Results

4.4.1. Longitudinal Acceleration

The longitudinal acceleration of participants was collected. The distribution of participants' longitudinal acceleration in different experimental scenarios is shown in Figure 7a. As the warning length increased, the participants' longitudinal acceleration showed an increasing trend and gradually converged. As the transition length increased, the participants' longitudinal acceleration showed an increasing trend. When the warning length was 1800 m, the absolute value of the participants' longitudinal acceleration was larger, indicating that the drivers had a large operating load and were eager to brake. It may be due to the shorter warning length, which shortened the deceleration distance.

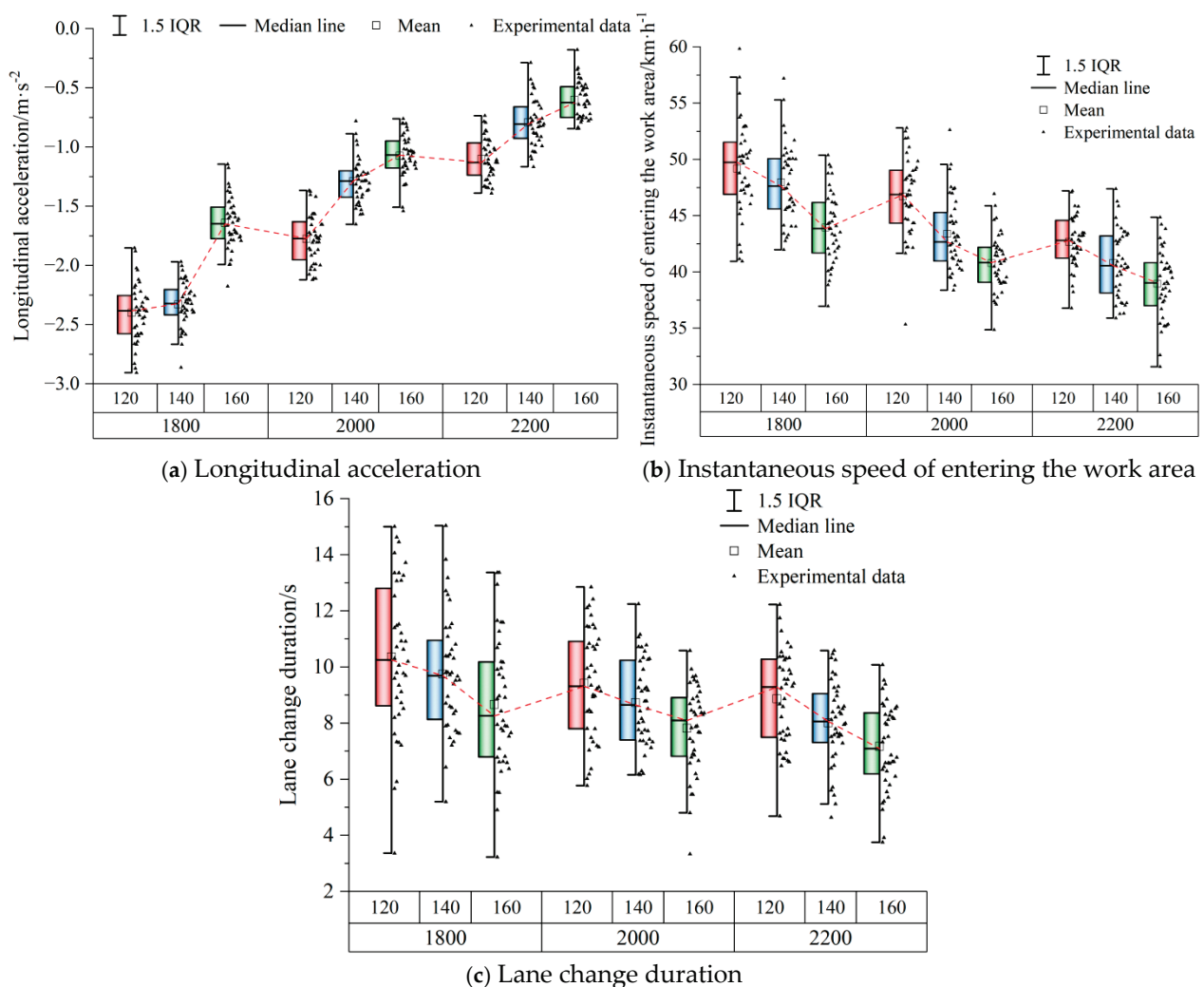


Figure 7. Comparison of driving behavior indicators under different conditions.

As shown in Table 4, the main effects of the warning and transition lengths and the interaction effect on the longitudinal acceleration were statistically significant. Therefore, simple effects comparisons were performed for the warning and transition lengths separately.

The pairwise comparison revealed that when the warning length was 1800 m, there were no significant differences in the longitudinal acceleration between the transition lengths of 120 m and 140 m ($M(120-140) = -0.072$, $SE = 0.05$, $p = 0.464$). In contrast, the pairwise comparisons showed significant differences between other transition lengths. For any given transition length, the differences in the longitudinal acceleration among different warning lengths were significant. When the warning length was 2200 m, the absolute value of the drivers' longitudinal acceleration was the smallest, indicating that the drivers slowed down gently.

4.4.2. Instantaneous Speed of Entering the Work Area

The participants' instantaneous speed of entering the work area was collected. The distribution of participants' instantaneous speed of entering the work area in different experimental scenarios is shown in Figure 7b. In the experiment, the speed limit of the work area was 60 km/h. Before entering the work area, all participants took deceleration measures, with an average speed between 40 km/h and 50 km/h, which met the speed limit requirements of the work zone. As the warning length increased, the participants' instantaneous speed of entering the work area gradually decreased. Also, as the transition length increased, the participants' instantaneous speed of entering the work area gradually decreased. It indicated that the longer the warning and transition areas, the longer the deceleration area could be provided for participants, which could reduce the deceleration pressure of the drivers.

As shown in Table 4, the main effects of the warning and transition lengths and the interaction effect on the instantaneous speed of entering the work area were statistically significant. Therefore, simple effects comparisons were performed for the warning and transition lengths one by one.

The pairwise comparison revealed that when the warning length was 1800 m, there were no significant differences in the instantaneous speed between the transition lengths of 120 m and 140 m ($M(120-140) = 1.212$, $SE = 0.788$, $p = 0.396$). When the warning length was 2200 m, there were no significant differences in the instantaneous speed between the transition lengths of 140 m and 160 m ($M(120-140) = 1.872$, $SE = 0.762$, $p = 0.056$). In contrast, the pairwise comparisons showed significant differences between other transition lengths. For any given transition length, the differences in the instantaneous speed among different warning lengths were significant.

4.4.3. Lane Change Duration

The lane change duration of participants was collected. The distribution of the participants' lane change durations in different experimental scenarios is shown in Figure 7c. As the warning length increased, the participants' lane change duration gradually decreased, and the change trend of different transition lengths was similar. As the transition length increased, the participants' lane change duration steadily reduced, and the changing trend of different warning lengths was similar. The shorter warning or transition length may increase the cross-section traffic volume, and then the space for drivers to change lanes is compressed, making lane changing more difficult.

As shown in Table 4, the main effects of warning and transition lengths on lane change duration were statistically significant. Still, the interaction between the warning and transition lengths on lane change duration was not statistically significant. Therefore,

post hoc multiple comparisons were performed for within-group effects where significant differences existed, as shown in Table 7.

Table 7. Post hoc multiple comparisons tests of participants' lane change duration (s).

Warning or Transition Length	Warning or Transition Length	Mean Difference (I–J)	SE	Sig.	95% CI for the Difference
Warning length (I)	Warning length (J)				
1800 m	2000 m	0.933 *	0.234	<0.001	(0.349, 1.518)
	2200 m	1.573 *	0.264	<0.001	(0.913, 2.233)
2000 m	2200 m	0.640 *	0.249	0.042	(0.018, 1.262)
Transition length (I)	Transition length (J)				
120 m	140 m	0.720 *	0.278	0.040	(0.026, 1.414)
	160 m	1.667 *	0.260	<0.001	(1.016, 2.318)
140 m	160 m	0.947 *	0.253	0.002	(0.314, 1.579)

* The significance level of the mean value difference was 0.05.

As shown in Table 7, all lane change durations between different warning lengths exhibited significant differences, with the lane change duration being most pronounced when the warning length was 1800 m, indicating that lane changes were the most difficult. Driving safety was the worst for the drivers. Similarly, all lane change durations between different transition lengths showed significant differences, with the lane change duration being most pronounced when the transition length was 120 m, indicating that lane changes were the most difficult. Driving safety was also the worst for the drivers.

4.5. Results of Weights and Ranking Quality

4.5.1. Comprehensive Weights of Indicators

Seven experts in relevant fields were invited to assess the relative importance of various evaluation indicators on a scale of 1 to 9. The subjective weight of each evaluation indicator was calculated using the Super Decision software (version 8.5.6) using the ANP method. Using the entropy weight method, the objective weight was calculated according to Equations (2)–(5) based on the mean values of various evaluation indicators obtained from the driving simulation experiment. The comprehensive weights of the indicators were calculated from Equation (6), as shown in Table 8.

Table 8. Comprehensive weights of safety and comfort evaluation indicator system (U).

Primary Indicators	Subjective Weights	Objective Weights	Comprehensive Weights	Secondary Indicators	Subjective Weights	Objective Weights	Comprehensive Weights
Visual effects (U ₁)	0.317	0.161	0.161	Pupil area (U ₁₁)	0.634	0.535	0.666
				Saccade frequency (U ₁₂)	0.366	0.465	0.334
				Absolute power of α wave (U ₂₁)	0.304	0.348	0.318
Cognitive effects (U ₂)	0.203	0.489	0.312	Absolute power of β wave (U ₂₂)	0.349	0.167	0.175
				θ/β value (U ₂₃)	0.347	0.484	0.506
				Longitudinal acceleration (U ₃₁)	0.583	0.369	0.621
Driving effects (U ₃)	0.480	0.350	0.527	Instantaneous speed of entering the work area (U ₃₂)	0.204	0.336	0.198
				Lane change duration (U ₃₃)	0.213	0.295	0.181

4.5.2. Grading Criteria for Indicators

To eliminate the inconsistency of units and dimensions between indicator data, the original data were forward normalized so that the data range fell within the interval of [0, 1]. Then, the K-means clustering algorithm was used to classify the evaluation indicators through SPSS software [58]. The grading criteria of the evaluation indicators obtained through the K-means clustering algorithm are shown in Table 9.

Table 9. The grading criteria of evaluation indicators.

Evaluation Indicators	The Grading Criteria of Safety and Comfort				
	Lowest	Lower	Medium	Higher	Highest
Pupil area (U_{11})	0–0.29	0.29–0.45	0.45–0.59	0.59–0.74	0.74–1
Saccade frequency (U_{12})	0–0.27	0.27–0.53	0.53–0.69	0.69–0.83	0.83–1
Absolute power of α wave (U_{21})	0–0.18	0.18–0.35	0.35–0.55	0.55–0.77	0.77–1
Absolute power of β wave (U_{22})	0–0.30	0.30–0.50	0.50–0.66	0.66–0.81	0.81–1
θ/β value (U_{23})	0–0.20	0.20–0.32	0.32–0.47	0.47–0.66	0.66–1
Longitudinal acceleration (U_{31})	0–0.23	0.23–0.39	0.39–0.56	0.56–0.71	0.71–1
Instantaneous speed of entering the work area (U_{32})	0–0.24	0.24–0.42	0.42–0.56	0.56–0.71	0.71–1
Lane change duration (U_{33})	0–0.26	0.26–0.44	0.44–0.59	0.59–0.74	0.74–1

4.5.3. Fuzzy Comprehensive Evaluation Results

Based on the experimental data and the established evaluation criteria, the membership degrees of the factor set were determined at each evaluation level, and the membership degree matrices were constructed. Combined with the comprehensive weight vector obtained by the ANP-EWM method, the fuzzy comprehensive evaluation result vector was calculated according to Equation (8). Then, the comprehensive evaluation grade eigenvalue G was calculated by Equation (9). Finally, the ranking of the experimental scenario was determined according to the eigenvalue G . The results are shown in Table 10.

Table 10. Results of fuzzy comprehensive evaluation and ranking of experimental scenarios.

Scenario	Membership Degree of Safety and Comfort Level					Safety and Comfort Level	G	Comprehensive Sort
	Lowest	Lower	Medium	Higher	Highest			
A	0.688	0.000	0.000	0.000	0.312	Lowest	1.682	8
B	0.643	0.045	0.021	0.267	0.024	Lowest	1.449	9
C	0.000	0.466	0.484	0.050	0.000	Medium	2.527	6
D	0.276	0.583	0.141	0.000	0.000	Lower	1.871	7
E	0.151	0.113	0.423	0.306	0.007	Medium	3.115	4
F	0.250	0.062	0.000	0.487	0.201	Higher	3.550	2
G	0.312	0.127	0.078	0.457	0.026	Higher	2.999	5
H	0.306	0.006	0.007	0.348	0.332	Higher	3.478	3
I	0.271	0.041	0.000	0.000	0.688	Highest	4.457	1

Table 10 presents the comprehensive ranking of safety and comfort for the nine experimental scenarios as follows: $I > F > H > E > G > C > D > A > B$. The higher the comprehensive evaluation grade eigenvalue G was, the higher the driver's safety and comfort in this scenario was. The comprehensive evaluation grade eigenvalue of Scenario I (2200 m warning length $\times$ 160 m transition length) was the highest, indicating that the drivers' safety and comfort were the best. The comprehensive evaluation grade eigenvalue of Scenario I (1800 m warning length $\times$ 140 m transition length) was the lowest, indicating that the drivers' safety and comfort were the worst.

According to the evaluation results in Table 10, it is concluded that the comprehensive evaluation grade eigenvalue increased as the warning length increased when the transition length was fixed, indicating that the driver's safety and comfort were improved. When the warning lengths were 2000 m and 2200 m, the comprehensive evaluation grade eigenvalue increased as the transition length increased, indicating improved driver safety and comfort. Therefore, in practical engineering applications, when the road space conditions permit, it is recommended to use a combination of 2200 m warning length $\times$ 160 m transition length.

The safety and comfort level of Scenario A (1800 m warning length $\times$ 120 m transition length) was higher than Scenario B (1800 m warning length $\times$ 140 m transition length), the

safety and comfort level of Scenario C (1800 m warning length $\times$ 160 m transition length) was higher than Scenario D (2000 m warning length $\times$ 120 m transition length), the safety and comfort level of Scenario E (2000 m warning length $\times$ 140 m transition length) was higher than Scenario G (2200 m warning length $\times$ 120 m transition length), and the safety and comfort level of Scenario F (2000 m warning length $\times$ 160 m transition length) was higher than Scenario H (2200 m warning length $\times$ 140 m transition length). Therefore, a reasonable combination of lengths can also achieve high driving safety and comfort when the road space is limited.

5. Discussion

5.1. Results

Although the standards stipulate the work zone layout, only the minimum value is recommended for the length of various work zones [1]. Although the optimal length of work zones has been explored by some studies [59,60], few scholars have considered the influence of different length combinations of warning/transition zones. This study evaluated nine warning–transition length combinations on urban expressways using a comprehensive safety and comfort assessment system. The system evaluated the drivers' visual, cognitive, and driving effects. This system is more comprehensive and systematic compared with the evaluation system that only uses road traffic characteristic indicators [36,61], eye movement indicators [62], or driving behavior indicators [34,35].

The results showed several key findings. First, the optimal combination: Scenario I (2200 m warning $\times$ 160 m transition) achieved the highest safety/comfort score ($G = 4.457$), with significantly improved visual stability (e.g., 61% lower saccade frequency vs. shortest lengths) and smoother driving behavior (longitudinal acceleration closest to zero). Second, shorter length combinations could also achieve higher driving safety and comfort (e.g., 1800 m $\times$ 120 m, ranked #8, but 1800 m $\times$ 140 m, ranked #9), proving that balanced combinations can mitigate spatial constraints. Finally, longer warnings increased the EEG-measured cognitive load (β wave power rose 67% from 1800 m to 2200 m), highlighting a trade-off between driver preparedness and stress.

Longer warnings consistently enhanced safety (e.g., 2000 m and 2200 m warnings reduced lane change duration by 9–16% vs. 1800 m). The results showed that the main effects of the warning/transition lengths on the participants' eye movement and driving behavior indicators were statistically significant (Tables 2 and 4). As the warning length increased, the participants' visual effects and driving effects improved. Additionally, as the transition length increased, the visual effects and driving effects on the participants improved. This may be because longer work zones can provide drivers with a better visual search environment and longer deceleration and lane change distances [38,42]. The analysis results of EEG behavior data showed that the main effects of the warning and transition lengths, and the interaction effect on participants' EEG indicators, were statistically significant (Table 3). As the warning length increased, the participants' cognitive effects decreased. This may be because the longer warning area increased the drivers' cognitive load, resulting in a sense of tension and pressure [44–47].

The proposed ANP-EWM method, which combines the weights of evaluation indicators, avoids the limitations of the single-weight method [57,63]. This method objectively reflects the important relationship between each evaluation indicator. It enhances the scientific validity of the results, providing a new approach for evaluating driving safety and comfort. The fuzzy theory was introduced using a combined weighting method, and the K-means clustering algorithm was employed to classify the evaluation indicators. The theory addresses the uncertainty of the evaluation system of different length combinations of warning and transition lengths. The trapezoidal membership degree function was se-

lected to calculate the membership matrix values of evaluation indicators at all levels. The fuzzy comprehensive evaluation grade eigenvalues were calculated and combined with the comprehensive weight vector. The results showed that the drivers' safety and comfort in Scenario I (2200 m warning length $\times$ 160 m transition length) was the best, consistent with the findings of others [38,64,65].

5.2. Practical Implications

The findings of this study offer actionable insights for optimizing work zone design on urban expressways. Below are key recommendations for practitioners:

- (a) Optimal length combination: Where the road space permits, adopt the 2200 m warning length $\times$ 160 m transition length (Scenario I) to maximize driver safety and comfort. This combination yielded the highest evaluation score and is recommended for new designs or major retrofits.
- (b) Space-constrained conditions: When right-of-way is limited, balanced combinations can mitigate spatial constraints. For example, use 1800 m warning $\times$ 120 m transition (Scenario A, ranked #8) instead of 1800 m $\times$ 140 m (Scenario B, worst-ranked). In addition, for the 2000 m warning, pair with the 160 m transition (Scenario F, ranked #2) rather than longer options (e.g., 2200 m $\times$ 140 m, ranked #3). This option would be vital in dense urban areas, leveraging dynamic message signs (DMSs) to create longer warning zones, thereby reducing physical space needs.
- (c) Interaction effects: The results indicated that there were interaction effects involving warning and transition zones (Tables 2 and 4). Therefore, one should refrain from optimizing warning/transition lengths in isolation. For example, shorter length combinations (e.g., 2000 m $\times$ 140 m, ranked #4) became safer than longer combinations (e.g., 2200 m $\times$ 120 m, ranked #5), thereby reducing the need for excessive land acquisition.
- (d) Human-centered evaluation framework: The study's physiology–psychology–behavior framework can be integrated into safety audits for high-risk work zones. In the future, designers can pilot this approach in intelligent work zones with IoT sensors to monitor real-time driver stress and adjust traffic controls dynamically.
- (e) Regulatory and design updates: The design standards, for example [1,12], should be updated to replace the minimum length requirements with context-dependent optimal ranges. Additionally, region-specific guidelines can be developed using the proposed framework to account for traffic volume, driver demographics, and alignment complexity.

6. Conclusions

This study examined the effects of various combinations of warning and transition lengths in urban expressway work zones on driving safety and comfort. Nine simulation scenarios were created using three warning lengths and three transition lengths. The research focused on the interaction between these lengths and their impact on driving performance by utilizing a multidimensional indicator system for evaluation. The analytic network process and entropy weighting method (ANP-EWM) were employed to determine indicator weights, while fuzzy comprehensive evaluation ranked each scenario. Key conclusions from the study include:

1. Optimal design implementation should prioritize a combination of 2200 m warning length $\times$ 160 m transition length (Scenario I), which yielded the highest safety and comfort scores. This configuration significantly improved driver visual stability and smoother driving behavior. Where right-of-way is limited, practitioners should emphasize extending warning areas over transition zones, adopting balanced combi-

nations like 2000 m × 160 m (Scenario F) or 1800 m × 120 m (Scenario A) to maintain safety without excessive land acquisition.

2. Critical findings revealed that longer warning lengths (≥ 2000 m) consistently enhanced safety performance, reducing the lane change duration by 9–16% and minimizing abrupt deceleration compared to shorter warnings. Notably, EEG data confirmed a cognitive load trade-off: while longer warnings improved operational safety, they increased driver cognitive tension (67% rise in β wave power from 1800 m to 2200 m).
3. This study demonstrated several notable contributions to traffic safety and work zone design. First, unlike traditional single evaluation index methods, this study developed a multidimensional evaluation system, incorporating eye movement data, EEG signals, and driving behavior indicators to assess driver safety and comfort. This holistic approach allowed for a more nuanced understanding of how combined warning and transition lengths impact drivers, capturing physiological and behavioral responses. By integrating these diverse data sources, the study provided a robust assessment of driver performance in work zones, which is more comprehensive than studies that rely solely on traffic characteristics or isolated driver behavior metrics.
4. The methodology used in this study was rigorous. It employed advanced analytical techniques, including the ANP-EWM, to determine the weights of evaluation indicators. This combination of subjective and objective weighting methods reduced the influence of bias and enhanced the objectivity of the results. Furthermore, the introduction of fuzzy theory and the use of a trapezoidal membership function addressed the inherent uncertainty in evaluating safety and comfort levels, providing a more flexible and realistic assessment. Additionally, the study benefited from a relatively large sample size of 45 participants, which enhanced the reliability and generalizability of the findings. The study's rigorous methodology and comprehensive evaluation framework make it a valuable contribution to the field, offering practical insights for improving work zone safety and comfort on urban expressways.
5. For practical implementation, transportation agencies should revise standards, such as China's [12], to recommend evidence-based length combinations. Dynamic message signs (DMSs) can simulate longer warning zones in space-constrained urban corridors. At the same time, the study's multidimensional framework should be adopted for safety audits in high-risk work zones.
6. This study has several limitations that present opportunities for future research. First, the study only examined the types of closed work zones on the outer lanes of urban expressways. Therefore, the results may not apply to other types of work zones. Although the nine length combinations of warning and transition areas were practical, this study did not test longer combinations. Hence, Scenario I is optimal only within the tested range. Second, young drivers (aged 20–28 years) were selected as subjects due to their limited driving experience, which necessitates more operating space. Therefore, future work may focus on drivers of varying ages, including older drivers. Lastly, future research may address nighttime conditions, higher traffic volumes (>2000 pcu/h), and dynamic length adjustments employing IoT-enabled adaptive signage to maximize safety in variable traffic environments.

Author Contributions: A.H., writing—original draft, methodology, project administration, funding acquisition; R.H., writing—methodology and editing; Y.Y., writing—review and editing, supervision, conceptualization, validation; I.E.-D., writing—review and editing, validation; S.M.E., writing—review and editing, analysis, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the [2021 project of the “14th five-year plan” of Educational Science in Fujian Province], approved by the Fujian Provincial Education Science Planning Leading Group Office [Project Approval No. FJJKBK21-042], and the [Central Government’s Special Fund for Local Scientific and Technological Development] (Project No. 2023L3033). And The APC was funded by [2021 project of the “14th five-year plan” of Educational Science in Fujian Province].

Institutional Review Board Statement: This research complied with the American Psychological Association Code of Ethics. All participants signed informed consent.

Data Availability Statement: Data will be made available upon request.

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

References

1. GB 5768.4-2017; Road Traffic Signs and Markings Part 4: Work Zone. Standardization Administration of the People’s Republic of China: Beijing, China, 2017.
2. Pesti, G.; Brydia, R.E. Work zone impact assessment methods and applications. *Transp. Res. Rec.* **2017**, *2617*, 52–59. [CrossRef]
3. Xu, W.X.; Zhao, X.H.; Chen, Y.F.; Bian, Y.; Li, H.G. Research on the relationship between dynamic message sign control strategies and driving safety in freeway work zones. *J. Adv. Transp.* **2018**, *1*, 9593084. [CrossRef]
4. Li, D.F.; Yang, J.R.; Xu, Y.; Wang, Z.G. Study on length of construction warning area of four lane highway. *J. Luoyang Inst. Sci. Technol. Nat. Sci. Ed.* **2017**, *4*, 41–45.
5. Ge, H.; Yang, Y. Research on calculation of warning zone length of freeway based on micro-simulation model. *IEEE Access* **2020**, *8*, 76532–76540. [CrossRef]
6. Liu, X.Q.; Lv, L.; Fan, X.X.; Meng, L.X.; Liu, X.Z.; Bao, Q.W.; Xu, H.D. Study on the Lengths of Upstream Transition Area in Freeway Reconstruction Guidance Section. *J. East China Jiaotong Univ.* **2020**, *37*, 75–81, 95.
7. Liu, S.S. Study on the Length of Construction Control Zone On Urban Road Based on Safety and Efficiency. Master’s Thesis, Chang’an University, Xi’an, China, 2021.
8. Wang, X.F.; Ding, Z.Z.; Liu, Y.; Wang, X.D. Study on spatial security of expressway work area upstream transition zone with mixed vehicles based on VISSIM modeling. *J. Highw. Transp. Res. Dev.* **2022**, *39*, 163–170.
9. Zhang, L.H.; Zhang, B.C. Research on length optimization of urban road maintenance and construction operation area based on VISSIM Simulation. *Highw. Eng.* **2018**, *6*, 207–211.
10. Singh, M.K.; Pandey, G.; Rao, K.R. Optimization of work zone segments on urban roads using cellular automata model in mixed traffic. *Period. Polytech. Transp. Eng.* **2019**, *47*, 29–35. [CrossRef]
11. Han, Y.J.; Du, L.L.; Wei, Q.C. Research on the reasonable length of construction area based on VISSIM simulation experiment. *Shanghai Highw.* **2024**, *02*, 104–111.
12. JTG H30-2015; Ministry of Transport. Highway Maintenance Safe Operation Regulations. People’s Communications Press: Beijing, China, 2015.
13. Omkar, B.; Shriniwas, A.; Gaurang, J.; Said, M.E. Effect of Construction Work Zone on Rear-End Conflicts by Vehicle Type under Heterogeneous Traffic Conditions. *J. Transp. Eng. Part A Syst.* **2023**, *149*, 05023001.
14. Tang, W.Y.; Wang, H.B.; Ma, J.X.; Yang, C.Y.; Yin, C.Y. Vehicle collision risk assessment method in highway work zone based on trajectory data. *Traffic Inj. Prev. Online First* **2025**, 1–8. [CrossRef] [PubMed]
15. Garber, N.J.; Zhao, M. Distribution and characteristics of crashes at different locations within work zones in Virginia. *Transp. Res. Rec. J. Transp. Res. Board* **2002**, *1794*, 19–25. [CrossRef]
16. Zhang, C.; Qin, J.; Zhang, M.; Zhang, H.; Hou, Y.D. Practical road resistance functions for expressway work zones in occupied lane conditions. *Sustainability* **2019**, *11*, 382. [CrossRef]
17. Zhang, C.; Wang, B.; Yang, S.; Zhang, M.; Zhang, H. The driving risk analysis and evaluation in rightward zone of expressway reconstruction and extension engineering. *J. Adv. Transp.* **2020**, *2020*, 8943463. [CrossRef]
18. Ishak, S.; Qi, Y.; Rayaprolu, P. Safety evaluation of joint and conventional lane merge configurations for freeway work zones. *Traffic Inj. Prev.* **2012**, *13*, 199–208. [CrossRef]
19. Umar, I.K.; Bashir, S. Assessment of work zone safety. *Nat. Eng. Sci.* **2018**, *3*, 62–71.
20. Li, Y. Analysis on Traffic Characteristics and the Safety Risk of Freeway Workzone. Master’s Thesis, Chang’an University, Xi’an, China, 2014.
21. Li, Y.; Zhang, S.R.; Ru, M.B. Car following rear-end conflict risk of freeway work zone. *J. Chang’an Univ. (Nat. Sci. Ed.)* **2017**, *37*, 81–88.
22. Li, X.M.; Yan, X.D.; Wong, S.C. Effects of fog, driver experience and gender on driving behavior on s-curved road segments. *Accid. Anal. Prev.* **2015**, *77*, 91–104. [CrossRef]

23. Heaslip, K.; Louisell, C.; Collura, J. Modeling impacts of driver behavior on highway performance and safety during forced merges in work zones. In Proceedings of the World Congress on Intelligent Transport Systems and ITS America Annual Meeting, London, UK, 8–12 October 2006.
24. Ulak, M.B.; Ozguven, E.E.; Ren, M.; Sando, T.; Boot, W.; Abdelrazig, Y.; Sobanjo, J.O. Assessment of traffic performance measures and safety based on driver age and experience: A microsimulation based analysis for an unsignalized T-intersection. *J. Traffic Transp. Eng. (Engl. Ed.)* **2019**, *6*, 455–469. [CrossRef]
25. Karacasu, M.; Er, A. An analysis on distribution of traffic faults in accidents, based on driver's age and gender: Eskisehir case. *Procedia-Soc. Behav. Sci.* **2011**, *20*, 776–785. [CrossRef]
26. Khan, R.A. Driver's Comfort Level in Construction Zones with Reduced Transition Taper Length: Case Study for Pakistani Conditions. Master's Thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 2011.
27. Zhao, X.H.; Ju, Y.J.; Li, J.; Tao, P.P.; Wang, X.S.; Zhou, R.G. Evaluation of the effect of RPMs in extra-long tunnels based on driving behavior and visual characteristics. *China J. Highw. Transp.* **2020**, *33*, 29–41.
28. Cheng, W.D.; Fu, R.; Ma, Y.; Zhou, Y.; Liu, J.K. Research on driver's cognitive distraction in mobile phone call behavior based on image recognition. *China J. Highw. Transp.* **2021**, *34*, 168–181.
29. Belyusar, D.; Reimer, B.; Mehler, B.; Coughlin, J.F. A field study on the effects of digital billboards on glance behavior during highway driving. *Accid. Anal. Prev.* **2016**, *88*, 88–96. [CrossRef]
30. Costa, M.; Simone, A.; Vignali, V.; Lantieri, C.; Palena, N. Fixation distance and fixation duration to vertical road signs. *Appl. Ergon.* **2018**, *69*, 48–57. [CrossRef] [PubMed]
31. Li, Y.S.; Wang, W.J.; Wang, M.S. Optimization design of urban tunnel lighting. *J. Chongqing Jiaotong Univ. (Nat. Sci.)* **2020**, *39*, 76–80.
32. Zhou, Y.; He, S.; Li, X.; Li, Y.; Du, G.; Hailu, Y. Study on drivers' visual load features in lighting environments of interior zones of extra-long tunnels over 10 km. *Adv. Civ. Eng.* **2022**, *2022*, 4840334. [CrossRef]
33. Wang, X.Y. Study on the Relationship Between Traffic Safety and Biomass Index at the Entrance of Highway Tunnel. Master's Thesis, Chongqing Jiaotong University, Chongqing, China, 2024.
34. Hu, M.W.; Wu, W.L.; Tian, Q.Y.; Guo, Z.X.; Yang, W.J. Evaluating tunnel lighting impact on driver safety and comfort via driving simulation. *J. Nanjing Univ. Inf. Sci. Technol.* **2024**, *5*, 654–666.
35. Qiao, J.G.; Yang, C.; Chen, Y.X. Evaluation of driving comfort on longitudinal slope section of highway based on driver's heart rate. *Sci. Technol. Eng.* **2022**, *22*, 11691–11707.
36. Peng, Y.H.; Wang, X.Y.; Lu, J.N.; Wang, W.; Zhang, H.F. Length determination method for expressway work zone based on service level. *China J. Highw. Transp.* **2016**, *29*, 130–136.
37. Ren, S.B. Research on Fine-grained Traffic Organization and Safety Assurance Technologies in the Construction Zones of Highway Reconstruction and Expansion Projects. Master's Thesis, Chang'an University, Xi'an, China, 2024.
38. Yang, Y.S. Research on the Length Optimization of Traffic Bottleneck Zone for Freeway Maintenance Work—A Case Study of the Outer Lane Closed. Master's Thesis, Jiangsu University, Zhenjiang, China, 2020.
39. Lei, J.; Yu, C.J.; Zhao, F.L. Study on transition area length of expressway work zone in half single closed lane of highways. *J. China Foreign Highw.* **2021**, *41*, 370–375.
40. Song, X.W.; Zang, X.S.; Dong, Y.; Zhu, P.Y.; Cheng, Y.Z. Research on the Length of Advance Warning Area for Maintenance Work at Vehicle-Bicycle Mixed Highway. *J. Huaiyin Inst. Technol.* **2022**, *31*, 10–15.
41. Chow, S.C.; Shao, J.; Wang, H.S. *Sample Size Calculations in Clinical Research*, 3rd ed.; Chapman Hall/CRC Biostatistics Series; CRC Press: Boca Raton, FL, USA, 2017.
42. Orden, K.V.; Jung, T.; Makeig, S. Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biol. Psychol.* **2000**, *52*, 221. [CrossRef] [PubMed]
43. Foy, H.J.; Chapman, P. Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Appl. Ergon.* **2018**, *73*, 90–99. [CrossRef]
44. Guan, W.; Yang, L.; Jiang, S.X.; Zhang, W.Y. Review on the application of EEG in traffic driving behavior study. *J. Transp. Syst. Eng. Inf. Technol.* **2016**, *16*, 35–44.
45. Brookhuis, K.A.; Waard, D.D.; Janssen, W.H. Behavioural impacts of advanced driver assistance systems; an overview. *Eur. J. Transp. Infrastruct. Res.* **2001**, *1*, 245–253. [CrossRef]
46. Yang, Y.Q.; Chen, Y.; Easa, S.M.; Lin, J.; Chen, M.F.; Zheng, X.Y. Evaluation of driver's situation awareness in freeway exit using backpropagation neural network. *Transp. Res. Part F Psychol. Behav.* **2024**, *105*, 42–57. [CrossRef]
47. Yang, Y.Q.; Ye, Z.H.; Easa, S.M.; Yang, F.; Zheng, X.Y. Effect of driving distractions on driver mental workload in work zone's warning area. *Transp. Res. Part F Psychol. Behav.* **2023**, *95*, 112–128. [CrossRef]
48. Ren, L.H.; Nie, Z.L.; Yu, X.; Chen, K.X.; Jiang, C.Y. Review on Recognition of Unsatisfactory Driving Status Based on Electroencephalogram. *China J. Highw. Transp.* **2024**, *37*, 216–230.
49. Xu, X.Q.; Feng, Z.X.; Li, J.Y. Research progress of drivers taking over autonomous vehicles. *Traffic Inf. Saf.* **2019**, *37*, 1–6.

50. Guo, F.X.; Li, G.S.; Ni, D.A.; Yang, W.C.; Tian, B.J. Reliability analysis of driver's driving behavior in the non-signalized intersection warning system. *J. Xi'an Univ. Technol.* **2023**, *39*, 208–215.
51. Li, X.; Sun, J.Q. Studies of vehicle lane-changing dynamics and its effect on traffic efficiency. *Saf. Environ. Impact* **2017**, *467*, 41–58.
52. Cui, G. Research on security risk assessment and application for highwaybridge during construction based on the theory of analytic network process. Master's Thesis, Changsha University of Science Technology, Changsha, China, 2014.
53. Su, Z.; Wang, G.; Yan, J.; Dong, C. Evaluation model of rural drinking water project based on entropy weight and fuzzy comprehensive evaluation method. *Asian Agric. Res.* **2022**, *14*, 62–65.
54. Wu, X.X.; Shen, S.; Ma, S.L.; Jiang, S.F.; Lin, X.W.; Liu, F.S.; He, X.B.; Zhang, P.X.; Wu, S.F. Improved method for safety assessment of bridge underwater pile foundation based on entropy method determining defect weight. *J. Fuzhou Univ. Nat. Sci. Ed.* **2022**, *50*, 111–119.
55. Zhang, J.W.; Zheng, Y.H.; Li, M.B. Analysis of safe driving behavior based on combination weighting and fuzzy comprehensive evaluation. *Sci. Technol. Eng.* **2022**, *22*, 11255–11261.
56. Zhang, Q.; Li, G.Q.; Zhuge, Y.S.; Yu, X.; Tan, H.W.; Du, Q. Application of improved fuzzy comprehensive method to water quality assessment in Erhai. *J. China Inst. Water Resour. Hydropower Res.* **2019**, *17*, 226–232.
57. Zhang, B.; Jiang, C.L.; Zheng, L.G.; Zhao, Q.; Su, W.; Chen, Y.P. Application of improved fuzzy comprehensive evaluation method in water quality assessment in a coal mine area. *Adm. Technol. Environ. Monit.* **2022**, *34*, 27–32.
58. Liu, M.M. Design and Implementation of Medical ServicePerformance Evaluation System Based on Fuzzy evaluation Model. Master's Thesis, Hebei Normal University, Shijiazhuang, China, 2021.
59. Mi, R.L.; Lee, C.; Kim, D.G. The proper length of transition area for work zones on urban freeways. *J. Korean Soc. Transp.* **2013**, *31*, 58–66.
60. Yu, G.Y. Risk Assessment and Length Setting Study of Mountain Highway Maintenance Work Area. Master's Thesis, Chongqing Jiaotong University, Chongqing, China, 2023.
61. Jiao, N.S.; Hu, C.; Gao, L.Y.; Ding, N.K. Research on segment length of expressway reconstruction and expansion construction area based on information entropy. *Transp. Sci. Technol.* **2019**, *6*, 10–13.
62. Zhou, S. Research on Safety Evaluation of Highway Maintenance Operations Based on Driver Visual Characteristics. Master's Thesis, Shandong Jiaotong University, Jinan, China, 2024.
63. Zhou, H.Y.; Bao, Q.G.; Ye, M.; Zhu, S.W.; Lin, Y.D. Fuzzy comprehensive evaluation of urban bridge reliability based on combination weighting method. *China Saf. Sci. J.* **2023**, *33*, 155–161.
64. Li, F. Length study on maintenance construction operating area of urban road based on simulation. *China Munic. Eng.* **2017**, *5*, 65–68.
65. Wang, J. Research on Traffic Conflict and Driving Risk in Nighttime Traffic Control Zone for Maintenance Work of Expressway. Master's Thesis, Chongqing Jiaotong University, Chongqing, China, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Micro-Driving Behavior Analysis of Drivers in Congested and Conflict Environments Using Graph Theory

Rongjun Cheng, Xudong An, Weiqi Fan and Dan Zhao *

Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China;
chengrongjun@nbu.edu.cn (R.C.); 2211120001@nbu.edu.cn (X.A.); 236000522@nbu.edu.cn (W.F.)

* Correspondence: zhaodan@nbu.edu.cn

Abstract: Many traffic conflicts on the roads are caused by a small proportion of drivers. Currently, there are few studies exploring the time-varying patterns of driving behavior among these drivers. This paper proposes a generic time-series analytical framework and uses it to analyze the driving behavior patterns of many high-risk drivers, which provides a theoretical and targeted basis for vehicle warning systems. Specifically, the natural trajectory time-series data in the rear-end conflict process from congested highway sections were first obtained. Secondly, K-medoid clustering was utilized to obtain the quantitative driving behavior sequence from the trajectory. Thirdly, the driving behavior sequence was transformed into a graph structure by the co-occurrence matrix. Graph theory and Markov theory were used to analyze the obtained graph to achieve the goal of analyzing the time-varying patterns of driving behavior. The analysis found that the driving behavior transition graph network of high-risk drivers on congested highway sections does not exhibit the small-world property and this suggests that during the conflict process, the driver is unable to quickly transition between states. Additionally, vehicles consistently evolve into a rear-end conflict state along a fixed driving behavior transition route, which indicates that the causes of conflicts in congested road sections are similar. Finally, the state change of the conflict process follows the Markov property, proving that the state during the conflict process can be predicted and controlled.

Keywords: rear-end conflict analysis; K-medoid clustering; Markov chain; complex network theory

1. Introduction

The real-world traffic environment is highly complex, presenting significant challenges for conventional traffic management. Targeted traffic management policies offer a potential solution. However, developing effective policies requires detailed classification of traffic environments or driver types, enabling in-depth exploration and characterization of patterns associated with specific scenarios or driver cohorts.

Traditional traffic safety analysis and management have predominantly relied on crash data. Yet, such data are scarce and require extensive time to collect [1,2]. Studies have demonstrated the feasibility of using traffic conflict techniques for road safety analysis [3–5], as a strong correlation exists between conflicts and crashes [6]. Compared to crash data, traffic conflict data are significantly more accessible [6–8]. Consequently, an increasing number of researchers are investigating traffic safety through traffic conflict technologies.

Extensive research utilizing traffic conflict techniques has demonstrated significant correlations between traffic states on road segments—including traffic flow, occupancy,

average speed, and vehicle composition—and environmental factors such as weather, with the occurrence of traffic conflicts [9–12]. Generally, higher traffic volumes and occupancies, coupled with greater fluctuations in average speed and increased traffic flow heterogeneity, are associated with a higher incidence of conflicts and potential accidents. While conflict frequency may rise under congested conditions, only a small proportion of vehicles are typically involved in such events. This suggests that deteriorating traffic conditions predominantly impact a limited subset of drivers. Consequently, analyzing the driving behaviors of those involved in rear-end conflicts offers valuable insights. Such investigations can elucidate characteristic behavioral patterns among these drivers and inform targeted traffic safety interventions.

Previous studies have examined conflict evolution from diverse perspectives. While investigations [13–15] have analyzed the propagation chain of inter-vehicle conflicts, they paid limited attention to the microscopic behavior of individual vehicles. In Ref. [9], the authors segmented road sections and examined risk evolution both within and between these segments. Similarly, the authors in [16] subdivided sections into finer spatial blocks to explore spatio-temporal risk evolution across blocks. A recent work [17,18] further investigated the influence of factors like speed variance and average speed of involved vehicles on conflict likelihood. However, these studies collectively failed to characterize the entire, fine-grained conflict process. Critically, existing research has primarily focused on conflict risk evolution within specific spatio-temporal road segments, overlooking the specific drivers involved in conflicts and the time-varying patterns of driving behavior among these high-risk drivers.

To address this research gap, we conducted a comprehensive investigation into the rear-end conflict process within congested highway segments. This study explicitly considers the temporal evolution of driver behavior throughout the entire conflict sequence and accounts for the heterogeneity inherent in congested traffic flows. The primary contributions of this work are summarized as follows:

1. An explainable time-series analysis framework: We propose a novel framework designed for the explainable analysis of time-varying patterns, enabling fine-grained interpretation of temporal dynamics.
2. Heterogeneity-aware conflict analysis under congestion: Leveraging the proposed framework, we analyzed a substantial dataset of conflict processes occurring under heterogeneous congested traffic conditions. The resulting actionable insights provide more detailed and targeted guidance for enhancing vehicle collision warning systems.
3. A transferable analytical paradigm: The developed framework establishes a methodological paradigm that can be readily adapted to support time-series analysis challenges in other research domains.

The rest of this paper is organized as follows. Section 2 reviews some related works about traffic conflict technology. In Section 3, graph theory and Markov theory are briefly introduced. The proposed explainable analysis framework for rear-end conflicts is described in detail in Section 4. Section 5 conducts experiment on a real traffic dataset and presents discussions about the results. This paper is summarized, and some future directions are discussed in Section 6.

2. Related Work

Traffic conflict techniques have gained widespread recognition as valuable tools in traffic safety research. Current investigations into these techniques primarily focus on areas such as traffic conflict prediction, safety evaluation, and conflict analysis. Common methodological approaches include statistical models, machine learning, and deep learning. These methods universally rely on extensive datasets, extracting relevant features and

employing suitable algorithms to address specific research challenges. Findings derived from such studies inform the formulation of evidence-based traffic management measures.

2.1. Statistical Methods

Traditional statistical models usually receive some traffic characteristics of a specific road section or period as input to the model and then obtain a classifier that can output a conflict probability. These methods include logistic regression, linear regression, and so forth. Sensitivity analysis and feature importance are often used to identify factors that cause conflict. Their advantage is that they have good interpretability, which is why they are often used for factor analysis [19–24]. Simultaneously, these methods are often employed to deal with macroscopic traffic features, as they are unable to handle high-dimensional microscopic traffic characteristics such as vehicle trajectories. Therefore, they are difficult to use for fine-grained time-varying pattern analysis, and another reason is that they cannot handle time-series data.

2.2. Machine Learning and Deep Learning Methods

Machine learning (ML) and deep learning (DL) methodologies aim to extract latent features or approximate data distributions from large-scale traffic datasets. These approaches are frequently employed for microscopic conflict prediction and analysis. For instance, studies leverage camera image data for real-time conflict probability prediction [25–31], while others utilize neural networks to model the data distribution of vehicle trajectories for conflict severity assessment and prediction [32–34]. However, ML and DL methods face practical challenges, including insufficient generalization capability and limited interpretability [6,35]. Although adept at processing high-dimensional data like vehicle trajectories, the specific factors driving model determinations often remain obscure.

Risk field theory (RFT) is commonly applied to assess the spatiotemporal evolution of risk for traffic conflict safety evaluation [21]. While RFT quantifies risk levels within defined spatiotemporal boundaries, it typically does not account for the specific microscopic evolution process of rear-end conflicts. Regarding conflict formation, some research has investigated the macro-scale spatial and temporal evolution of conflict risk [15,36,37]. Nevertheless, due to traffic flow heterogeneity, these findings may lack applicability in congested scenarios. Furthermore, these identified spatiotemporal risk patterns are derived from statistical aggregation across numerous vehicles, potentially overlooking the specific behavioral patterns of vehicles directly involved in conflicts.

As summarized in Table 1, while numerous researchers have proposed novel methods addressing various traffic safety problems and employed technical means to infer the causative factors of conflicts, limited attention has been given to the precise microscopic details characterizing rear-end conflict processes. This includes, for example, the evolution of driver behavior patterns during the conflict event itself.

Table 1. Summary of commonly used methods in related work and introduction.

	Input	Output
Statistical methods	Cross-sectional data	Conflict probability
Machine learning	Cross-sectional data	Conflict probability and number
Deep learning methods	Time series	Conflict probability and number
Spatial Markov	Time series	Conflict Probability distribution

2.3. Analysis of the Causes of Conflict

A conflict causality analysis and risk prevention and control are core intelligent transport system research areas. In line with conventional car-following conflict modeling, the

Random Parameters Logit Model (RPLM) was used at a two-way-stop controlled intersection in Aachen, Germany. It highlighted the significant impact of lead vehicle dynamic variables like speed and acceleration and first spotted the “unnecessary intentional deceleration” phenomenon, offering a theoretical basis for conflict intervention strategies [38].

To address rear-end collision risks, an evaluation framework combining deep neural networks and deep reinforcement learning was proposed. By dynamically adjusting the following vehicle speed, it effectively reduced rear-end collision probability, with its effectiveness verified by both simulation and real-vehicle data [39]. Further, an analytical approach based on time-to-collision (TTC) and its derivative (TTCD) was presented. It quantified the vehicle transition mechanism from safe to dangerous states, identified three typical transition modes, and confirmed that the following vehicle acceleration, speed difference, and acceleration difference between the two vehicles are significant TTCD-influencing factors [40].

To enhance the safety-distance model’s versatility, an improved car-following safety model introduced driver heterogeneity parameters. Using fuzzy theory to dynamically correct reaction time, it achieved precise safety-distance calculation under multiple traffic and driver-type conditions, underpinning active traffic-accident early warning [41].

In autonomous driving safety research, a pre-collision scenario classification method mapped and identified 24 high-incidence scenarios (98.1% accuracy), with rear-end and intersection conflicts identified as key risk scenarios. Causal- and correlation-based key-factor mining provided targeted paths for autonomous driving system optimization [42].

Special research on advanced driving systems (ADSs/ADASs) revealed that rear-end collisions accounted for 76.1% of cases in the NHTSA accident dataset. Ego vehicle speed, collision object motion state, and road type were identified as significant influencing factors. These findings are highly instructive for system design standards and road-user training [43].

For improved mixed-traffic-flow safety, a hybrid control framework was macro-simulated. It demonstrated a positive correlation between connected and automated vehicle (CAV) penetration rate and road capacity. Under high penetration rates, the framework effectively suppressed traffic oscillations and reduced collision risks [44].

Regarding lane-changing maneuver safety early warning, a hybrid model combining Mexican Hat Wavelet (MHW) transform and genetic-algorithm-optimized XGBoost (GA-XGBoost) achieved precise failed lane-changing detection (94.55% accuracy). It revealed a strong link between high collision risks of target-lane trailing vehicles and failed lane changes [45].

3. Preliminary

3.1. The Driving Behavior Time-Varying Pattern Analysis Based on a Graph

The purpose of the paper is to explore the patterns of vehicle state changes during rear-end collision processes. Our strategy involves converting the trajectory time-series data during conflicts into a network view structure. Initially, we utilize clustering methods on trajectory data to obtain quantified behavioral states of vehicles, followed by the analysis of transitions between these quantified behavioral states. In terms of constructing the graph structure, we represent the conflict process in the entire dataset by establishing a directed weighted graph, with the vehicle behavioral states as nodes, the transitions between behavioral states as edges, and the frequency of these transitions as edge weights.

Subsequently, we must analyze the obtained network graph, which includes assessing the significance of behavioral states, identifying patterns in state transitions, and examining the overall characteristics of the graph. Specifically, we employed complex network theory and Markov theory. Next, we will delve into the underlying technical principles.

Complex networks are commonly used to model the evolution process of systems. There are some concepts related to graph and edge property analysis in complex network theory:

PageRank: The larger the PageRank value, the more important the node is. If a driving state node has a high PageRank score, it indicates that this driving state is crucial throughout the entire conflict process. Its formula is shown in (1).

$$PR(v_i) = \sum_{v_j \in M(v_i)} \frac{PR(v_j)}{L(v_j)}, i = 1, 2, \dots, n \quad (1)$$

where $M(v_i)$ represents the set of nodes to node v_i , and $L(v_j)$ represents the number of directed edges connected by node v_j .

Closeness centrality: Its calculation formula is shown in (2). A driving behavior state node with a high closeness centrality can more quickly influence other driving states.

$$\begin{cases} d_i = \frac{1}{N-1} \sum_{j=1}^N d_{ij} \\ CC_i = \frac{1}{d_i} \end{cases} \quad (2)$$

where d_i represents the average distance from node v_i to the other node v_j .

Small-world property: The small-world characteristics are determined by the shortest path and the aggregation coefficient of graph. The shortest path is calculated by Formula (3) and it represents the average of paths between any two nodes in the network. The aggregation coefficient is calculated by Formula (4) and it indicates whether a node is densely connected to its neighboring nodes. When the shortest path is long and the aggregation coefficient is small, it indicates that the network does not have small-world characteristics. In this paper, we utilize these two metrics to determine whether the conflict process network exhibits small-world characteristics.

$$L = \frac{1}{n(n-1)} \sum_{i \geq j} d_{ij} \quad (3)$$

where n is node number and d_{ij} is the shortest distance between any state nodes.

$$cl_T = \frac{(A + A^T)_{ii}^3}{2[d_i^{tot}(d_i^{tot} - 1) - 2(A^2)_{ii}]} \quad (4)$$

where cl_T is the aggregation coefficient, A represents the adjacency matrix of the network and d_i^{tot} represents the total degree of state node i .

Through these analysis methods, we can obtain the critical driving states during the conflict process and the transfer relationships between these states and conduct an explainable analysis of the conflict process accordingly.

3.2. Markov Property Analysis of Behavior Sequence

To explore the stochastic process of vehicle driving behavior during rear-end collisions, we conducted a statistical analysis of the changes in driving direction and frequency, resulting in the derivation of a co-occurrence matrix for driving behavior. The co-occurrence matrix, resembling a transition matrix, requires assessment to ascertain its adherence to Markovian properties. The Markov test can be used to analyze whether a discrete sequence satisfies a stochastic process, often using chi square statistics for testing.

Based on frequency transition matrix $F_{m \times m}$ and probability transition matrix $P_{m \times m}$, we can construct chi square statistics as Formula (5).

$$\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \log \frac{p_{ij}}{p_{.j}} \right| \quad (5)$$

where m is the number of states, f_{ij} and p_{ij} are the transition probabilities and frequencies from state i to state, respectively, and $p_{.j}$ is the boundary probability, which is calculated by Formula (6).

$$P_{.j} = \frac{\sum_{j=1}^m f_{ij}}{\sum_{i=1}^m \sum_{j=1}^m f_{ij}} \quad (6)$$

If the chi square statistic of the driving behavior sequence during the conflict process is less than the p value, it indicates that the conflict process satisfies the Markov property.

4. Methodology

4.1. Problem Definition

Analyzing conflict dynamic change mechanism under temporal effects can provide a better understanding of traffic conflict. Its purpose is to analyze the time-varying relationship during the process based on the natural trajectory times series data of vehicles before the conflict occurs by using relevant technical methods, such as to analyze the quantitative time-varying relationship of driving patterns. More specifically, the fine-grained variation patterns of headway, velocity, and acceleration in the time dimension during the conflict process are explored in this paper.

4.2. The Explainable Analysis Framework of Conflict Process

Figure 1 illustrates the conflict analysis framework proposed in this paper. The graph structure generation flowchart is shown in the left half, while the interpretability analysis of the driving state sequences and graph structure is shown in the right half.

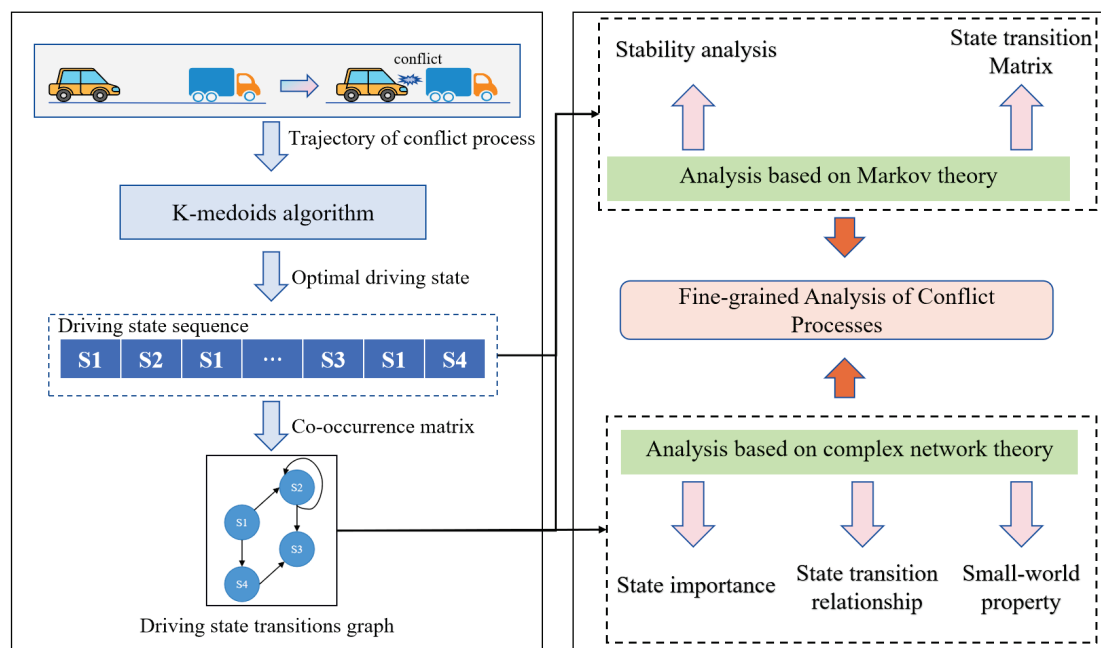


Figure 1. The explainable analysis framework for rear-end conflicts.

Upon identifying a conflict, we preserved the trajectories leading up to the conflict and systematically traversed the entire dataset, yielding hundreds of statistically significant trajectory records. Subsequently, we clustered the mean values of three features—vehicle speed, headway distance, and acceleration—within a 0.5 s time window to obtain quantitative driving behavior states for that time frame. Furthermore, by converting trajectory data into a sequence of driving states and conducting statistical analysis on this sequence, we derived a co-occurrence matrix. This matrix allowed us to construct the network structure of the conflict process in the dataset, unveiling its intricate complexity through a network graph. To obtain driving states during conflict processes, K-medoids are employed in unsupervised clustering. The ideal number of clusters is then ascertained using the Calinski–Harabasz score. Better clustering is indicated by a higher Calinski–Harabasz score.

On the right side of the illustration, the upper section features an analysis based on Markov theory, while the lower section presents an analysis rooted in complex network theory. These analyses primarily focus on the significant driving behavior states during the conflict process, the stochastic nature of driving behavior transitions, and the overall characteristics of these transitions. On the right side of the illustration, the upper section features an analysis based on Markov theory, while the lower section presents an analysis rooted in complex network theory. These analyses primarily focus on the significant driving behavior states during the conflict process, the stochastic nature of driving behavior transitions, and the overall characteristics of these transitions.

5. Experiments and Results

5.1. Dataset Introduction

DJI's Aerial Dataset for China Congested Highway and Expressway (AD4CHE) provides detailed natural trajectory data of vehicles on multiple highways and expressways in China [46]. Its frame rate is 30 per second and the final distance error does not exceed 10 cm. It has been used by many scholars for research in areas such as trajectory prediction and traffic safety analysis.

Figure 2a shows a schematic diagram of the overhead shooting of drone, in which over 60 h of videos were captured in different locations using this method. In Figure 2b, a specific road section is shown, and the vehicles obtained through YOLO target recognition show dense traffic and slow vehicle speeds, indicating congestion on the road section.

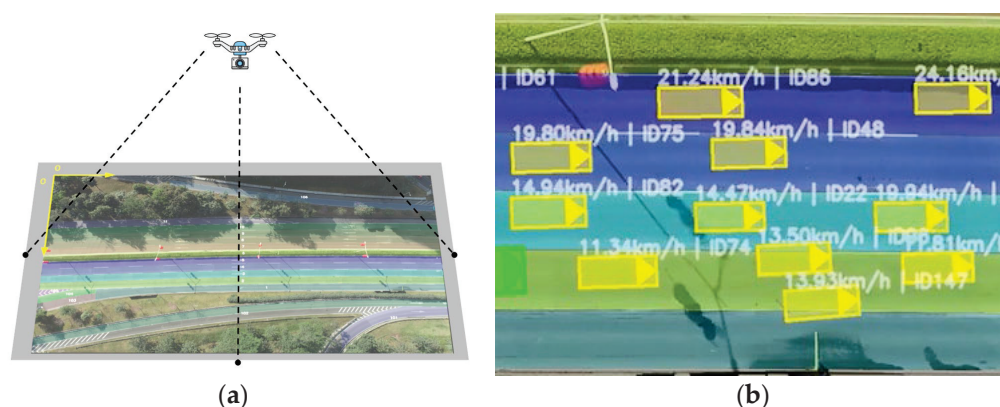


Figure 2. AD4CHE data schematic diagram. (a) Schematic diagram of drone shooting; (b) congested traffic flow. The rectangular box here represents the vehicle target and the triangle represents the front of the vehicle.

Table 2 shows partial records of AD4CHE, which records various data of each frame of the vehicle and encompasses lateral and longitudinal velocity and acceleration (x is

longitudinal and y is lateral), headway time and TTC. When TTC is less than 0, it indicates that the distance between the two vehicles is gradually widening.

Table 2. Partial fields of AD4CHE.

Frame	x Velocity	y Velocity	x Acceleration	y Acceleration	dhw	TTC
0	1.98	0.01	0.48	0.02	4.88	−37.65
1	1.98	0.01	0.48	0.02	4.95	−38.23
2	2.00	0.01	0.45	0.00	4.91	−41.08
3	2.01	0.01	0.45	0.00	4.91	−45.00
4	2.03	0.01	0.43	0.00	4.88	−49.09
5	2.04	0.01	0.41	0.00	4.72	−52.27

For conflict identification, when TTC of vehicle is less than 1.5 s, we consider it a conflict [46]. In total, we obtained 257 rear-end conflicts from AD4CHE. We selected acceleration, headway, and lateral acceleration to describe driving behaviors of the rear vehicle. The reason is that they can fully record the following behavior during the conflict process, which was inspired by the car-following model. Finally, we take the moment of conflict occurrence as the zero-time point and then select the trajectory data within 5 s before the conflict [10].

5.2. Driving State Clustering

The optimal clustering number is determined by the Calinski–Harabasz score. Figure 3 shows the Calinski–Harabasz score of different clustering numbers. When the cluster number is 8, the Calinski–Harabasz score is the largest, which indicates that the optimal cluster number is 8. It means that the driving states during the conflict process can be roughly divided into 8 types. Therefore, we set the number of clusters for K-medoids to 8.

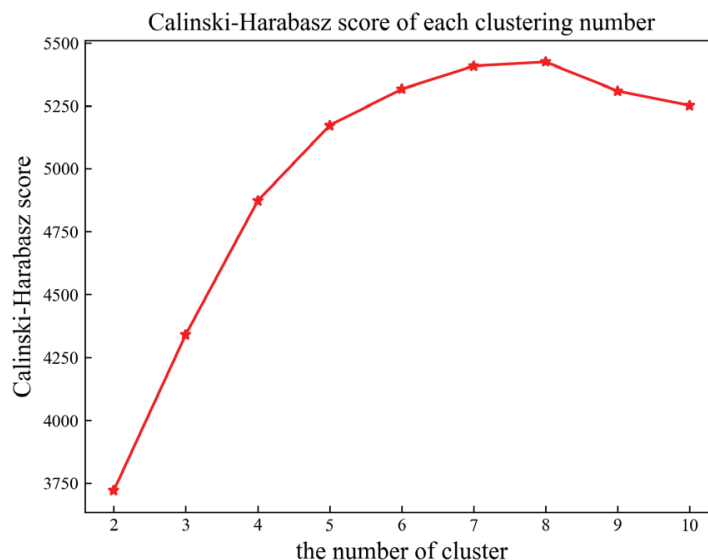


Figure 3. The Calinski–Harabasz score of each clustering number.

From Figure 4, it can be observed that the hierarchy between every cluster is clear, and the samples in each cluster are concentrated, proving that the clustering results are appropriate. Furthermore, the center of each cluster is shown in Table 3.

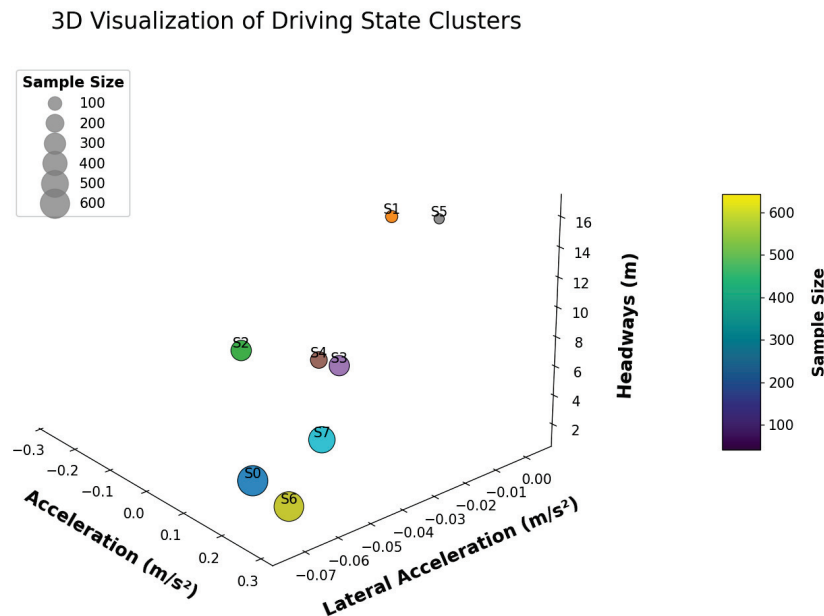


Figure 4. The clustering results.

Table 3. Cluster center for each driving state.

Driving State	Mean of Acceleration (m/s ²)	Mean of Lateral Acceleration (m/s ²)	Mean of Headways (m)	Total Sum
S0	0.232	−0.074	4.074	643
S1	−0.044	0.003	11.970	74
S2	−0.264	−0.020	1.521	273
S3	0.135	−0.036	7.107	270
S4	0.181	−0.048	9.096	172
S5	0.221	−0.016	16.354	41
S6	0.289	−0.070	2.846	615
S7	0.255	−0.056	5.517	482

In Figure 5, the characteristics of different driving states are significantly different: S2 has the strongest deceleration (median acceleration -0.264 m/s^2) and S6 has the most obvious acceleration (0.289 m/s^2); S5 has the largest headway (16.35 m) and S2 has the smallest (1.52 m); lateral acceleration is mostly negative, where only S1 is slightly positive (0.003 m/s^2). The sample size varies greatly (S0/S6 has more than 600 cases, S5 has only 41 cases), and the box plot distribution is consistent with the cluster center, verifying the distinguishability of the state. S1 has the smallest lateral acceleration fluctuation and the most stable lateral movement. The distribution morphology of the box plots of each state is consistent with the cluster center, verifying the physical distinguishability of the driving state. Among them, the lateral acceleration of the S1 state has the smallest range of change, indicating that its lateral movement is the most stable.

Figure 6 shows the total number of each category, with S0 and S6 having the highest number. These two states are also the closest states to conflict, indicating that most drivers will experience these two states before a conflict occurs.

Figure 7 shows a bar chart of the average longitudinal acceleration, average lateral acceleration, and average distance between vehicles for each clustering state, which clearly shows the value size of each category. Among them, state 3 has the smallest headway and a larger deceleration, indicating that this state is closest to collision.

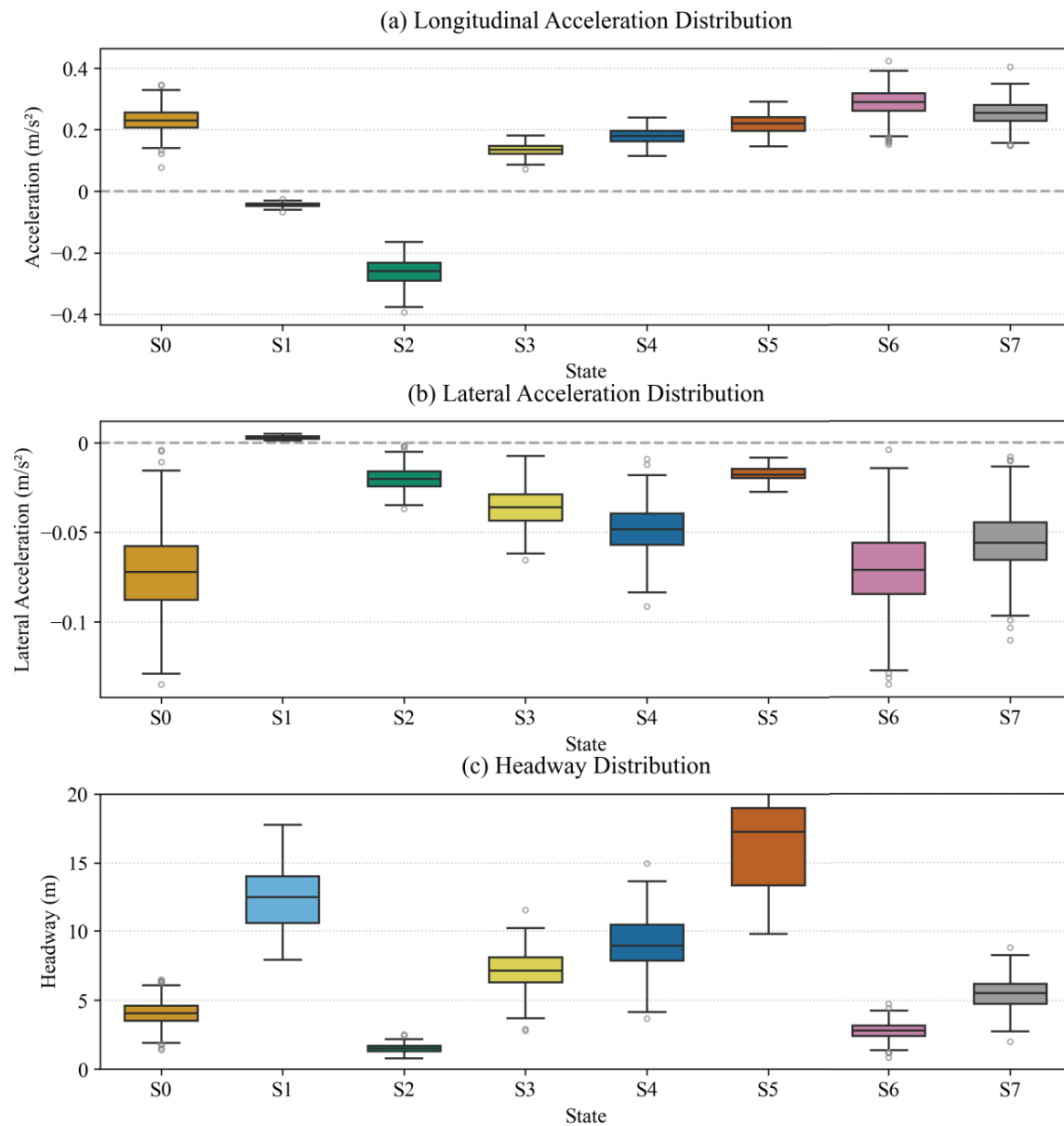


Figure 5. Box plots of different states.

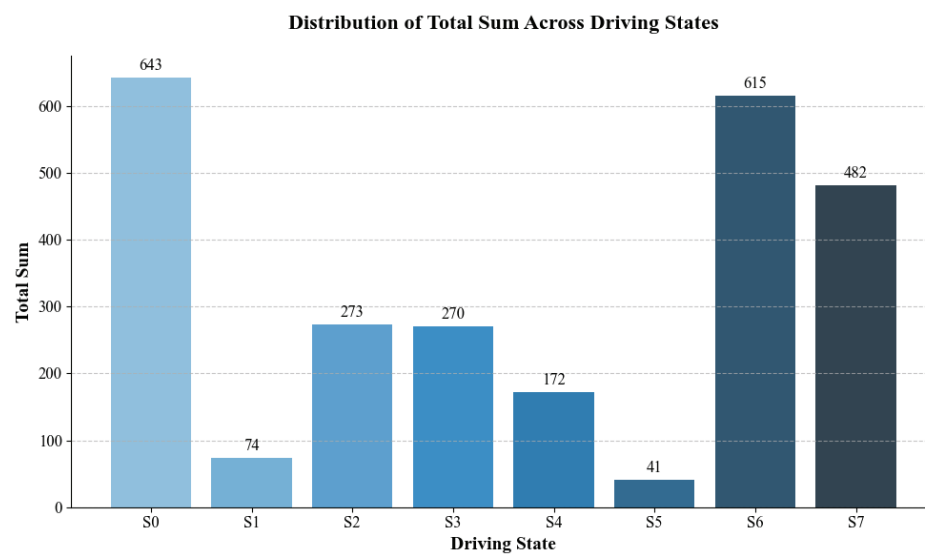


Figure 6. The sum for each driving state.

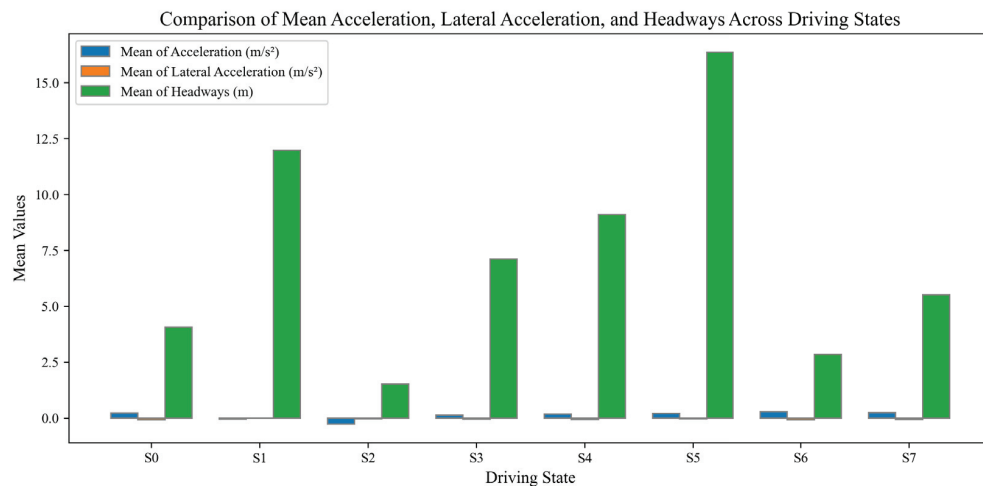


Figure 7. The variable values each driving state.

5.3. Graph Analysis Results

5.3.1. Graph Construction and State Node Transition Analysis

To conduct a detailed study of the conflict process, we can explore the transfer patterns of driving states during the conflict process. Firstly, we need to count the number of transitions between driving states and construct a state transition matrix. We conducted statistics on the state transitions of each driving state sequence. For example, if a sample transitions from state S1 to state S4, add 1 to the position corresponding to the co-occurrence matrix. Then, we obtain the transition matrix between driving states by a co-occurrence matrix that is obtained through statistics, which is shown in Table 4.

Table 4. Driving state transition co-occurrence matrix.

	S0	S1	S2	S3	S4	S5	S6	S7
S0	437	0	8	0	0	0	120	20
S1	0	47	0	0	27	0	0	0
S2	0	0	194	0	0	0	19	0
S3	2	0	0	171	4	0	0	89
S4	0	0	0	53	118	0	0	0
S5	0	13	0	0	0	28	0	0
S6	32	0	47	0	0	0	426	0
S7	122	0	0	15	0	0	0	321

From Table 4, it can be observed that the key nodes (key driving states) are driving states 0, 2, 6, and 7. These key driving states are critical during conflict process because they have a more significant page rank and closeness centrality. During the conflict process, the driving state continues to change, and during the transformation process, it will either go through these critical key states or ultimately stay in them. Further, based on the state transition relationships in the co-occurrence matrix, we indirectly convert the conflict process into a graph structure.

In Figure 8, each node represents a driving state, and the weight of the edge is the number of transitions between two states. Next, we use complex network theory to analyze the graph structure to obtain key nodes (critical driving states) and determine whether the driving states' transition process satisfies characteristics such as the small-world property.

In Figure 8, a thicker edge indicates more transitions, while a thinner edge indicates fewer transitions. We can also observe the patterns of state transitions. When the vehicle drives in the same state, the edge forms a self-loop. However, when the self-loop is broken

and the current driving state transfers to another driving state, the driving state undergoes a sudden change. Here, “Free movement” generally refers to a vehicle moving slowly and steadily behind the vehicle ahead and “Sudden braking” refers to sudden emergency braking while following the vehicle ahead, which is often the result of congestion spreading from the lead vehicle to the rear.

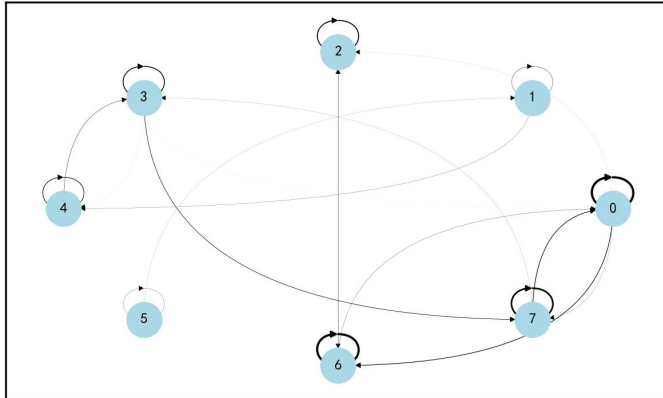


Figure 8. Driving state transition graph structure.

From the thickness of the edges in Figure 6, S3–S7–S0–S6–S2 is the most prominent transition path. After maintaining S6 for a period, some vehicles transfer to S2, which is a state of emergency braking.

In this path, the headway continuously decreases. As the headway gradually decreases from 7 m to less than 1.5 m, it will likely transfer to a more dangerous driving state 2. This situation is often caused by incorrect acceleration behavior in the early stage (such as aggressive acceleration, cutting in, etc.), which leads to rapid deceleration in the later stage. Figure 9 shows the changing trends of the three variables in the state. It can be seen that as the front of the vehicle continues to decrease, the acceleration of the vehicle first increases and then decreases, and there is a trend of turning the steering wheel.

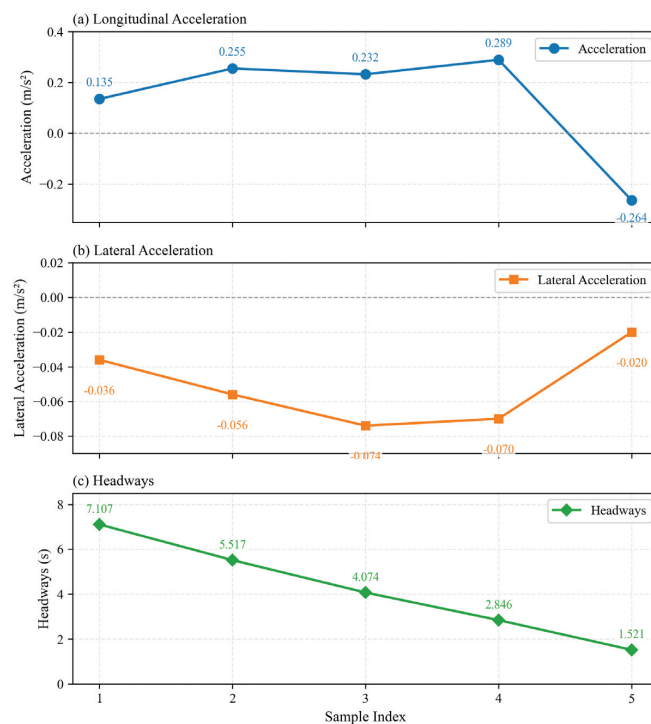


Figure 9. The changing trends of the three variables in the state.

To prevent conflicts, this state transition path can be blocked in advance by always maintaining a headway of at least 6 m when following the front vehicle in a congested section on the highways.

5.3.2. Graph Node Importance and Small-World Characteristics

Through graph structure analysis, some key nodes (important driving states) and important state transition patterns can be obtained. Table 5 shows the graph network characteristics of each driving state node.

Table 5. Closeness centrality and page rank of driving status nodes.

	Closeness Centrality	Page Rank
S0	0.50	0.18
S1	0.14	0.07
S2	0.39	0.17
S3	0.44	0.11
S4	0.39	0.10
S5	0.00	0.04
S6	0.39	0.21
S7	0.44	0.13

Figure 10 shows the closeness centrality and page rank of each state, where S0, S2, and S3 are the states with higher closeness centrality and page rank, indicating that these states are crucial in the conflict process.

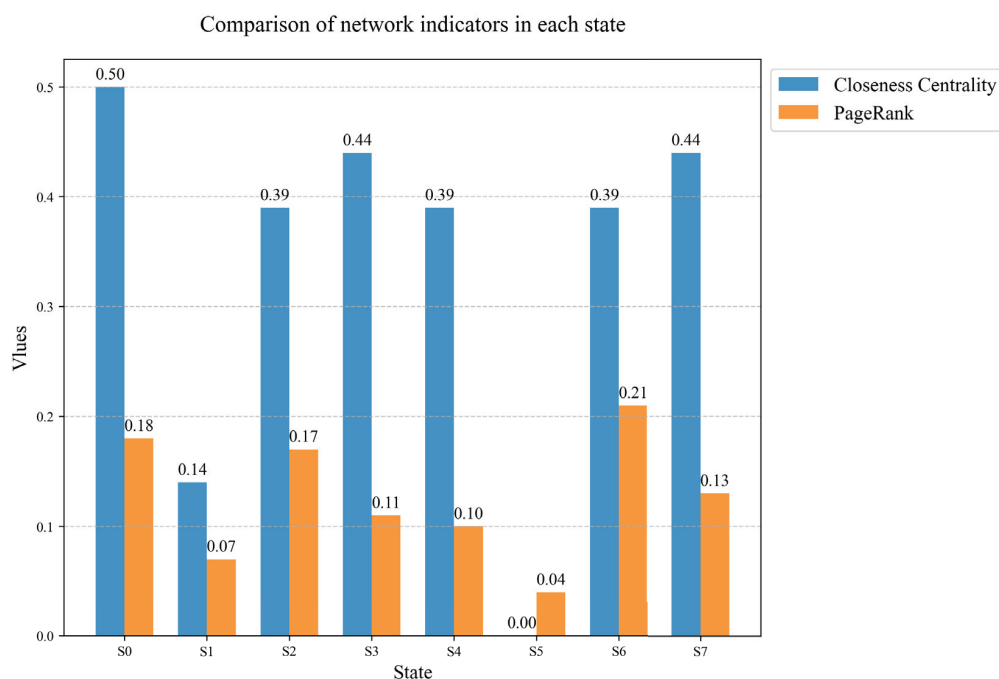


Figure 10. Closeness centrality and page rank comparison of each state.

In addition, the state transition graph of the conflict process does not have a small-world property. So, the following conclusions can be inferred: the path of driving state transitions is lengthy, suggesting that multiple driving states must be experienced before reaching the final conflict. This indicates a slow transition between driving states, leading drivers unable to perceive dangerous headway early enough. Prior to the final conflict, the vehicle consistently remains in a self-loop state before abruptly entering the conflict state.

5.4. Markov Characteristic Analysis Results

Furthermore, we transform the driving state transition matrix (Table 4) into a state probability transition matrix, as shown in Table 6.

Table 6. States' probability transition matrix.

	S0	S1	S2	S3	S4	S5	S6	S7
S0	0.747	0.000	0.013	0.000	0.000	0.000	0.205	0.034
S1	0.000	0.635	0.000	0.000	0.365	0.000	0.000	0.000
S2	0.000	0.000	0.911	0.000	0.000	0.000	0.089	0.000
S3	0.007	0.000	0.000	0.643	0.015	0.000	0.000	0.335
S4	0.000	0.000	0.000	0.310	0.690	0.000	0.000	0.000
S5	0.000	0.317	0.000	0.000	0.000	0.683	0.000	0.000
S6	0.063	0.000	0.093	0.000	0.000	0.000	0.844	0.000
S7	0.266	0.000	0.000	0.033	0.000	0.000	0.000	0.701

Next, we calculate the power of this state transition matrix. After the 39th time, the matrix gradually reaches a stable state. The stable matrix is shown in Table 7.

Table 7. Matrix under power stability.

	S0	S1	S2	S3	S4	S5	S6	S7
S0	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150
S1	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150
S2	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150
S3	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150
S4	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150
S5	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150
S6	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150
S7	0.1198	0.0000	0.4500	0.0015	0.0001	0.0000	0.4100	0.0150

This indicates that the driving state transition process during the conflict process corresponds to the Markov property, and the conflict process can be described using a Markov chain. Furthermore, when the power of the matrix reaches stability, it ultimately remains in the four key driving states analyzed above, i.e., 0, 2, 6, and 7, and it is most likely to stay in driving states 2 and states 6. And the calculated p value of the chi square statistic is close to 0, which is much smaller than the given error probability α of 0.05. We can further reject the zero hypothesis; the conflict process satisfies Markov properties. This indicates that the driving state eventually tends to stabilize during the conflict process, and the driving state is predictable.

6. Conclusions

In this paper, a framework based on graph theory was proposed to analyze the time-varying patterns of time-series variables. And a fine-grained and detailed analysis of the rear-end conflict process on the congested highway sections was conducted using this framework. Specifically, we based our analysis on the micro-trajectory time series in the conflict process and used complex network theory to analyze the driving behavior switch time-varying patterns during the conflict process. The main results are as follows:

1. The transition of driving behavior states during the conflict process conforms to the Markov property, which means that the driving behavior during the conflict process is predictable and can be altered.

2. The network of the driving behavior state transition graph does not conform to the small-world property. This means that before the final conflict occurs, high-risk vehicles always continuously switch states between a few states or remain in a self-looping state until they abruptly transition into the final conflict state.
3. On congested highway sections, rear-end collisions often transition to a conflict state along a fixed path within the graph structure, indicating that the majority of conflicts stem from similar causes.

Based on the careful analysis of traffic conflicts under congested traffic flow, these conclusions can be used to develop targeted vehicle auxiliary warning systems. In the future, research on vehicle collision avoidance control based on this process analysis is worth further exploration.

Author Contributions: R.C. Writing—review and editing; Supervision. X.A. Writing—original draft; Methodology. W.F. Conceptualization; Validation, D.Z. Investigation; Data processing; Investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Program of Humanities and Social Science of Education Ministry of China (Grant No. 24YJA630013) and the APC was funded by the Ningbo Natural Science Foundation of China (Grant No. 2024J125).

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors report no declarations of interest.

References

1. Kuang, Y.; Qu, X. A review of crash surrogate events. In *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*; American Society of Civil Engineers: Reston, WV, USA, 2014; pp. 2254–2264.
2. Zheng, L.; Sayed, T.; Mannering, F. Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. *Anal. Methods Accid. Res.* **2021**, *29*, 100142. [CrossRef]
3. Guo, Y.; Liu, P.; Wu, Y.; Yang, M. Traffic conflict model based on Bayesian multivariate Poisson-lognormal normal distribution. *China J. Highw. Transp.* **2018**, *31*, 101–109.
4. Sacchi, E.; Sayed, T.; Deleur, P. A comparison of collision-based and conflict-based safety evaluations: The case of right-turn smart channels. *Accid. Anal. Prev.* **2013**, *59*, 260–266. [CrossRef]
5. Tarko, A.P. A unifying view on traffic conflicts and their connection with crashes. *Accid. Anal. Prev.* **2021**, *158*, 106187. [CrossRef]
6. Fu, C.; Sayed, T. Bayesian dynamic extreme value modeling for conflict-based real-time safety analysis. *Anal. Methods Accid. Res.* **2022**, *34*, 100204. [CrossRef]
7. Fu, C.; Sayed, T. A multivariate method for evaluating safety from conflict extremes in real time. *Anal. Methods Accid. Res.* **2022**, *36*, 100244. [CrossRef]
8. Sun, Z.; Chen, Y.; Wang, P.; Fang, S.; Tang, B. Vision-based traffic conflict detection using trajectory learning and prediction. *IEEE Access* **2021**, *9*, 34558–34569. [CrossRef]
9. Wang, L.; Zou, L.; Abdel-Aty, M.; Ma, W. Expressway rear-end crash risk evolution mechanism analysis under different traffic states. *Transp. B Transp. Dyn.* **2023**, *11*, 510–527. [CrossRef]
10. Yu, R.; Han, L.; Zhang, H. Trajectory data based freeway high-risk events prediction and its influencing factors analyses. *Accid. Anal. Prev.* **2021**, *154*, 106085. [CrossRef]
11. Yuan, C.; Li, Y.; Huang, H.; Wang, S.; Sun, Z.; Li, Y. Using traffic flow characteristics to predict real-time conflict risk: A novel method for trajectory data analysis. *Anal. Methods Accid. Res.* **2022**, *35*, 100217. [CrossRef]
12. Ji, Q.; Lyu, H.; Yang, H.; Wei, Q.; Cheng, R. Bifurcation control of solid angle car-following model through a time-delay feedback method. *J. Zhejiang Univ.-Sci. A* **2023**, *24*, 828–840. [CrossRef]
13. Hu, J.; Bai, J.; Yang, J.; Lee, J.J. Crash risk prediction using sparse collision data: Granger causal inference and graph convolutional network approaches. *Expert Syst. Appl.* **2025**, *259*, 125315. [CrossRef]
14. Zhong, H.; Wang, L.; Su, Z.; Liu, G.; Ma, W. Characteristics identification and evolution patterns analyses of road chain conflicts. *Accid. Anal. Prev.* **2024**, *195*, 107395. [CrossRef]
15. Li, L.; Lyu, H.; Wang, T.; Cheng, R. STdi4DMPC: Distributed Model Predictive Control for Connected and Automated Truck Platoon with Mixed Traffic Flow Based on Spatiotemporal Trajectory Prediction. *IEEE Trans. Veh. Technol.* **2024**, *73*, 14563–14579. [CrossRef]

16. Hu, Y.; Li, Y.; Huang, H. Spatio-temporal dynamic change mechanism analysis of traffic conflict risk based on trajectory data. *Accid. Anal. Prev.* **2023**, *191*, 107203. [CrossRef] [PubMed]
17. Chen, B.; Zhao, X.; Li, Y.; Liu, X. Exploring the associations of demographics and scale measures with cognitive driving behavior among older drivers in China. *Accid. Anal. Prev.* **2024**, *200*, 107542. [CrossRef]
18. Xie, N.; Yu, R.; He, Y.; Li, H.; Li, S. Unveiling pre-crash driving behavior common features based upon behavior entropy. *Accid. Anal. Prev.* **2024**, *196*, 107433. [CrossRef]
19. Arun, A.; Haque, M.; Bhaskar, A.; Washington, S.; Sayed, T. A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. *Anal. Methods Accid. Res.* **2021**, *32*, 100180. [CrossRef]
20. Arun, A.; Haque, M.; Bhaskar, A.; Washington, S.; Sayed, T. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accid. Anal. Prev.* **2021**, *153*, 106016. [CrossRef]
21. Arun, A.; Haque, M.; Washington, S.; Mannering, F. A physics-informed road user safety field theory for traffic safety assessments applying artificial intelligence-based video analytics. *Anal. Methods Accid. Res.* **2023**, *37*, 100252. [CrossRef]
22. Cheng, R.; Ge, H.; Wang, J. An extended continuum model accounting for the driver's timid and aggressive attributions. *Phys. Lett. A* **2017**, *381*, 1302–1312. [CrossRef]
23. Arun, A.; Haque, M.; Washington, S.; Sayed, T.; Mannering, F. How many are enough?: Investigating the effectiveness of multiple conflict indicators for crash frequency-by-severity estimation by automated traffic conflict analysis. *Transp. Res. Part C Emerg. Technol.* **2022**, *138*, 103653. [CrossRef]
24. Cheng, R.; Lyu, H.; Zheng, Y.; Ge, H. Modeling and stability analysis of cyberattack effects on heterogeneous intelligent traffic flow. *Phys. A Stat. Mech. Appl.* **2022**, *604*, 127941. [CrossRef]
25. Cao, J.; Chen, J.; Guo, X.; Wang, L. Trajectory data-based severe conflict prediction for expressways under different traffic states. *Phys. A Stat. Mech. Appl.* **2023**, *621*, 128595. [CrossRef]
26. Choi, J.G.; Kong, C.W.; Kim, G.; Lim, S. Car crash detection using ensemble deep learning and multimodal data from dashboard cameras. *Expert Syst. Appl.* **2021**, *183*, 115400. [CrossRef]
27. Gu, X.; Abdel-Aty, M.; Xiang, Q.; Cai, Q.; Yuan, J. Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. *Accid. Anal. Prev.* **2019**, *123*, 159–169. [CrossRef]
28. Li, P.; Abdel-Aty, M.; Yuan, J. Real-time crash risk prediction on arterials based on LSTM-CNN. *Accid. Anal. Prev.* **2020**, *135*, 105371. [CrossRef] [PubMed]
29. Li, L.; Lin, Y.; Du, B.; Yang, F.; Ran, B. Real-time traffic incident detection based on a hybrid deep learning model. *Transp. A Transp. Sci.* **2022**, *18*, 78–98. [CrossRef]
30. Li, Q.; Cheng, R.; Ge, H. Short-term vehicle speed prediction based on BiLSTM-GRU model considering driver heterogeneity. *Phys. A Stat. Mech. Appl.* **2023**, *610*, 128410. [CrossRef]
31. Ye, Y.; Xiao, Y.; Zhou, Y.; Li, S.; Zang, Y.; Zhang, Y. Dynamic multi-graph neural network for traffic flow prediction incorporating traffic accidents. *Expert Syst. Appl.* **2023**, *234*, 121101. [CrossRef]
32. Lu, J.; Grembek, O.; Hansen, M. Learning the representation of surrogate safety measures to identify traffic conflict. *Accid. Anal. Prev.* **2022**, *174*, 106755. [CrossRef]
33. Sattar, K.; Oughali, F.C.; Assi, K.; Ratrou, N.; Jamal, A.; Rahman, S.M. Transparent deep machine learning framework for predicting traffic crash severity. *Neural Comput. Appl.* **2023**, *35*, 1535–1547. [CrossRef]
34. Zhang, S.; Abdel-Aty, M.; Cai, Q.; Li, P.; Ugan, J. Prediction of pedestrian-vehicle conflicts at signalized intersections based on long short-term memory neural network. *Accid. Anal. Prev.* **2020**, *148*, 105799. [CrossRef] [PubMed]
35. Wang, T.; Cheng, R.; Wu, Y. Stability analysis of heterogeneous traffic flow influenced by memory feedback control signal. *Appl. Math. Model.* **2022**, *109*, 693–708. [CrossRef]
36. Cheng, Z.; Zu, Z.; Lu, J. Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. *Sustainability* **2018**, *11*, 160. [CrossRef]
37. Li, J.; Cheng, R. A real-time adaptive signal control method for multi-intersections in mixed connected vehicle environments. *J. Zhejiang Univ.-Sci. A Appl. Phys. Eng.* **2025**, *1*, 189.
38. Nasr, H.A.; Huang, H.; Jin, J.; Zhou, H. Rear-end conflicts analysis at non-signalized intersection based on vehicles trajectory data. *Eur. Transp. Res. Rev.* **2025**, *17*, 27. [CrossRef]
39. Sheikh, M.S.; Peng, Y. Assessment of Rear-end Collision Risk based on a Deep Reinforcement Learning Technique: A Break Reaction Assessment Approach. *IEEE Access* **2025**, *13*, 20171–20190. [CrossRef]
40. Li, Y.; Wu, D.; Lee, J.; Yang, M.; Shi, Y. Analysis of the transition condition of rear-end collisions using time-to-collision index and vehicle trajectory data. *Accid. Anal. Prev.* **2020**, *144*, 105676. [CrossRef]
41. Luo, Q.; Chen, X.; Yuan, J.; Zang, X.; Yang, J.; Chen, J. Study and Simulation Analysis of Vehicle Rear-End Collision Model considering Driver Types. *J. Adv. Transp.* **2020**, *2020*, 7878656. [CrossRef]
42. Li, Y.; Wang, X.; Wang, T.; Liu, Q. Characteristics analysis of autonomous vehicle pre-crash scenarios. *arXiv* **2025**, arXiv:2502.20789.

43. Huang, C.; Wen, X.; He, D. Characteristics of rear-end collisions: A comparison between automated driving system-involved crashes and advanced driving assistance system-involved crashes. *Transp. Res. Rec.* **2024**, 2678, 771–782. [CrossRef]
44. Peng, Y.; Liu, D.; Wu, S.; Yang, X.; Wang, Y.; Zou, Y. Enhancing Mixed Traffic Flow with Platoon Control and Lane Management for Connected and Autonomous Vehicles. *Sensors* **2025**, 25, 644. [CrossRef] [PubMed]
45. Xing, Y.; Wu, Y.; Wang, H.; Wang, L.; Li, L.; Peng, Y. Failed lane-changing detection and prediction using naturalistic vehicle trajectories. *Transp. Res. Part C Emerg. Technol.* **2025**, 170, 104939. [CrossRef]
46. Zhang, Y.; Wang, C.; Yu, R.; Wang, L.; Quan, W.; Gao, Y.; Li, P. The ad4che dataset and its application in typical congestion scenarios of traffic jam pilot systems. *IEEE Trans. Intell. Veh.* **2023**, 8, 3312–3323. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Analyzing New Operation Strategy of Demand-Responsive Transports Using Discrete-Event Simulation Framework

Seung-Wan Cho ¹, Yeong-Hyun Lim ¹, Seong-Hyeon Ju ¹ and Kyung-Min Seo ^{2,*}

¹ Department of Industrial & Management Engineering, Hanyang University, 222, Wangsimni-ro, Seong-dong-gu, Seoul 04763, Republic of Korea; swcho@nsl.hanyang.ac.kr (S.-W.C.); yhl@nsl.hanyang.ac.kr (Y.-H.L.); shju@nsl.hanyang.ac.kr (S.-H.J.)

² Department of Industrial & Management Engineering, Hanyang University ERICA, 55, Hanyangdaehak-ro, Sangnok-gu, Ansan 15588, Republic of Korea

* Correspondence: kmseo@hanyang.ac.kr

Abstract: Demand-responsive transport (DRT) provides flexible ride-sharing by dynamically adjusting routes based on real-time user demand, making it suitable for complex urban mobility needs. This study proposes a modular simulation framework based on the DEVS (Discrete Event System Specification) formalism and introduces an “express service” strategy that enables direct trips without intermediate stops. The framework supports scenario-based analysis using key performance indicators (KPIs) and allows for flexible testing of operational strategies. Two experiments were conducted: the first validated the simulation model under varying demand and fleet conditions; and the second assessed the impact of the express service. Results showed that express passengers experienced significantly shorter waiting and riding times, while standard passenger service remained stable. The strategy also improved operational efficiency under constrained resources. This study contributes to a configurable simulation platform for evaluating differentiated DRT services and provides practical insights for adaptive service planning, especially in urban settings where tiered mobility solutions are increasingly needed.

Keywords: dispatching; simulation; on-demand mobility; ride-sharing; DEVS (discrete event system specification)

1. Introduction

Demand-responsive transport (DRT) is a form of shared mobility that dynamically adjusts vehicle routes based on users’ real-time demands rather than following a fixed schedule [1–3]. The flexibility of DRT makes it an increasingly sought-after transportation option globally, as it effectively caters to the needs of expanding populations and increasing urban population densities [4,5]. Such tailor-made DRT services have the potential to significantly offset reliance on conventional public transport systems and provide more dynamic planning [6,7]. This shift towards demand-responsive operations not only accommodates the unique mobility needs of urban dwellers, but also promotes a more efficient use of road space [8]. By facilitating ride-sharing based on real-time demand, DRT can substantially decrease the overall number of vehicles in operation, ease urban congestion, and contribute to sustainable transportation [9,10].

Efforts to optimize DRT operations have been undertaken using various approaches, particularly focusing on the dynamic allocation of vehicles in response to demand [11,12]. This has led to the development of dispatch algorithms to efficiently match vehicles with passenger requests [13,14]. To assess and evaluate such newly developed strategies, the

creation of a foundational simulation model is necessary [15]. DRT simulation research can be broadly categorized into two types: tool- and methodology-based simulation studies [16–18]. Both approaches offer valuable insights: tool-based simulations benefit from maturity and scalability of existing software, while methodology-based simulations offer greater flexibility to incorporate novel rules and system behaviors [19]. In this study, we pursue a methodology-based simulation approach. This choice is motivated by the need to easily integrate and evaluate a new operational strategy for DRT, which would be difficult to test in a black-box commercial simulator.

This study introduces a new operational strategy and methodology-based simulation framework designed to enhance the efficiency of DRT systems. A new operational strategy called express service allows for the direct transportation of passengers to their destinations without intermediate stops, significantly enhancing user convenience and efficiency. The simulation models were developed using the DEVS (Discrete Event System Specification) formalism [20], which is particularly well-suited for modeling the stochastic, event-driven nature of DRT operations—such as customer calls, passenger boardings, and drop-offs. Leveraging the inherent modularity and hierarchy of DEVS, we structured the DRT simulation into independently functioning components, including demand generation, vehicle movement, and routing/dispatch logic. Each of these components can be developed, replaced, or extended without modifying the entire system. This modular design enables flexible testing of new operational strategies and facilitates seamless integration of additional services such as tiered DRT offerings or specialized shuttles. Furthermore, the hierarchical organization supports scalable modeling across different service layers or geographical regions, improving both the adaptability and reusability of the simulation framework [21,22]. This approach not only ensures a realistic representation of DRT processes, but also enhances the framework’s applicability to a wide range of scenarios, making it a robust tool for both theoretical analysis and practical service planning.

The proposed simulation framework is structured into four phases: data pre-processing, simulation modeling, operational strategy identification, and simulation data analysis. The pre-processor is subdivided into three essential functions: mapping the network based on foundational data prior to the simulation, categorizing the demand data, and selecting essential stops for the simulation. In this study, we present a DEVS-based simulation model that further branches into an experimental framework for conducting and analyzing experiments, a control system for managing operations, and a physical system designed for modeling both DRT and passengers. Furthermore, the experimental framework incorporates a visualizer model, enabling real-time observation of the experimental proceedings. This feature aids in tracking the progress of the experiment with clarity and precision. Based on this model, we propose ‘express service’ operation strategy aimed at optimizing efficiency for passengers and standard DRT operations. This strategy targets specific guests and increases efficiency by not allowing other passengers to pick up or drop off between their origin and destination.

We propose a set of key performance indicators (KPIs) to evaluate and analyze the data and operational strategies thoroughly. Using the proposed framework, we conducted two simulation experiments. The first experiment validated the baseline DRT model by simulating standard operations without the express service. Results showed expected trends in passenger wait and ride times under varying demand and vehicle availability, confirming the model’s reliability. The second experiment evaluated the impact of the proposed express service strategy, which enables selected trips to bypass intermediate stops. Simulation results showed that express passengers experienced significantly shorter wait and ride times, while standard passengers’ service levels remained stable. These findings demonstrate that the express service can enhance user experience without compromising

overall system performance, highlighting the practical value of the proposed strategy and the simulation framework [23].

The remainder of this paper is organized as follows: Section 2 reviews related to research on DRT and simulation methods. Section 3 explains the simulation framework and the DEVS-based model. Section 4 presents the experimental setup and results. Section 5 concludes the paper and discusses implications and future research.

2. Related Work

Simulation-based studies on demand-responsive transport (DRT) systems can be broadly categorized into two types: those using existing simulation tools (tool-based) and those developing custom modeling frameworks (methodology-based). While tool-based approaches emphasize accessibility and usability through predefined components, methodology-based approaches often offer higher flexibility and extensibility for tailored experimentation. Table 1 summarizes representative studies from both categories, highlighting their modeling methods, research goals, and key limitations to help position our work within this evolving research landscape.

Table 1. Comparison of simulation studies on demand-responsive transport (DRT) systems.

Related Research	Approach	Modeling Method	Problems to Be Solved	Limitations
David et al. (2023) [24]	Tool-based	SUMO-based tool	Assignment and route optimization	Static demand; not real-time
Ronald et al. (2017) [25]	Tool-based	Delphi, SUMOoD, MATSim	Comparison of DRT simulation tools	No dynamic demand modeling; limited scenario specificity
Kagho et al. (2021) [26]	Tool-based	MATSim	DRT in car-dependent areas	Limited control over logic; high vehicle mileage
Fielbaum et al. (2024) [27]	Methodology-based	Hybrid route planning	Combine DRT and fixed routes	Fixed routing; lacks re-routing flexibility
Kim et al. (2022) [28]	Methodology-based	Event-based simulation	KPI analysis of DRT systems	No visualization; no service-type comparison
Alomrani et al. (2023) [29]	Methodology-based	Bipartite matching	Real-time dispatch optimization	Lacks service-tier differentiation; not modular
Wang et al. (2022) [30]	Methodology-based	Soft windows and compensation	Tiered service and user satisfaction	High model complexity; user assumptions
Our study	Methodology-based	DEVS formalism	New DRT service (Express service) analysis	No real-world validation; synthetic demand used

Tool-based approaches typically rely on established simulation platforms such as SUMO, MATSim, or other agent-based environments that facilitate modeling of urban mobility systems. For example, David et al. (2023) developed a SUMO-based tool that computes vehicle assignments and route planning using static demand inputs [24]. While effective in structured experimental conditions, this tool lacks responsiveness to real-time demand fluctuations, limiting its applicability to dynamic urban systems. Ronald et al. (2017) compared several DRT simulation tools—including Delphi, SUMOoD, and MATSim—focusing on their

ease of use and integration capabilities [25]. However, the study did not explore scenario-specific demand structures or dispatching algorithms, making it less suitable for evaluating new service models. Similarly, Kagho et al. (2021) employed MATSim to simulate DRT for car-dependent suburban regions and demonstrated potential benefits in accessibility [26]. Still, issues such as excessive vehicle mileage and limited scheduling adaptability were noted. While these tools are valuable for initial prototyping or policy simulation, their black-box nature often limits the flexibility required to test novel operational strategies such as tiered or express services.

Methodology-based studies, in contrast, focus on building models from the ground up with greater control over system behavior, often incorporating discrete-event logic, optimization techniques, or hybrid system architectures. Fielbaum et. al (2024) proposed a hybrid model that integrates fixed-route planning with demand-responsive service to improve network efficiency [27]. However, the resulting route structure becomes fixed after optimization, limiting its responsiveness to temporal or spatial demand variability. Kim et al. (2022) constructed an event-based simulation that captures the stochastic nature of passenger demand and evaluates key performance indicators (KPIs) of DRT systems. Despite its strengths in KPI tracking, the study did not incorporate modular structures or service-type variations [28]. Recent studies such as Alomrani et al. (2023) proposed graph-based bipartite optimization for real-time dispatching, improving matching efficiency, but without differentiating service levels between passengers [29]. Wang et al. (2022) simulated soft time windows and compensation policies to examine service differentiation, but implemented these within a fixed optimization context, limiting reusability across alternative strategies or geographic contexts [30].

Table 1 offers a comparative summary of these representative works. While prior research has contributed significantly to understanding DRT design and operational policies, many focus on static or specialized configurations that are not easily adapted for comparative testing of new service strategies. Very few studies explicitly compare standard shared-ride models with express or prioritized service variants within a modular simulation environment that supports flexible experimentation and performance analysis.

In contrast, this study introduces a simulation framework based on the DEVS formalism, which enables modular, hierarchical, and event-driven modeling. The proposed framework allows for plug-and-play updates of key components, such as routing algorithms and dispatch policies, making it suitable for dynamic policy testing. Our model explicitly supports tiered service differentiation by assigning express passengers direct, non-stop routes, enabling controlled experiments on resource allocation and user prioritization. Unlike prior models tightly coupled to fixed optimization goals, this framework enables scenario-based, KPI-driven comparisons of multiple operational strategies under realistic constraints. This approach addresses several limitations in existing DRT simulation studies and provides a practical platform for researchers and planners to explore service innovation in diverse urban contexts.

3. Proposed Method

This section describes the proposed method and overall structure of the system. The system is broadly divided into the pre-processor and the proposed models. Notably, the proposed model is designed as a discrete-event system that can be abstracted through complexity and logic to suit the purpose of the simulation. Additionally, considering an actual mobile system, a simulation model was designed with a hierarchical structure. Furthermore, it has a modular structure that makes it easy to replace and add schedules or other control systems.

The overall structure of the system is shown in Figure 1. The pre-processors are segmented into three principal phases: map networking, demand data categorization, and essential stop selection. These phases constitute a pre-processing phase that transforms various data into a format amenable to mobility simulations before they are input into the proposed model. The three pre-processors are described in Section 4.2.

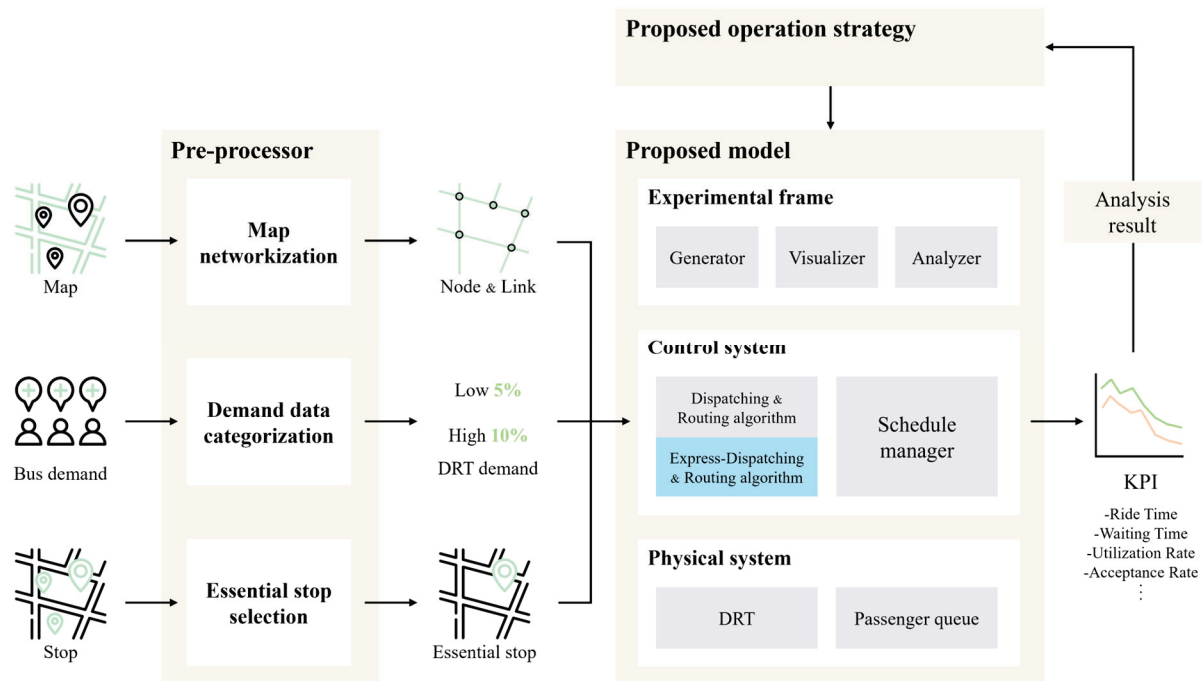


Figure 1. The overall structure of the proposed simulation framework.

The pre-processed data are subsequently input into the proposed simulation model, which is structured around three primary components—an Experimental Frame, a Control System, and a Physical System—delineated through an analysis of the mobility field. This model facilitates simulations based on the interactions among its sub-models, each of which is assigned a specific function. Through these interactions, simulation outcomes were generated and analyzed. Detailed discussions on the underlying algorithms and model frameworks are provided in the following sections: Section 3.1 elucidates the foundation and express dispatching algorithm, and Section 3.2 delves into the design of each model and its corresponding sub-models.

The experiment implemented the operational strategy proposed in this study within the developed model from which the KPIs were calculated. On the passenger side, the KPIs encompass ride time and waiting time, whereas the DRT side focuses on call acceptance and utilization rates. The experimental scenario is outlined in Section 4.1. The definitions of the resultant KPIs are provided in Section 4.3, and the analysis of these results is further elaborated in the subsections of Sections 4.4 and 4.5.

The proposed framework encompasses a model structure that is conducive to the development of various algorithms that initiate data pre-processing. This facilitated the simulation of diverse operational strategies, thereby enabling a thorough analysis of the results. This approach ensures a comprehensive examination of how different strategies impact the overall performance, allowing for optimized decision-making based on the simulated outcomes.

3.1. DRT Dispatching Algorithm

One of the most critical considerations in shared vehicle services such as demand-responsive transport systems is the assignment of vehicles to passengers. Figure 2 illustrates the dispatching scenario for a demand-response vehicle when a new service call is received. In this scenario, ‘S+’ represents the pick-up stop for Passenger S, and ‘S-’ denotes the drop-off stop for the same passenger. Currently, DRT X operates on route [A+, B+, A-, B-], whereas DRT Y serves route [C+, C-]. Upon receiving a new service request from Passenger D, the DRT control system must determine whether to assign the call to DRT X or DRT Y and identify an optimal insertion point within the existing routes.

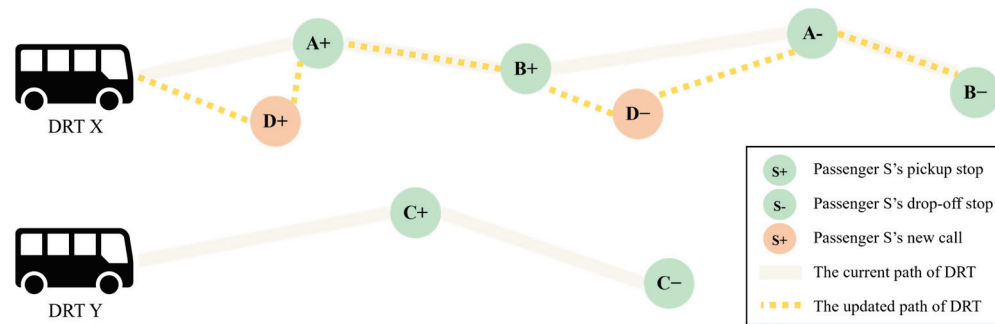


Figure 2. Dispatching and routing scenario for a new call (standard call).

Now, we introduce a novel service designed for periods of low vehicle demand or when excess vehicles are available. This service, tailored for passengers, ensures direct transportation to one's destination without intermediate stops upon request for demand-responsive transport. Implementing this service can significantly enhance passenger satisfaction by minimizing both waiting and riding times, thereby reducing vehicle idle times [31]. For instance, DRT X in Figure 3 follows the [A+, B+, A-, B-] route, while DRT Y operates on the [C+, C-] route. Upon receiving a new request from express passenger D, DRT X's itinerary is adjusted to [A+, D+, D-, B+, A-, B-], facilitating direct transport to D's destination without altering the sequence to [D+, A+, B+, D-, A-, B-]. This express service feature exemplifies how direct routing can be efficiently incorporated into existing schedules, prioritizing speed and convenience for passengers seeking immediate transportation to their destinations.

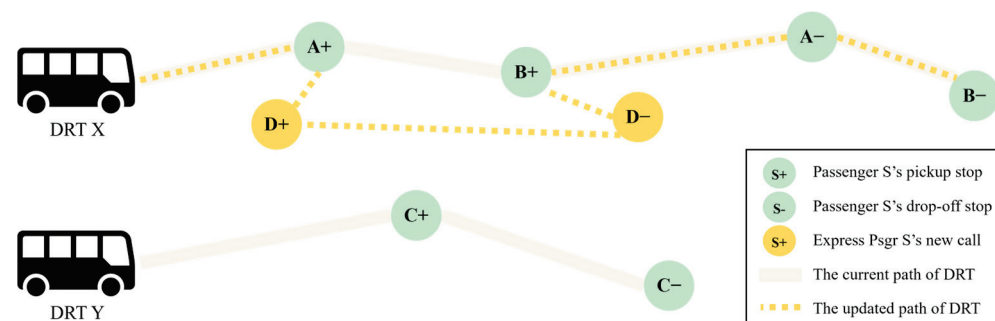


Figure 3. Dispatching and routing scenario for a new call (express call).

In the simulation model, vehicle assignment and routing are implemented using adapted versions of established algorithms, such as a greedy dispatching rule and a Dijkstra-based shortest path search [32–34]. These algorithms were customized to incorporate real-time operational constraints specific to DRT services, such as dynamic passenger requests and vehicle capacity limitations. While the fundamental logic of these baseline algorithms underpins standard operations, the express routing algorithm extends this structure by

adding constraints and optimization criteria tailored to express passengers—namely, direct routing without intermediate stops and prioritized insertion into vehicle schedules.

Modifications to the optimal insertion algorithm are necessary in the context of express DRT dispatching, which differs from the conventional approach used for standard passengers. Algorithm 1 illustrates the express optimal insertion algorithm, which aims to minimize both the waiting time and ride time for express passengers by prioritizing direct, non-stop service. To ensure this, specific constraints are applied when inserting the departure and arrival points of express passengers into the existing route. For instance, express passengers cannot be inserted before other express stops, and standard passengers cannot be inserted between the departure and arrival points of an express passenger.

Algorithm 1 *Pseudo Code of Express Optimal Insertion Algorithm*

```

1. //Initialization
2. best_path := NULL
3. best_total_time := ∞
4. insertable_positions := []
5. existingDst := targetDRT.existingDst
6. psgrCount := targetDRT.psgrCount

7. IF NOT expressLst[0] THEN
8.     insertable_positions.append((0, 0))
9. END IF

10. FOR i FROM 1 TO length(expressLst) DO
11.     IF NOT expressLst[i - 1] AND expressLst[i] THEN
12.         insertablePositions.append((i, i))
13.     ELSE IF NOT expressLst[i] THEN
14.         insertablePositions.append((i, i + 1))
15.     END IF
16. END FOR

17. insertablePositions.append((len(existing_dst), len(existing_dst)))

18. IF targetPsgr.psgrExpress THEN
19.     FOR EACH (i, j) IN insertablePositions DO
20.         temp_path := existingDst[0:i] + [targetDeparture] + [targetArrival] +
            existingDst[i:end]
21.         count_path := psgrCount [0:i] + [+targetPsgr.psgrNum] + [-targetPsgr.psgrNum]
            + psgrCount [i:end]
22.         possible_count := curPsgrNum
23.         over_capacity := FALSE

```

Algorithm 1 Cont.

```

24.  FOR  $k$  FROM 0 TO  $\text{length}(\text{count\_path}) - 1$  DO
25.     $\text{possibleCount} := \text{possibleCount} + \text{count\_path}[k]$ 
26.    IF  $\text{possibleCount} > \text{DRTMax}$  THEN
27.       $\text{over\_capacity} := \text{TRUE}$ 
28.      BREAK
29.    END IF
30.  END FOR

31.  IF  $\text{over\_capacity}$  THEN
32.    CONTINUE
33.  END IF

34.   $\text{total\_time} := \text{find\_shortest\_path}(\text{temp\_path})$ 
35.    IF  $\text{total\_time} < \text{best\_total\_time}$  THEN
36.       $\text{best\_total\_time} := \text{total\_time}$ 
37.       $\text{best\_path} := \text{temp\_path}$ 
38.    END IF
39.  END FOR
40. END IF
41. RETURN  $\text{best\_point}$ 

```

The algorithm evaluates all possible insertion points that satisfy these constraints and computes the corresponding passenger loads along the route to ensure vehicle capacity is not exceeded. Among the feasible insertion paths, the one with the shortest overall riding time is selected as optimal. Although the algorithm has the worst-case time complexity of $O(n^3)$ due to repeated evaluation and shortest-path calculations, its computational impact is limited in practice. This is because the algorithm is applied only when an express passenger request is generated, which represents a small portion of overall system activity. Additionally, the number of destinations per DRT is typically moderate in real-world operations, keeping execution time reasonable even under high-demand conditions.

3.2. System Modeling

A mobility system characterized by its response to passenger, call, and ride-off events can be effectively modeled as a discrete-event system. This approach allows for the intricate interactions and logic inherent in such systems to be represented despite their complexity. In event-based modeling, it is crucial to abstract certain characteristics to align with the goals of the simulation. Acknowledging this, we propose a mobility simulation model that leverages DEVS formalism. This model is designed for adaptability, enabling its reconfiguration into a hierarchical and modular setup to accommodate various operational strategy simulations. This modifiable structure not only facilitates the exploration of the system's behavior under different conditions, but also enhances the model's utility across diverse simulation scenarios.

Figure 4 illustrates the simplified architecture of the DEVS-based coupled model, which is segmented into three main components: the experimental frame, control system, and physical system. These components interact via well-defined input/output ports and message flows, supporting the modular and hierarchical structure of the simulation. Each component internally consists of atomic sub-models that handle specific tasks such as

dispatching, routing, and passenger interaction. To maintain clarity and focus, the following subsections highlight the most critical atomic models that govern key system behaviors and operational logic, including the implementation of the proposed express service.

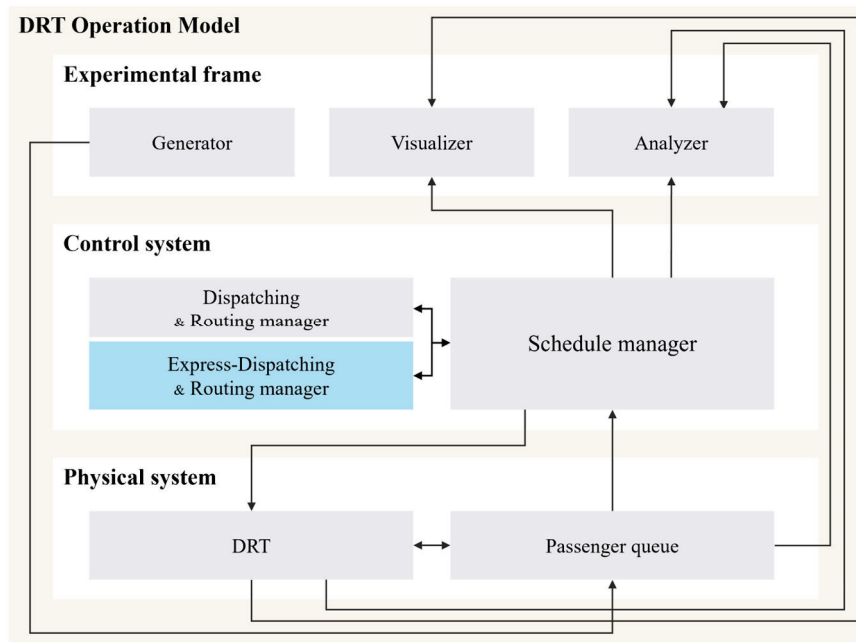


Figure 4. Overall system architecture based on DEVS coupled model.

DEVS Atomic Modeling

This section presents key DEVS atomic models that constitute the core behavior of the proposed simulation framework. While the complete system includes many atomic models, we focus here on the most essential ones: the Visualizer model from the experimental frame; the Schedule Manager and Dispatching and Routing Manager from the control system; and the DRT model from the physical system. These models were selected for their central roles in managing simulation execution, coordinating DRT operations, and representing system dynamics.

- Visualizer

The visualizer is tasked with receiving up-to-date schedule data from the DRT and Schedule Manager, and dynamically visualizes key information such as passenger boarding/dropping points and the DRT's current route in real time. The DEVS set is defined by Equations (1)–(8), with a distinctive characteristic being the presence of an input message set Y , while the internal transition functions δ_{int} , remains an empty set. This configuration begins in the WAIT state and functions by external transition functions set δ_{ext} , which are defined as functions. During this operation, the state of the simulation was visualized in real time, adhering to the rendering time specified in Equation (16), to ensure that the dynamics of the simulation were accurately reflected as they unfolded. Initially, the visualizer remains in the WAIT state and awaits the command. Upon receiving pair commands from the Schedule Manager, it shifts to the VISUALIZE state. In this state, the visualizer actively renders the schedule information contained in the messages for a period known as the rendering time, continuing to process and display the CurrentNode data from the DRT until a simDone message from the Schedule Manager signals the end of the visualization task.

$$Visualizer_{Atomic} = \langle X, Y, S, \delta_{ext}, \delta_{int}, \lambda, TA \rangle \quad (1)$$

$$X = \{Schedule, CurrentNode, simDone\} \quad (2)$$

$$Y = \{\}$$
 (3)

$$S = \{VISUALIZE, WAIT\}$$
 (4)

$$\delta_{ext} : \begin{cases} (WAIT) \times (Schedule) \rightarrow (VISUALIZE) \\ (VISUALIZE) \times (CurrentNode) \rightarrow (VISUALIZE) \\ (VISUALIZE) \times (simDone) \rightarrow (WAIT) \end{cases}$$
 (5)

$$\delta_{int} : \{\Phi\}$$
 (6)

$$\lambda : \{\Phi\}$$
 (7)

$$TA : \begin{cases} (WAIT) \rightarrow \infty \\ (VISUALIZE) \rightarrow Rendering\ time \end{cases}$$
 (8)

This visualization employed Python's TK Widget package (Tkinter 8.6) for implementation. As the simulation is initiated, the visualizer begins by loading the map data, displaying all nodes and paths, as depicted in Figure 5. Subsequently, upon receiving schedule information comprising the DRT Id, Current Node, Boarding Node, and Dropping Node from the Schedule Manager and DRT Atomic Model, the map is updated accordingly.



Figure 5. Implementation result of Visualizer model.

A closer look at Path A in Figure 5 reveals a route traversed by a circular DRT with two boarding stops, denoted by red + symbols, and two planned drop-off stops, marked with blue symbols. In addition, paths that have been previously traversed or have already expired are represented by dotted lines, whereas the currently scheduled path is depicted by a solid line. For instance, Path B in Figure 5 illustrates an initial plan for passengers dropping off to stop at the upper right side. However, the addition of new boarding and drop-off stops within the red circle indicates that the previously expired path is indicated by a dotted line, whereas the updated current path is displayed as a solid line. Through this visualization model, it becomes feasible to verify the effective development of dispatching and routing algorithms and monitor the progress of the simulation experiment more efficiently, ensuring a comprehensive overview of the system's dynamic operations.

The control system oversees and directs the operation of the entire system. It is segmented into the Schedule Manager and the Dispatching and Routing Manager. The Schedule Manager orchestrated a comprehensive scheduling process. Upon receiving a call

from a passenger, the system forwards a calculation request to the dispatching and routing managers to preliminarily assess the schedule. In response, the Dispatching and Routing Manager assumes the roles of vehicle dispatching and route planning, submitting a range of potential schedules to the Schedule Manager. Subsequently, the most fitting schedule is chosen, and this selected schedule is then assigned to the respective DRT, ensuring efficient and optimized transportation planning. By distinctly architecting the Schedule Manager and the Dispatching and Routing Manager, the system leverages the modular nature of DEVS to facilitate partial modifications, especially when updates or replacements of algorithms are necessary.

- Schedule Manager

The Schedule Manager is engineered to craft schedules utilizing incoming passenger calls and the real-time positions of the DRTs, as depicted in Figure 6 and outlined by Equations (9)–(16) in the DEVS set, which illustrates the operational diagram. At the beginning of the simulation, as expressed by Equation (12), the Schedule Manager remained in an IDLE state. It then transitioned to the DSP_PP state, as indicated by Equations (10) and (13), when it received a call message from the passenger queue. In the DSP_PP state, the Schedule Manager sends a DispatchRoute_Req message to the Dispatching and Routing managers. This step, aimed at assigning passengers to DRTs based on call specifics, precedes their return to the IDLE state, as expressed in Equations (11) and (14). The transition to the SCHEDULE state occurs upon the receipt of a candidate schedule from the Dispatching and Routing manager model, during which the scheduling information is relayed to the DRT model, as described in Equation (15). The Schedule Manager progresses to the ANALYZE state if a transit message from the DRT suggests that the last passenger has dropped off. Throughout these transitions, time shifts, indicated by Equation (16), occur instantaneously.

$$\text{Schedule Manager}_{Atomic} = \langle X, Y, S, \delta_{ext}, \delta_{int}, \lambda, TA \rangle \quad (9)$$

$$X = \{DispatchRoute_{Res}, Call, Transit\} \quad (10)$$

$$Y = \{DispatchRoute_{Req}, Schedule, simDone\} \quad (11)$$

$$S = \{IDLE, SCHEDULE, ANALYZE, DSP_{PP}\} \quad (12)$$

$$\delta_{ext} : \begin{cases} (IDLE) \times (Call) \rightarrow (DSP_{PP}) \\ (IDLE) \times (DispatchRoute_{Req}) \rightarrow (SCHEDULE) \\ (IDLE) \times (Transit) \rightarrow (ANALYZE) \end{cases} \quad (13)$$

$$\delta_{int} : \begin{cases} (DSP_{PP}) \rightarrow (IDLE) \\ (SCHEDULE) \rightarrow (IDLE) \\ (ANALYZE) \rightarrow (IDLE) \end{cases} \quad (14)$$

$$\lambda : \begin{cases} (DSP_{PP}) \rightarrow (DispatchRoute_{Req}) \\ (SCHEDULE) \rightarrow (Schedule) \\ (ANALYZE) \rightarrow (simDone) \end{cases} \quad (15)$$

$$TA : \begin{cases} (IDLE) \rightarrow \infty \\ (DSP_{PP}) \rightarrow 0 \\ (SCHEDULE) \rightarrow 0 \\ (ANALYZE) \rightarrow 0 \end{cases} \quad (16)$$

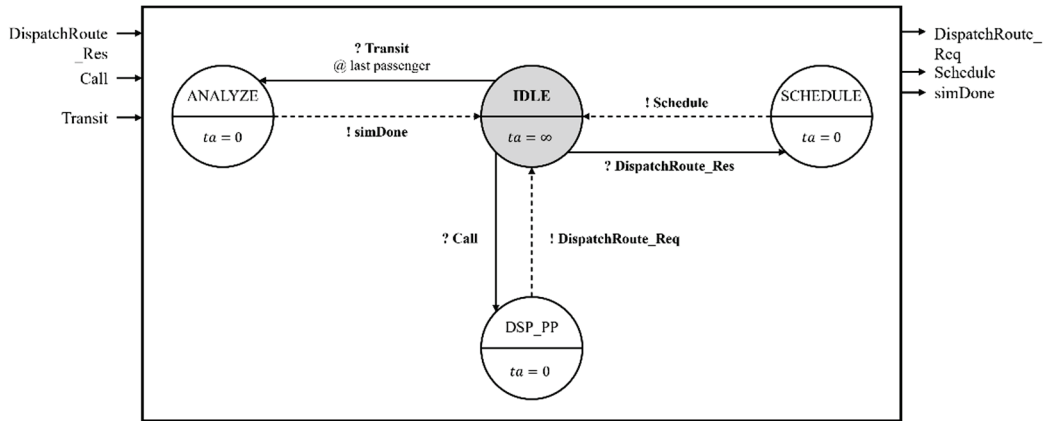


Figure 6. DEVS diagram of Schedule Manager atomic model.

- Dispatching and Routing Manager

The Routing Manager is responsible for executing the dispatching and routing algorithms requested by the scheduler. Its primary function is to dispatch the optimal DRT based on the boarding and dropping locations of passengers, and to calculate the DRT's best route, thereby generating the candidate schedule. The function of this model is encapsulated within the DEVS set using Equations (17)–(24). Upon the occurrence of an input event, as specified in Equation (18), the model undergoes an external transition to the ALGORITHM state, as shown in Equation (21). Following the time-lapse defined by Equation (24), the model exports an output event, as described in Equation (23), leading to an internal transition in accordance with Equation (22). The Routing Manager model is a crucial component of the operational strategy and is pivotal for efficient vehicle dispatching and path planning. This model offers a unique advantage: when adjustments or modifications to various operational strategies are necessary, only the atomic model must be redesigned or modified. The remaining system modules remain unchanged, thereby maintaining the integrity and continuity of the entire framework. This attribute underscores the overall structure of the DEVS model in enabling flexible and dynamic responses to changing operational requirements, while ensuring broader system stability and effectiveness.

$$Dispatching\&RoutingManager_{Atomic} = \langle X, Y, S, \delta_{ext}, \delta_{int}, \lambda, TA \rangle \quad (17)$$

$$X = \{DispatchRoute_{Req}\} \quad (18)$$

$$Y = \{DispatchRoute_{Res}\} \quad (19)$$

$$S = \{IDLE, ALGORITHM\} \quad (20)$$

$$\delta_{ext} : \{(IDLE) \times (DispatchRoute_{Req}) \rightarrow (ALGORITHM)\} \quad (21)$$

$$\delta_{int} : \{(ALGORITHM) \rightarrow (IDLE)\} \quad (22)$$

$$\lambda : \{(ALGORITHM) \rightarrow (DispatchRoute_{Res})\} \quad (23)$$

$$TA : \begin{cases} (IDLE) \rightarrow \infty \\ (Algorithm) \rightarrow 0 \end{cases} \quad (24)$$

The physical system encompasses both the DRT and passenger queue, and functions as a tangible component of the operation. It primarily interacts with the scheduler and generator, executing transportation tasks based on the received directives. In this setup, the passenger queue operates by forwarding passenger calls generated by the generator

to the Schedule Manager. Similarly, the DRT performs its tasks upon receiving a schedule from the Schedule Manager. If the model under the physical system is divided into separate entities, such as DRTs and passengers, it can be easily expanded to other strategies and transportation in future studies. This enables the system to be easily expanded or refined, particularly when incorporating simulations of additional forms of transportation such as buses and taxis. This classification not only enhances the system's adaptability to new simulation requirements, but also ensures that the framework remains scalable and versatile.

- DRT

The DRT operates based on the schedule assigned by the Schedule Manager and performs its designated task. DRT model's operational logic is defined by Equations (25)–(32) within the DEVS dataset. According to Equation (28), the DRT remains in the IDLE state upon initiating the simulation. It transitions to the MOVE state upon receiving a schedule message from the Schedule Manager, as indicated by Equations (26) and (29). In the MOVE state, the DRT navigates according to the assigned route and reports its location. When the DRT reaches a passenger's pickup or drop-off stop, it shifts into the BOARD state to manage pickup and drop-off activities over time, as specified in Equation (32). During this phase, the model exports a transit output message as outlined in Equations (27) and (31). If there are additional pick-up or drop-off tasks, the DRT reverts to the MOVE state, as described in Equation (30), to proceed to the next location. Upon completing all the tasks detailed in its Schedule, the DRT returns to the IDLE state, awaiting further instructions. Moreover, if the DRT receives an updated schedule message from the Schedule Manager in either the MOVE or BOARD state, it adjusts its current schedule accordingly. This dynamic allows the DRT to adapt to real-time changes and ensure the management of its routing and passenger service tasks.

$$Shuttle_{Atomic} = \langle X, Y, S, \delta_{ext}, \delta_{int}, \lambda, TA \rangle \quad (25)$$

$$X = \{Schedule\} \quad (26)$$

$$Y = \{Transit, CurrentNode\} \quad (27)$$

$$S = \{IDLE, MOVE, BOARD\} \quad (28)$$

$$\delta_{ext} : \begin{cases} (IDLE) \times (Schedule) \rightarrow (MOVE) \\ (MOVE) \times (Schedule) \rightarrow (MOVE) \\ (BOARD) \times (Schedule) \rightarrow (BOARD) \end{cases} \quad (29)$$

$$\delta_{int} : \begin{cases} (MOVE) \rightarrow (MOVE) \\ (MOVE) \rightarrow (BOARD) \\ (BOARD) \rightarrow (MOVE) \\ (BOARD) \rightarrow (IDLE) \end{cases} \quad (30)$$

$$\lambda : \begin{cases} (MOVE) \rightarrow (CurrentNode) \\ (BOARD) \rightarrow (Transit) \end{cases} \quad (31)$$

$$TA : \begin{cases} (IDLE) \rightarrow \infty \\ (MOVE) \rightarrow t_{move} \\ (BOARD) \rightarrow t_{BOARDING} \end{cases} \quad (32)$$

The model design was implemented using PyDEVS, an engine developed to run DEVS models crafted in Python 3.10, adhering to DEVS formalism [35]. This implementation, specifically chosen for Python's flexibility and ease of visualization, holds potential for

future integration with various AI packages, broadening the scope of its application and utility [36].

4. Experiment

The framework for the experimental scenario is outlined in Section 4.1. Section 4.2 details the pre-processor, while Section 4.3 defines the key performance indicators crucial to this study. Section 4.4 presents Experiment 1, which assesses the current scenario and compares the number of vehicles to the demand. Finally, in Section 4.5, we conducted Experiment 2 for a new operational strategy, offering a comprehensive analysis of its efficacy.

4.1. Experimental Scenario

The experiments were conducted in Dongtan 1 New Town and Dongtan 2 New Town, Hwaseong, Gyeonggi Province, Republic of Korea. Dongtan 1 New Town covers an area of 9.03 km², and Dongtan 2 New Town spans 24.01 km², as illustrated in Figure 7. These areas were selected because of the city's mobility challenges, exacerbated by the scarcity of taxis and buses, coupled with ongoing construction that limits parking availability. Given Dongtan Station's proximity and the need for DRT services to connect surrounding areas, new transportation options may be required due to the city's ongoing development. Therefore, we conducted experiments in an urban setting to evaluate various DRT operation strategies and collect data to address these transportation challenges.

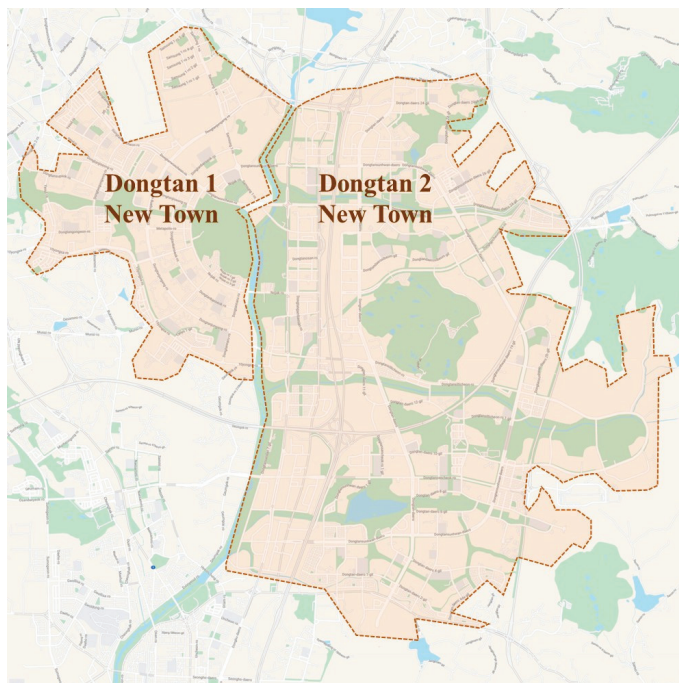


Figure 7. Target region for simulation experiments: Dongtan New Town city.

4.2. Pre-Processor

To effectively utilize the data in simulation experiments, the foremost task is to collect and pre-process the base data. This section delineates the three primary pre-processing steps. Section 4.2.1 elaborates on the networking process, which utilizes GIS map data for its foundation. Following this, Section 4.2.2 discusses the classification of DRT data, an effort grounded in bus demand data analysis. Finally, Section 4.2.3 details the process for selecting essential stops from all available options, pinpointing those crucial for conducting DRT experiments.

4.2.1. Map Networkization

Pre-processing is vital for effective utilization of GIS map data. To this end, we transformed the road and intersection data into a network format composed of nodes and links. This transformation facilitates the application of routing algorithms through the adoption of efficient data formats. Figure 8 shows the transformation, or ‘networkization’, using GIS data from Dongtan New Town. The network consisted of 1088 nodes representing intersections or destinations, connected by 3166 links. Each node includes details on adjacent nodes and links, including x- and y-coordinates. Links provide essential details, such as the link length, start and end nodes, and maximum allowable speed on that road segment, all of which are crucial for optimizing routing algorithms.

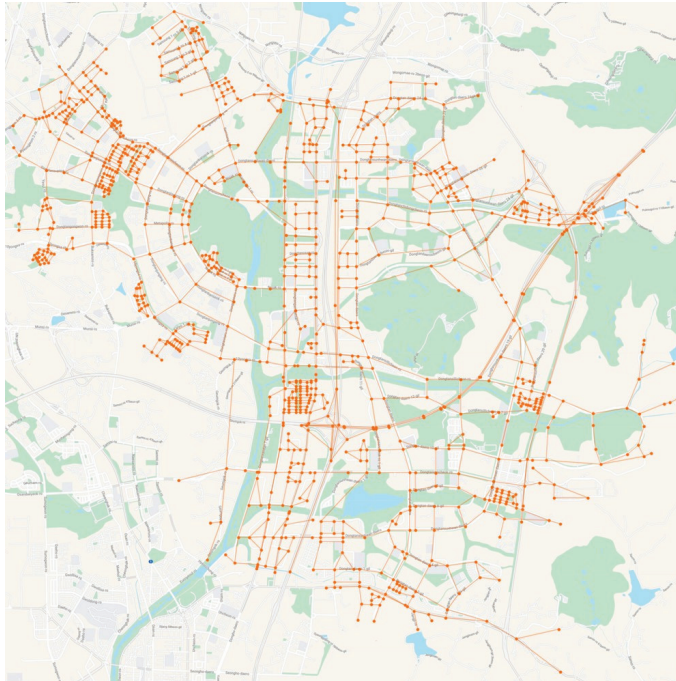


Figure 8. Network nodes and links of target city.

4.2.2. Demand Data Categorization

Given the practical challenges of gathering real DRT demand data, we chose an alternative approach. We collected bus stop data for Dongtan New Town in March 2024, using the Republic of Korea’s Transportation Card Big Data System [37]. To estimate the DRT demand from the bus data, we applied the public transportation share ratio provided by the Ministry of Land, Infrastructure, and Transport of the Republic of Korea [38]. As illustrated in Figure 9, the share of public transport varies annually, and based on this data, Table 2 summarizes national taxi usage rates ranging from 10.7% to 19.5%. Reflecting these values and following practices in early-stage DRT planning, we assumed DRT demand to be 5% (low) and 10% (moderate) of existing bus demand. Passenger requests were generated dynamically using probabilistic distributions to simulate time-varying demand, and multiple simulation runs were performed for each scenario to ensure robustness in the evaluation of key performance indicators (KPIs).

Table 2. Yearly demand rate for taxis compared to bus demand.

Taxi/Bus	2018	2019	2020	2021
Rate (%)	10.7	12.1	18	19.5

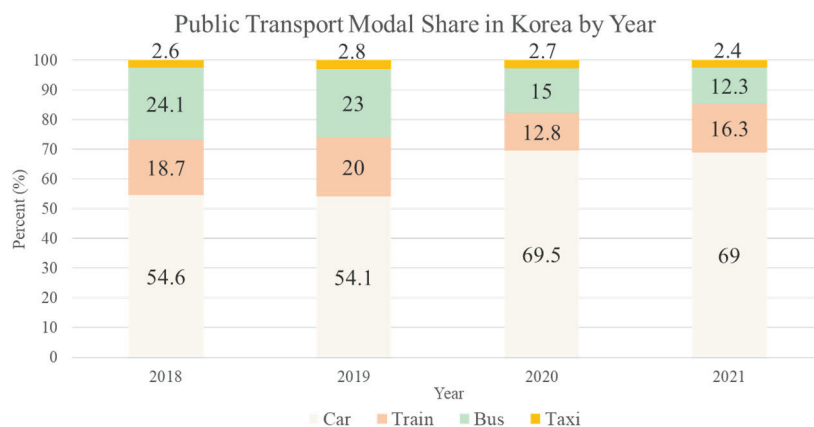


Figure 9. Public transport modal share in Korea by year.

4.2.3. Essential Stop Selection

Given that bus stops are overly abundant for DRT services, strategic selection of essential stops is necessary. By leveraging bus stop data, we applied specific criteria to identify these pivotal stops. The area of interest was divided into eighty-one zones, arranged on a 9×9 grid. The essential stops were identified within each zone based on the following five weighting criteria:

1. The presence of more than one elementary, middle, or high school within a 10-m radius;
2. The proximity to an industrial complex or distribution center within 10 m;
3. The existence of two or more apartment complexes or residential buildings within the same distance;
4. The availability of cultural and essential living facilities such as hospitals and parks within 10 m;
5. The location being within twenty meters of a busy street, restaurant street, or subway station.

Through this refinement process, the number of available stops was reduced from 282 to 188—a 33% reduction—as shown in Figure 10. The resulting list of essential stops was then utilized in the simulation experiments to establish a more focused and operationally realistic environment for evaluating the proposed DRT system. This selection not only enhances simulation efficiency, but also reflects the practical constraints of deploying DRT services in high-density urban settings.

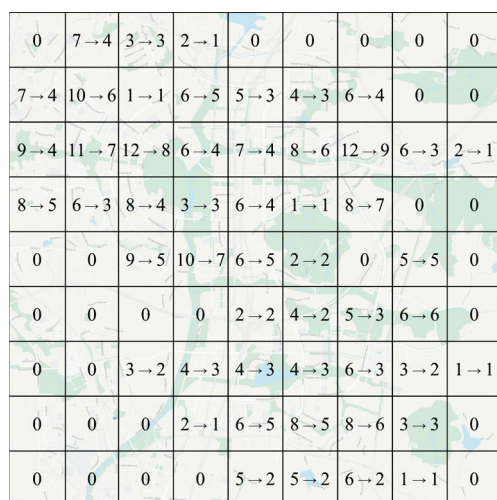


Figure 10. Refined list of essential stops.

4.3. Key Performance Indicator

Defining the KPIs is crucial for testing new operational strategies and evaluating algorithms. Table 3 lists the four essential KPIs from the passenger perspective and three key KPIs from the DRT perspective that we identified for evaluating our strategies. These KPIs are instrumental in assessing the system's performance and capturing data for subsequent analyses, ensuring a comprehensive evaluation of operational efficacy and strategic outcomes. More information to derive the KPIs is provided in Appendices A.1 and A.2.

Table 3. Summary of KPIs within an experiment.

Perspective of KPI	KPI Name	KPI Description
Passenger	AWm	Average waiting time to board a vehicle with m passengers
Passenger	ARm	Average ride time in a vehicle with m passengers
Passenger	AΔWm	Average of the difference between expected and actual waiting times for m passengers
Passenger	AΔRm	Average of the difference between expected and actual boarding time for m passengers
DRT	AURn	Average utilization rate for DRT n
DRT	APLn	Average ridership within the operating hours of DRT n
DRT	AAR	Average acceptance rate for calls from DRT n

4.4. Experiment 1

In Experiment 1, the number of vehicles varied from 1 to 10 in the current map, and the KPIs for the DRT operation were analyzed across two distinct demand stages: low and high. This strategic approach allows for an understanding of the optimal number of DRTs required prior to service deployment, thereby facilitating the acquisition of anticipatory data. Such a methodical analysis not only underscores the scalability of DRT operations, but also aids in forecasting the system's efficiency and effectiveness under varying demand conditions.

4.4.1. Experiment 1 Parameter

The setup parameters for Experiment 1 are outlined in Table 4, with Dongtan 1 and 2 New Towns as the primary areas of study, covering a total area of 33.04 km². The map incorporates 1088 nodes representing intersections or stops connected by 3136 links that form a road network. In this experimental framework, the number of DRTs varied from one to ten, with 188 designated stops for boarding and alighting. Passenger demand was segmented into two categories, low and high, corresponding to 5% and 10% of the bus demand in Dongtan, respectively.

The key operational parameters include a boarding time of 5 s per group of passengers and a maximum waiting time threshold of 10 min, beyond which passengers may decline calls. The DRT speed was set to 5 km/h below the maximum speed limit of each road; for instance, on a road with a speed limit of 60 km/h, the DRT operated at 55 km/h. DRT can accommodate up to nine passengers simultaneously. To ensure the reliability of the results, the Monte Carlo method was employed, averaging the outcomes over 30 trials.

Table 4. Input parameters for Experiment 1.

Parameter Name	Parameter Level	Parameter Description
City	Dongtan 1 and 2 New Town	The city that is the subject of the simulation
City area	33.04 km ² for combined area	The area of the city that is the subject of the simulation
Map node	1088 nodes	The number of nodes in the simulation map
Map link	3136 links	The number of links in the simulation map
Number of DRTs	1~10 (incrementing by 1)	The number of operational DRTs
Number of bus stops	188 stops	The number of bus stops on the map where boarding is possible
Passenger demand	5%, 10%	Passenger demand compared to bus demand
Passenger boarding time	5 s	The time it takes for passengers to board and alight
Passenger call Acceptance time	10 min	The maximum waiting time after calling for a DRT is set at a threshold beyond which passengers may decline the call
DRT speed	Each maximum road speed—5 km/h	The maximum speed that the DRT can travel on each road
DRT capacity	Nine passengers	The maximum number of passengers that can simultaneously be aboard the DRT
Monte Carlo simulations	Thirty simulations	The number of Monte Carlo simulations conducted

4.4.2. Experiment 1 Results

The results of the experiment were categorized based on the DRT demand levels of 5% (low) and 10% (high), with subsequent analysis of these outcomes. As discussed in Section 4.3, the analysis of the KPIs for passengers includes the average waiting time, average ride time, deviation in waiting time, and deviation in ride time. Meanwhile, the DRT KPIs focus on the average call acceptance rate, utilization rate, and passenger load per DRT.

In scenarios of low demand, the variations in the average waiting time for DRTs and passenger ride time in relation to the change in the number of vehicles are depicted in Figure 11. For instance, with the operation of six DRTs, passengers experienced an average waiting time of 4.48 min (268.9 s) from the moment the DRT was called for boarding. Furthermore, the journey to their destination post-boarding took an average of 7.76 min (465.8 s). This indicates a clear trend in which an increase in the number of DRTs leads to a reduction in both waiting and ride times for passengers.

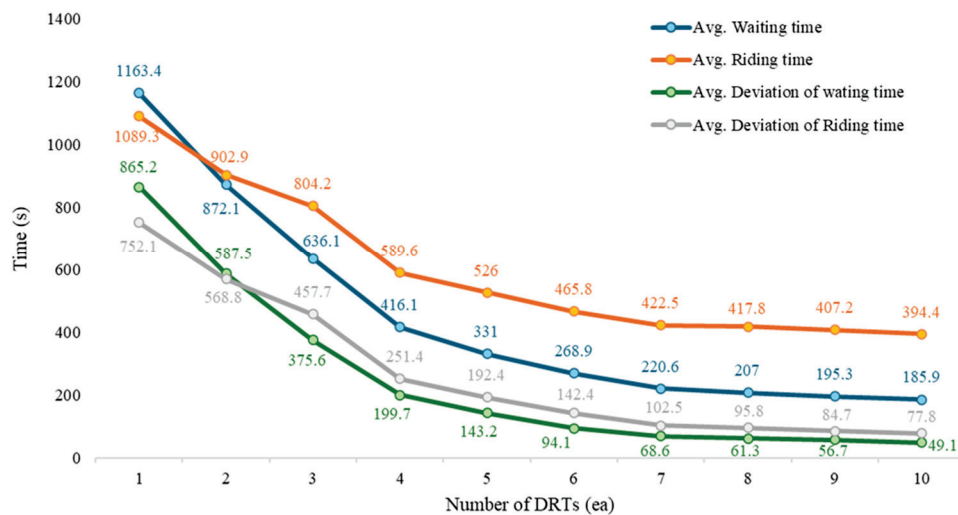


Figure 11. Results from Experiment 1 regarding passengers based on shuttle count: demand 5%.

The proposed DRT dynamic routing, which allows ride-sharing, leads to potential variations in both waiting and ride times beyond the initial estimates. These deviations are depicted in Figure 11 for both waiting time and ride time. For instance, with six DRTs in operation, the waiting time deviation was observed to be 1.56 min (94.1 s). Comparatively, based on the waiting time results, if a DRT had been able to proceed directly to a called passenger, the waiting could have been 2.92 min shorter (4.48 min–1.56 min deviation), indicating that the actual waiting time extended to 4.48 min due to mid route alterations. Similarly, the ride time deviation was recorded at 2.37 min (142.4 s), suggesting that a direct route to the destination post-boarding might have taken 5.39 min (7.76 min–2.37 min deviation). However, the necessity of adjusting dynamic routes for picking up or dropping off other passengers has led to increased ride times.

When a passenger call predicts a basic waiting time of approximately 10 min, such calls are automatically set for rejection. Figure 12 illustrates the effects on passenger request acceptance rate, DRT utilization rate, and DRT passenger load when the number of available DRTs is changed. Acceptance rates started at a significantly low level, with only one or two available DRTs. Nonetheless, this rate improved noticeably once the availability increased to four DRTs, with almost every call being accepted. When the availability reached six or more DRTs, the average acceptance rate was 100%. Regarding DRT utilization, when only one DRT was in operation, the average utilization rate was 85%; however, this rate decreased to 56% with four available DRTs. This pattern indicates that while more available DRTs lead to higher call acceptance rates, they also result in an increase in idle DRTs. Additionally, an analysis of DRT passenger load revealed that when only one to two DRTs were available, each DRT carried an average of five passengers. However, when the number of available DRTs increased from four to nine, the average number of passengers per DRT decreased to three.

As depicted in Figure 13, the high-demand scenario exhibited a trend similar to that of the low-demand scenario. However, the waiting and ride times observed at the sixth DRT iteration are notably longer, clocking in at 16.7 min (1001.1 s) and 16.5 min (995.2 s), respectively. This marks a significant increase from the low-demand scenario's times of 4.48 min for waiting and 7.76 min for riding. Moreover, the deviations in waiting and ride times are 13.1 min (786.7 s) and 10.8 min (651.4 s), respectively. These deviations are higher than those recorded in the low demand scenario, which were 1.56 min for waiting and 2.37 min for riding, indicating a pronounced impact of demand levels on DRT service efficiency.

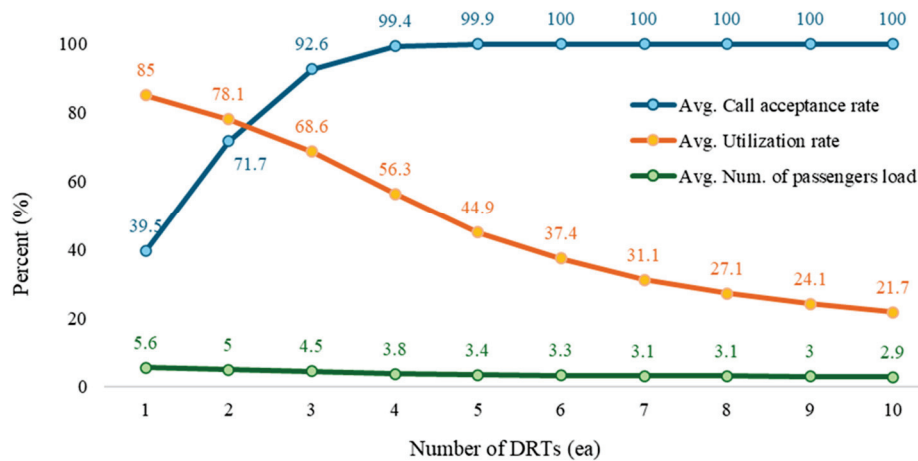


Figure 12. Results from Experiment 1 regarding shuttles based on shuttle count: demand 5%.

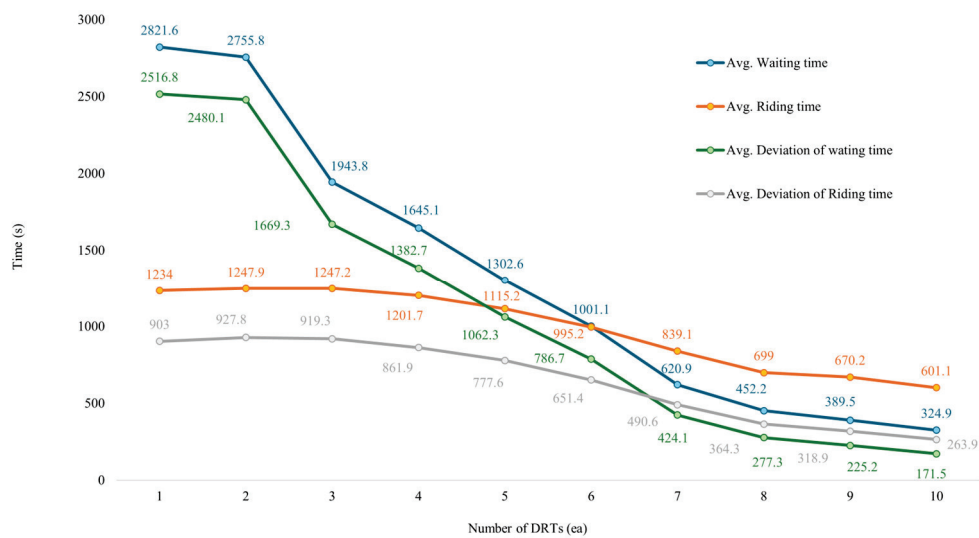


Figure 13. Results from Experiment 1 regarding passengers based on shuttle count: demand 10%.

The DRT KPIs demonstrate a consistent trend across different demand scenarios. However, in the low-demand scenario, as shown in Figure 12, the call acceptance rate was 100% when the number of available DRTs reached four. Conversely, in a high-demand scenario, achieving a similar call acceptance rate requires more than six available DRTs, as shown in Figure 14. This finding contrasts with the low-demand condition in which a 56% DRT utilization rate was observed with four available DRTs. For the high-demand scenario, the utilization rate remains significantly higher at 76.2% with the same number of available DRTs. In addition, the analysis of the DRT passenger load shows its variability with the number of available DRTs. With one to three DRTs available, an average of six passengers was observed. This average decreases to five passengers with four to six available DRTs and further decreases to four passengers when the availability exceeds seven DRTs.

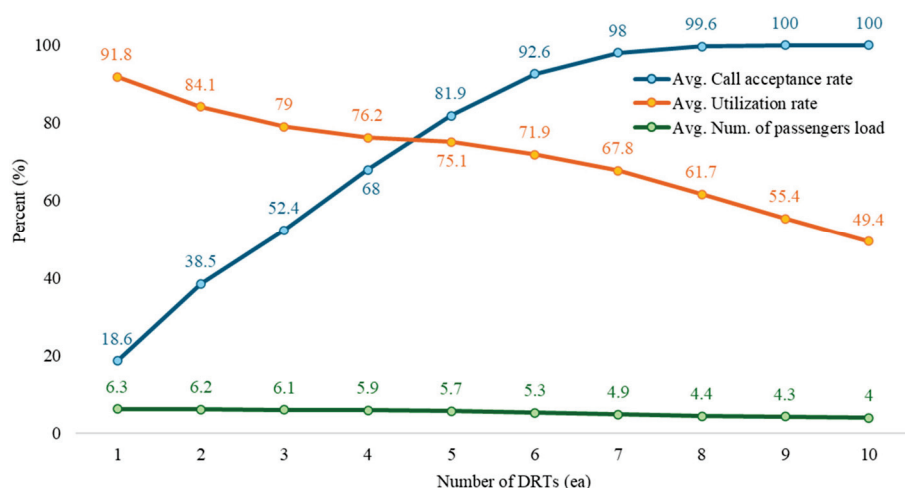


Figure 14. Results from Experiment 1 regarding Shuttles based on shuttle count: demand 10%.

4.4.3. Experiment 1 Discussion

Across both low- and high-demand scenarios, the deviation in waiting time and ride time was observed to decrease as the number of vehicles increased, indicating greater service stability under larger fleets. However, at fixed fleet sizes, demand levels had a significant impact. For example, when demand increased from 10% to 20%, waiting times surged by 3.7 times, and ride times increased by 2.12 times, compared to low-demand conditions. This difference highlights the added strain on routing and scheduling in high-demand environments.

To isolate the core service delays, we subtracted the deviation factors and found that the base waiting time under low demand was 2.92 min (from 4.48 min), while under high demand it was 3.6 min (from 16.7 min). Similarly, the base ride time was 5.39 min under low demand and 5.7 min under high demand. These results suggest that the increased total times in high-demand scenarios primarily stem from additional intermediate stops and route adjustments, rather than changes in trip distance.

In terms of call acceptance rate, both demand scenarios showed improvement as vehicle numbers increased. However, when the fleet size exceeded four vehicles under low demand and eight under high demand, the utilization rate dropped significantly, indicating overprovision. Notably, for a fixed number of vehicles, high-demand conditions naturally resulted in a higher average passenger load per DRT, which improves utilization, but may reduce passenger satisfaction due to potential overcrowding.

This comparison underscores the inherent trade-offs between service accessibility (e.g., higher acceptance rates) and operational efficiency (e.g., vehicle utilization). It also emphasizes the importance of demand-aware planning to prevent vehicle underuse while maintaining acceptable service quality. These insights directly address the need for clearer interpretation of KPI dynamics, as emphasized in prior literature and noted in reviewer feedback.

4.5. Experiment 2

In this section, we explore and analyze a novel operation strategy known as the express dispatch service. As previously described, Express Dispatching is a service designed to transport express passengers directly from their pickup location to their destination without any intermediate stops, especially during periods when the number of vehicles exceeds the demand or when additional vehicles are available. This service offers express passenger convenience similar to that of a taxi, providing a more appealing option for those who may otherwise avoid shared rides owing to concerns over extended waiting or ride times.

4.5.1. Experiment 2 Parameter

Experiment 2 followed a structure similar to Experiment 1, as detailed in Section 4.4, but was conducted only for the high DRT demand category. The specific parameters for this experiment are listed in Table 4 and presented in Section 4.4.1. Unlike Experiment 1, this experiment introduced a new variable by varying the proportion of express passengers within the overall passenger pool. Three distinct scenarios were considered: (1) a baseline with no express passengers, identical to the setup in Experiment 1; (2) a case where express passengers accounted for 10% of all passengers; and (3) a case with a 20% express passenger share.

Accordingly, analysis and experimentation were conducted across three scenarios for high-demand situations: no express passengers, 10% express passengers, and 20% express passengers. When passengers are generated, express passengers appear with a random probability of 10% or 20% throughout the entire simulation period. The remaining 90% of passengers are standard passengers, following the same conditions as in Experiment 1. This approach allows a comprehensive examination of how varying levels of express passengers affect the dynamic operation and effectiveness of DRT services under different demand conditions.

4.5.2. Experiment 2 Result

The analysis initially focused on the relationship between the number of available DRTs and the ratio of express passengers under high demand conditions. In scenarios without express passengers, the average number of passengers per DRT is consistently the highest. This occurs because express passengers follow a boarding pattern where the DRT bypasses intermediate stops until reaching their destination, leading to an increased overall passenger count. Furthermore, similar to the observations in Experiment 1, the average passenger load per DRT decreased as more DRTs became available, demonstrating the trade-off between service efficiency and resource utilization.

Figures 15 and 16 illustrate the effects of increasing the number of DRTs from five to ten on waiting and ride times. In these figures, the labels “Standard (10%)” and “Express (10%)” refer to standard and express passengers, respectively, when 10% of all DRT users are assigned to the express service. Similarly, “Standard (20%)” and “Express (20%)” represent the corresponding passenger groups when the express service adoption rate is 20%. Initially, with five DRTs available, the average waiting time stands at 21.71 min (1302.6 s). In a scenario with 10% express passengers, standard passengers have a waiting time of 19.76 min (1186 s), whereas express passengers have a waiting time of 11.85 min (711.1 s). In a 20% express scenario, standard passengers face a waiting time of 22.72 min (1363.4 s), whereas express passengers wait 12.04 min (722.6 s).

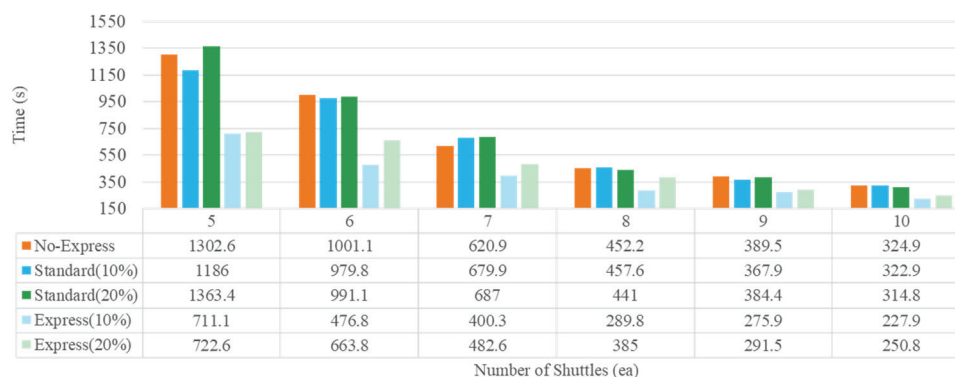


Figure 15. Express passenger shuttle waiting time based on shuttle count: demand 10%.

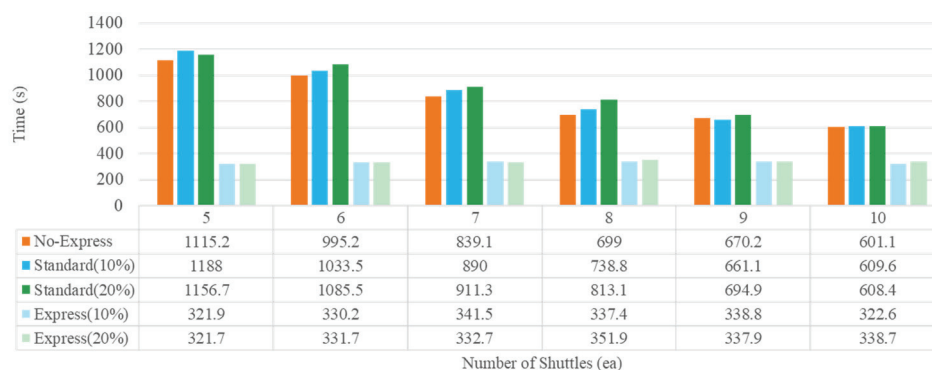


Figure 16. Express passenger shuttle ride time based on shuttle count: demand 10%.

As the DRT count reaches nine, the waiting time for standard passengers drops to 6.49 min (389.5 s), 6.13 min (367.9 s), and 6.40 min (384.4 s), with express passengers experiencing times of 4.59 min (275.9 s) and 4.85 min (291.5 s). This trend demonstrates that as more DRTs become available, the gap in waiting times between standard and express passengers narrows, regardless of the express passenger ratio.

Ride times similarly exhibit disparities: 18.58 min (1115.2 s) without express service, 19.8 min (1188 s) in a 10% express scenario, and 19.27 min (1156.7 s) in a 20% express scenario with five DRTs available. However, these differences shrink to 11.17 min (670.2 s), 11.01 min (661.1 s), and 11.58 min (694.9 s) when the number of available DRTs is nine, indicating a convergence in travel durations as DRT availability increases. When there is a large number of available vehicles, accommodating the convenience of express passengers has minimal impact on the waiting and ride times of standard passengers.

Specifically, with five DRTs available and express passengers making up 10% of the total, it was observed that express passengers enjoy a waiting time that is 40% shorter and a ride time that is 72% shorter than standard passengers. Moreover, when the proportion of express passengers increases to 20%, express passengers experience a 47% reduction in waiting time and maintain a 72% shorter ride time relative to standard passengers, demonstrating the effectiveness of express services in optimizing passenger transit times without significantly affecting standard passenger experience.

The deviations in the waiting and ride times followed a similar trend, as shown in the data depicted in Figure 17 for the waiting time and Figure 18 for the ride time deviation. With five vehicles available, the deviation in standard passenger waiting times registers at 17.70 min (1062.3 s), 15.61 min (937.1 s), and 18.43 min (1105.9 s), respectively. For express passengers, the waiting time deviation is significantly lower, at 2.21 min (132.8 s) and 7.40 min (444.4 s). Moreover, as the availability of DRTs increases to nine, the deviation decreases substantially for both standard and express passengers, falling to 3.75 min (225.2 s), 3.37 min (202.4 s), and 3.53 min (212 s) for standard passengers and to 1.53 min (92 s) and 1.8 min (108 s), respectively.

In the high-demand scenario concerning the DRT acceptance rate, as illustrated in Figure 19, the no-express scenario had an acceptance rate of 81.9%, whereas standard passengers in the 10% express scenario had an 80% acceptance rate, and those in the 20% express scenario had an 81.4% rate. However, the acceptance rates for express passengers were lower, at 75.5% and 77.6% in the 10% and 20% express scenarios, respectively. The primary reason for this difference is the challenge of consecutively integrating express passenger departures and destination points into the existing DRT route. This constraint often leads to call rejections for express passengers because of the difficulty in finding a suitable insertion point in the DRT journey. Furthermore, when the number of available

vehicles increased to eight, the acceptance rates for all scenarios approached an average of approximately 100%.

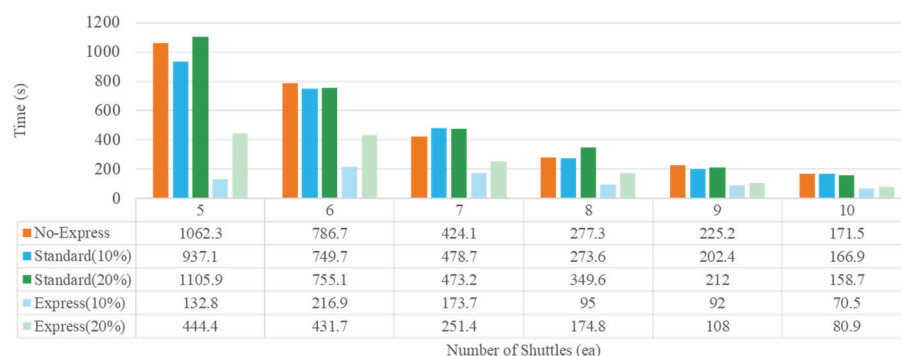


Figure 17. Express passenger shuttle waiting time deviation based on shuttle count: demand 10%.

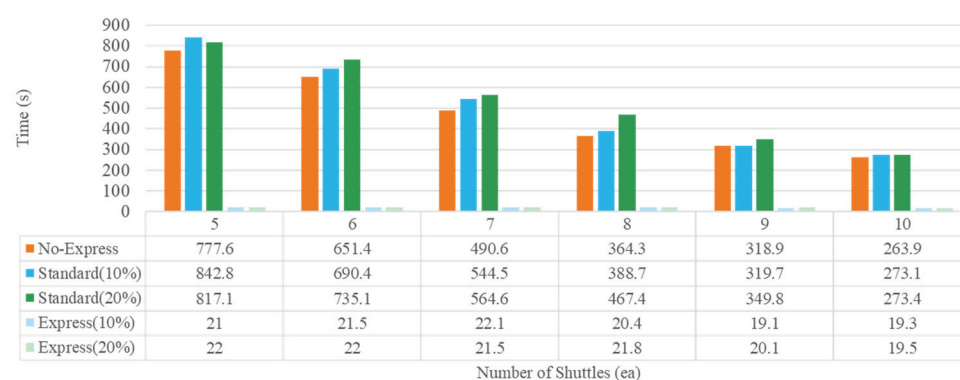


Figure 18. Express passenger shuttle ride time deviation based on shuttle count: demand 10%.

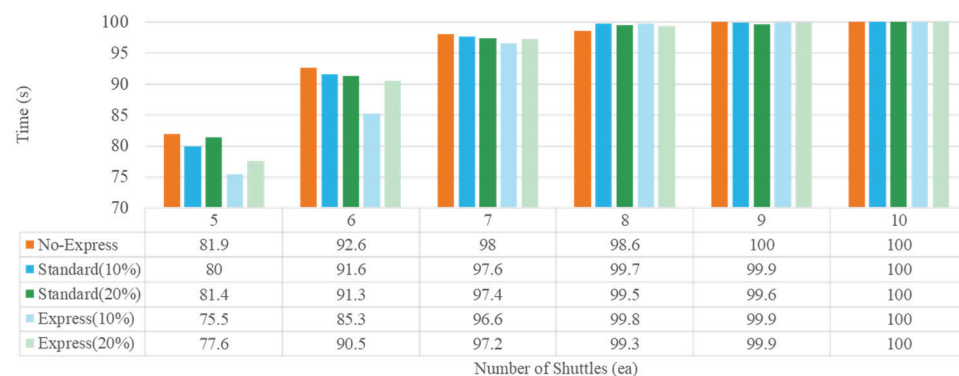


Figure 19. Average call acceptance rate of shuttles based on shuttle count and express rate: demand 10%.

The DRT utilization rate, as depicted in Figure 20, reveals that scenarios involving express passengers tend to exhibit slightly higher utilization rates than the no-express scenario, regardless of the number of vehicles. This increase is attributed to the incorporation of express passengers, which may have led to the creation of bypass routes. However, this trend shows that with five available vehicles, the utilization rates were 75.1%, 78.3%, and 75.3%; however, as the number of available vehicles increased to nine, the differences narrowed to 55.4%, 56.1%, and 56.3%, respectively.

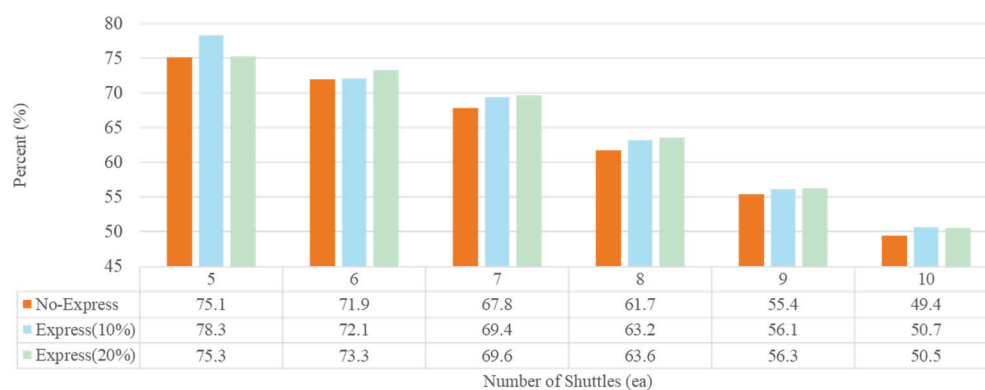


Figure 20. Average utilization rate of shuttles based on shuttle count and express rate: demand 10%.

Moreover, Figure 21 illustrates that the highest average number of passengers boarding the DRT occurred in the no-express scenario across all vehicle availabilities, and the lowest concurrent passenger numbers were observed in scenarios with 20% express. This is because passengers travel directly from their pickup locations to their destinations, bypassing other stops. In addition, the data indicate that as the number of available DRTs increases, passengers spread more across the fleet, leading to a decrease in the average passenger load per DRT. This dispersion effect underscores the impact of increased vehicle availability in reducing DRT crowding and optimizing passenger distribution.

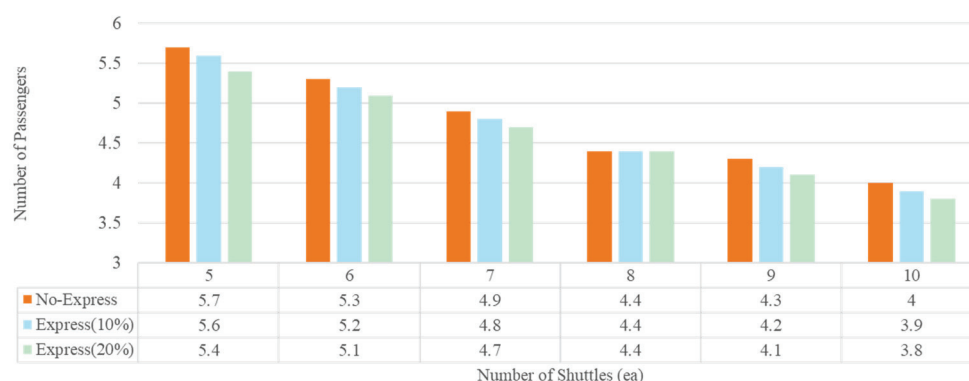


Figure 21. Average shuttle passengers load based on shuttle count and express rate: demand 10%.

4.5.3. Experiment 2 Discussion

In Experiment 2, we simulated a high-demand scenario and examined three DRT operation strategies: no express service, 10% express passenger ratio, and 20% express passenger ratio. When 20% of passengers used the express service, waiting times were reduced by 40%, and ride times decreased by 72%. These results indicate that express passengers consistently experienced improved KPIs compared to standard passengers, regardless of demand level, express ratio, or vehicle availability, demonstrating the efficiency of integrating express services into DRT operations.

Additionally, we analyzed the effect of express services on standard passengers' KPIs. When the number of vehicles was limited, the gap between the KPIs of express and non-express operations was more pronounced. For example, in a high-demand setting with five vehicles, the average waiting times for standard passengers under each scenario were 21.71, 19.76, and 22.72 min, respectively. However, with nine vehicles, these values converged to 6.49, 6.13, and 6.40 min, showing that increasing vehicle availability reduces the disparity. These findings suggest that an optimal fleet size, combined with a balanced ratio of express passengers, can improve service quality for express users without adversely affecting standard passengers.

From an operational standpoint, adding express passengers tends to lower the call acceptance rate or increase the utilization rate compared with scenarios without express passengers. This suggests that more resources are allocated to efficiently transport express passengers. This pattern became more pronounced as the proportion of express passengers increased, regardless of the overall demand and number of vehicles available. If the proportion of express passengers becomes too high, it could negatively impact the KPIs for both standard passengers and overall vehicle operation. Therefore, when considering an increase in the ratio of express passengers, it is necessary to find a suitable balance to effectively manage the consumption of vehicle resources.

These results suggest that the express service is not merely a reclassification of passengers, but rather a strategic operational mechanism that enables differentiated service delivery within DRT systems. By prioritizing a subset of passengers for direct, uninterrupted service, this strategy supports flexible resource allocation aligned with service level expectations. While moderate adoption of express services can enhance performance for targeted users without significantly compromising standard service quality, excessive allocation may lead to system imbalance. Therefore, the express service should be regarded as a configurable operational tool that requires careful calibration. Its effectiveness lies not only in numerical improvements, but also in its potential to support tiered service models, optimize resource distribution, and enhance user satisfaction in capacity-constrained environments.

Furthermore, although express service inevitably introduces a degree of resource inclination by prioritizing specific passengers, our analysis shows that this effect can be mitigated through adequate fleet sizing and demand balancing. When appropriately configured, the express strategy improves overall system efficiency without significantly degrading the service experience for standard passengers. This highlights its potential as a robust and adaptable solution for tiered mobility services in real-world urban operations.

5. Conclusions

This study introduced a DEVS-based simulation framework to evaluate a novel operational strategy for demand-responsive transport (DRT), focusing on an express service model. The modular framework consists of four components—pre-processor, simulation model, operational strategy module, and result analysis—and was used to assess DRT performance across varying vehicle numbers and demand levels. Key performance indicators (KPIs) were analyzed from both the passenger and system perspectives, including wait time, ride time, call acceptance rate, and vehicle utilization.

Simulation experiments demonstrated the feasibility and effectiveness of the proposed strategy. Under high-demand conditions, limited fleet size increased waiting and ride times, whereas expanding the fleet improved call acceptance at the cost of reduced vehicle utilization. The express service strategy, in particular, yielded notable benefits: when 10–20% of passengers used the express option, their waiting and ride times decreased by up to 40% and 72%, respectively, without significantly degrading the experience of standard passengers. These findings indicate that selective integration of express services can improve both operational efficiency and user satisfaction, provided that vehicle allocation is appropriately managed.

This study primarily focused on operational-level modeling and simulation-based evaluation. While KPIs were designed to reflect system performance and user experience, cost-related indicators—such as total operating expenditure—were not explicitly modeled. We recognize that such metrics are essential for balancing service quality and financial sustainability, and we plan to integrate cost analysis modules in future extensions of the framework.

Furthermore, because real-world DRT data were unavailable for the study area, demand was estimated based on existing bus usage data, and model validation was conducted using simplified scenarios. Nevertheless, the proposed DEVS-based framework remains highly modular and adaptable, enabling integration of diverse services, dispatching logic, and regional contexts. Future research will focus on incorporating dynamic demand elements, advanced scheduling algorithms, and economic feasibility analyses, with the goal of supporting DRT deployment in rural and mixed-modal environments. This research offers a flexible and extensible simulation platform for evaluating innovative DRT strategies in evolving mobility systems

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

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article.

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

Appendix A

Appendix A.1. Passenger KPI

Let $\mathbf{A}$ denote the set of all passengers (Equation (A1)). For each passenger m , the call time (C_m), pickup time (P_m), and drop-off time (D_m) are defined in Equations (A2)–(A4). These timestamps are used to compute the waiting time (W_m , Equation (A9)) and ride time (R_m , Equation (A10)). The minimum ride time between two nodes x and y is denoted as $T_{x,y}$ (Equation (A5)), while node i , node j , and node k represent the DRT's current location, the pickup node, and the drop-off node for passenger m , respectively (Equations (A6)–(A8)).

To measure deviation from ideal service, we define the waiting time deviation (ΔW_m) and ride time deviation (ΔR_m) in Equations (A11) and (A12), which quantify detours caused by shared rides. The average values of these indicators across all passengers are given by Equations (A13)–(A16) and used to evaluate overall passenger-level service performance.

$$\mathbf{A}: \text{Set of all passengers} \quad (\text{A1})$$

$$P_m: \text{Pickup time for passengers } m \quad (\text{A2})$$

$$C_m: \text{Call time for passengers } m \quad (\text{A3})$$

$$D_m: \text{Drop-off time for passengers } m \quad (\text{A4})$$

$$T_{x,y}: \text{Minimum riding time between nodes } x \text{ and } y \quad (\text{A5})$$

$$\text{node } i: \text{Current node of DRT} \quad (\text{A6})$$

$$\text{node } j: \text{Pickup node of passengers } m \quad (\text{A7})$$

$$\text{node } k: \text{Drop-off node of passengers } m \quad (\text{A8})$$

$$W_m: \text{Waiting time for passengers } m \quad W_m = P_m - C_m \quad (\text{A9})$$

$$R_m: \text{Riding time for passengers } m \quad R_m = D_m - B_m \quad (\text{A10})$$

$$\Delta W_m: \text{Deviation in waiting time for passengers } m \quad \Delta W_m = W_m - T_{i,j} \quad (A11)$$

$$\Delta R_m: \text{Deviation in riding time for passengers } m \quad \Delta R_m = R_m - T_{j,k} \quad (A12)$$

$$AW_m: \text{Average waiting time for passengers } m \quad \frac{\sum_{m \in A} W_m}{|A|} \quad (A13)$$

$$AR_m: \text{Average riding time for passengers } m \quad \frac{\sum_{m \in A} R_m}{|A|} \quad (A14)$$

$$A\Delta W_m: \text{Average deviation in waiting time for passengers } m \quad \frac{\sum_{m \in A} \Delta W_m}{|A|} \quad (A15)$$

$$A\Delta R_m: \text{Average deviation in riding time for passengers } m \quad \frac{\sum_{m \in A} \Delta R_m}{|A|} \quad (A16)$$

Appendix A.2. DRT KPI

Let B denote the set of all DRT vehicles (Equation (A17)). The number of accepted and rejected ride requests is denoted by N_{ac} and N_{rc} (Equations (A18) and (A19)), forming the basis of system responsiveness. The idle time ($S_{it,n}$) and operating time ($S_{ot,n}$) of each DRT are tracked (Equations (A20) and (A21)), along with the number of passengers per request (P_m , Equation (A22)).

Using these values, the utilization rate (UR_n) and passenger load (PL_n) of each DRT are calculated (Equations (A23) and (A24)). Finally, the average values across all DRTs provide key system-level KPIs: average utilization (AUR_n), average load (APL_n), and acceptance rate (AAR , Equations (A25)–(A27)).

$$B: \text{Set of all DRTs} \quad (A17)$$

$$N_{ac}: \text{Number of accepted calls} \quad (A18)$$

$$N_{rc}: \text{Number of rejected calls} \quad (A19)$$

$$S_{it,n}: \text{DRT idle time} \quad (A20)$$

$$S_{ot,n}: \text{DRT operating time} \quad (A21)$$

$$P_m: \text{Passenger count for } m \quad (A22)$$

$$UR_n: \text{Utilization rate for DRT } n \quad UR_n = \frac{S_{ot,n}}{S_{ot,n} + S_{it,n}} \times 100 \quad (A23)$$

$$PL_n: \text{Passenger load for DRT } n \quad PL_n = \frac{\sum_{m \in A} R_m P_m}{S_{it,n}} \quad (A24)$$

$$AUR_n: \text{Average utilization rate for DRT } n \quad \frac{\sum_{n \in B} UR_n}{|B|} \quad (A25)$$

$$APL_n: \text{Average passenger load for DRT } n \quad \frac{\sum_{n \in B} PL_n}{|B|} \quad (A26)$$

$$AAR: \text{Average DRT acceptance rate} \quad AAR = \frac{N_{ac}}{N_{ac} + N_{rc}} \times 100 \quad (A27)$$

References

1. Fatnassi, E.; Chaouachi, J.; Klibi, W. Planning and operating a shared goods and passengers on-demand rapid transit system for sustainable city-logistics. *Transp. Res. B Methodol.* **2015**, *81*, 440–460. [CrossRef]
2. Quadrifoglio, L.; Hall, R.W.; Dessouky, M.M. Performance and design of mobility allowance shuttle transit services: Bounds on the maximum longitudinal velocity. *Transp. Sci.* **2006**, *40*, 351–363. [CrossRef]
3. Atasoy, B.; Ikeda, T.; Ben-Akiva, M.E. Optimizing a flexible mobility on demand system. *Transp. Res. Rec.* **2016**, *2563*, 76–85. [CrossRef]

4. Vansteenwegen, P.; Melis, L.; Aktaş, D.; Montenegro, B.D.G.; Sartori Vieira, F.; Sörensen, K. A survey on demand-responsive public bus systems. *Transp. Res. C* **2022**, *137*, 103573. [CrossRef]
5. Burghout, W.; Rigole, P.J.; Andreasson, I. Impacts of shared autonomous taxis in a metropolitan area. In Proceedings of the 94th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 11–15 January 2015.
6. Goszczyński, J.A.; Blajer, W.; Krawczyk, M. The inverse simulation study of aircraft flight path reconstruction. *Transport* **2002**, *17*, 103–107.
7. Lu, C.; Maciejewski, M.; Wu, H.; Nagel, K. Demand-responsive transport for students in rural areas: A case study in Vulkaneifel, Germany. *Transp. Res. Part A Policy Pract.* **2023**, *178*, 103837. [CrossRef]
8. Linares, M.P.; Barceló, J.; Carmona, C.; Montero, L. Analysis and operational challenges of dynamic ride sharing demand responsive transportation models. *Transp. Res. Procedia* **2017**, *21*, 110–129. [CrossRef]
9. Sun, B.; Wei, M.; Zhu, S. Optimal design of demand-responsive feeder transit services with passengers' multiple time windows and satisfaction. *Future Internet* **2018**, *10*, 30. [CrossRef]
10. Nyga, A.; Minnich, A.; Schlüter, J. The effects of susceptibility, eco-friendliness and dependence on the Consumers' Willingness to Pay for a door-to-door DRT system. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 540–558. [CrossRef]
11. Li, X.; Wang, T.; Xu, W.; Li, H.; Yuan, Y. A novel model and algorithm for designing an eco-oriented demand responsive transit (DRT) system. *Transp. Res. E* **2022**, *157*, 102556. [CrossRef]
12. Galarza Montenegro, B.D.; Sörensen, K.; Vansteenwegen, P. A large neighborhood search algorithm to optimize a demand-responsive feeder service. *Transp. Res. C* **2021**, *127*, 103102. [CrossRef]
13. Horn, M.E.T. Procedures for planning multi-leg journeys with fixed-route and demand-responsive passenger transport services. *Transp. Res. C* **2004**, *12*, 33–55. [CrossRef]
14. Wang, L.; Zeng, L.; Ma, W.; Guo, Y. Integrating passenger incentives to optimize routing for demand-responsive customized bus systems. *IEEE Access* **2021**, *9*, 21507–21521. [CrossRef]
15. Costa, P.C.; Cunha, C.B.; Arbex, R.O. A simulation-optimization model for analyzing a demand responsive transit system for last-mile transportation: A case study in São Paulo, Brazil. *Case Stud. Transp. Policy* **2021**, *9*, 1707–1714. [CrossRef]
16. Agatz, N.; Erera, A.L.; Savelsbergh, M.W.P.; Wang, X. Dynamic ride-sharing: A simulation study in metro Atlanta. *Procedia Soc. Behav. Sci.* **2011**, *17*, 532–550. [CrossRef]
17. Dandl, F.; Bogenberger, K.; Mahmassani, H.S. Autonomous mobility-on-demand real-time gaming framework. In Proceedings of the 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Cracow, Poland, 5–7 June 2019; pp. 1–10. [CrossRef]
18. Djavadian, S.; Chow, J.Y.J. An agent-based day-to-day adjustment process for modeling 'Mobility as a Service' with a two-sided flexible transport market. *Transp. Res. B Methodol.* **2017**, *104*, 36–57. [CrossRef]
19. Cho, S.-W.; Seo, K.-M.; Yun, J.-M.; Kang, B.-G. Simulation-based analysis for verifying new certification standards of smart LED streetlight systems. *Mathematics* **2024**, *12*, 657. [CrossRef]
20. Zeigler, B.P.; Praehofer, H.; Kim, T.G. *Theory of Modeling and Simulation*; Academic Press: Cambridge, MA, USA, 2000.
21. Kang, B.G.; Park, H.-M.; Jang, M.; Seo, K.-M. Hybrid model-based simulation analysis on the effects of social distancing policy of the COVID-19 epidemic. *Int. J. Environ. Res. Public Health* **2021**, *18*, 11264. [CrossRef]
22. Sung, J.-H.; Ju, S.-H.; Cho, S.-W.; Joo, H.-J.; Seo, K.-M.; Kang, B.-G. Modeling and Exploratory Analysis of Discrete Event Simulations for Optimizing Overhead Hoist Transport Systems and Logistics in Semiconductor Manufacturing. *Mathematics* **2025**, *13*, 1167. [CrossRef]
23. Lu, C.; Maciejewski, M.; Nagel, K. Effective Operation of Demand-Responsive Transport (DRT): Implementation and Evaluation of Various Rebalancing Strategies. In Proceedings of the 27th ITS World Congress, Hamburg, Germany, 11–15 October 2021; Technische Universität Berlin: Berlin, Germany, 2023.
24. David, E.; Kurt, F.; Scheier, B.; Mittag, S. Optimization of DRT Concepts Using the SUMO Simulation Software: How Can the Service Quality Be Improved? *Transp. Res. Procedia* **2023**, *69*, 481–488. [CrossRef]
25. Ronald, N.; Thompson, R.G.; Winter, S. Simulating ad-hoc demand-responsive transportation: A comparison of three approaches. *Transp. Plan. Technol.* **2017**, *40*, 340–358. [CrossRef]
26. Kagho, G.O.; Azevedo, C.L.; Ben-Akiva, M. Demand Responsive Transit Simulation of Wayne County, Michigan. *Transp. Res. Rec.* **2021**, *2675*, 125–134. [CrossRef]
27. Fielbaum, A.; Jara-Díaz, S.; Alonso-Mora, J. Beyond the Last Mile: Different Spatial Strategies to Integrate On-Demand Services into Public Transport in a Simplified City. *Public Transp.* **2024**, *16*, 73–93. [CrossRef]
28. Kim, H.; Lee, J.; Choi, J. Identifying Different Sources of the Benefit: Simulation of DRT Operation in the Heartland and Hinterland Regions. *Sustainability* **2022**, *14*, 16519. [CrossRef]
29. Alomrani, M.A.; Moravej, R.; Khalil, E.B. Deep policies for online bipartite matching: A reinforcement learning approach. *arXiv* **2021**, arXiv:2109.10380. [CrossRef]

30. Wang, H.; Guan, H.; Qin, H.; Li, W.; Zhu, J. A slack departure strategy for demand responsive transit based on bounded rationality. *J. Adv. Transp.* **2022**, 9693949. [CrossRef]
31. Zhen, L.; Fan, T.; Li, H.; Wang, S.; Tan, Z. An optimization model for express delivery with high-speed railway. *Transp. Res. Part E* **2023**, 176, 103206. [CrossRef]
32. Park, H.M.; Cho, S.W.; Seo, K.M. Hybrid path planning framework to integrate improved A*-DWA algorithms for enhancing path safety and efficiency. *Appl. Ocean. Res.* **2025**, 157, 104497. [CrossRef]
33. Thabet, A.S.M.; Zengin, A. Optimizing Dijkstra's Algorithm for Managing Urban Traffic Using Simulation of Urban Mobility (Sumo) Software. *Ann. Math. Phys.* **2024**, 7, 206–213. [CrossRef]
34. Liu, M.; Li, Y.; Liu, X.; Chen, Y.; Hao, R. An Integrated Optimization Framework for Connected and Automated Vehicles and Traffic Signals in Urban Networks. *Systems* **2025**, 13, 224. [CrossRef]
35. Yun, T.S.; Moon, I.C. Housing market agent-based simulation with loan-to-value and debt-to-income. *J. Artif. Soc. Soc. Simul.* **2020**, 23, 5. [CrossRef]
36. Risco-Martín, J.L.; Mittal, S.; Henares, K.; Cardenas, R.; Arroba, P. xDEVS: A toolkit for interoperable modeling and simulation of formal discrete event systems. *Softw. Pract. Exp.* **2023**, 53, 748–789. [CrossRef]
37. Transportation Card Big Data System. Route and Bus Stop Indicators. 2024. Available online: <https://stcis.go.kr/pivotIndi/wpsPivotIndicator.do?siteGb=P&indiCls=IC03> (accessed on 1 March 2024).
38. National Index System. Public Transportation Share Ratio. 2024. Available online: <https://www.index.go.kr/unify/idx-info.do?idxCd=4259> (accessed on 1 March 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Algorithmic Identification of Conflicting Traffic Lights: A Large-Scale Approach with a Network Conflict Matrix

Sergio Rojas-Blanco ^{1,*}, Alberto Cerezo-Narváez ¹, Sol Sáez-Martínez ² and Manuel Otero-Mateo ¹

¹ Department of Mechanical Engineering and Industrial Design, University of Cádiz, Avda. Universidad de Cádiz, nº 10, Campus Universitario de Puerto Real, 11519 Puerto Real, Spain; alberto.cerezo@uca.es (A.C.-N.); manuel.otero@uca.es (M.O.-M.)

² Department of Mathematics, University of Cádiz, Avda. Universidad de Cádiz, nº 10, Campus Universitario de Puerto Real, 11519 Puerto Real, Spain; sol.saez@uca.es

* Correspondence: sergio.rojas@uca.es

Abstract: Efficient urban traffic management is crucial for mitigating congestion and enhancing road safety. This study introduces a novel algorithm, with code provided, to generate a traffic light conflict matrix, identifying potential signal conflicts solely based on road network topology. Unlike existing graphical approaches that are difficult to execute automatically, our method leverages readily available topological data and adjacency matrices, ensuring broad applicability and automation. While our approach deliberately focuses on topology as a stable foundation, it is designed to complement rather than replace dynamic traffic analysis, serving as an essential preprocessing layer for subsequent temporal optimization. Implemented in MATLAB, with specific functionality for Vissim users, the algorithm has been tested on various networks with up to 547 traffic lights, demonstrating high efficiency, even in complex scenarios. This tool enables focused allocation of computational resources for traffic light optimization and is particularly valuable for prioritizing emergency vehicles. Our findings make a significant contribution to traffic management strategies by offering a scalable and efficient tool that bridges critical gaps in current research. As urban areas continue to grow, this algorithm represents a step forward in developing sustainable solutions for modern transportation challenges.

Keywords: urban road network; traffic light control; intelligent transportation system; Vissim; traffic conflict; adjacency matrix

1. Introduction

Urban traffic management presents significant challenges, with congestion and accidents negatively impacting quality of life and economic productivity. In 2019, traffic accidents in the United States resulted in 36,500 deaths and a USD 340 billion economic cost [1]. In 2022, congestion caused average delays of 54 h per traveler, leading to USD 224 billion in losses [2]. Addressing these issues requires efficient traffic light control, especially in complex networks where signal conflicts occur within and between intersections [3]. Traffic analysis and optimization is influenced by internal factors, like network topology, and external factors, such as events and weather conditions.

To address these challenges, research has increasingly focused on automating traffic signal management. Coordinating traffic signals is a proven strategy for optimizing transportation systems [4,5] and enhancing road safety [6]. Effective coordination requires understanding the relationships among traffic signals, which must be established by manipulating available data.

The evolution of traffic analysis techniques has expanded from isolated intersections to complex spatio-temporal relationships across entire networks [7,8]. These methods require both spatial and temporal information, and they are often represented through graphs or adjacency matrices. For instance, Spatio-Temporal Multi-Agent Reinforcement Learning (STMARL) [9] constructs a directional traffic light adjacency graph based on spatial structures before analyzing temporal dependencies.

While some studies focus on isolated intersections and attempt to relate them through adjacency [10,11], our approach, similar to Wang et al.'s, focuses on the relationships between individual traffic lights across the entire network, allowing for a more comprehensive understanding of potential conflicts.

Our method identifies traffic light conflicts using network topology, offering several advantages:

1. Stable conflict detection that is unaffected by temporary changes in traffic patterns or signal timing.
2. Large-scale analysis without costly real-time data collection.
3. Proactive conflict identification and resolution based on structural relationships.

This topological approach is valuable for long-term urban planning and serves as a preprocessing step for detailed optimization techniques. Building on our previous work on traffic light adjacency list algorithms [12], we introduce a scalable method for identifying conflicting traffic lights using readily accessible GIS datasets.

Key contributions of this work include the following:

1. A scalable methodology to identify conflicting traffic lights using only topological data.
2. Utilization of widely available network infrastructure datasets.
3. Extension of adjacency list algorithms to address traffic light conflicts.
4. Enhanced traffic flow continuity through conflict identification and resolution.
5. Improved emergency vehicle prioritization through immediate spatial coordination.
6. A practical, sustainable tool for urban traffic management.

These contributions address critical gaps in traffic optimization and provide solutions for improving urban transportation networks. Our approach presents a foundational layer of information that is crucial for understanding and optimizing urban traffic patterns based on network topology.

To ensure reproducibility and facilitate further research, we have attached all the necessary MATLAB codes as complementary files. These codes allow for the complete reproduction of our algorithm on the four networks used in our experiments. By providing these implementation details, we aim to support transparency in research and enable other researchers to build upon or validate our work.

Section 2 reviews related works in traffic signal control and network analysis. Section 3 provides the key definitions used in our methodology. Section 4 details the proposed algorithm for generating the traffic light conflict matrix. Section 5 presents the experimental setup and results, demonstrating the algorithm's performance across four different urban networks. Section 6 discusses the implications of our findings, and Section 7 offers conclusions and recommendations for future research. Finally, Appendix A provides detailed examples, illustrating the methodology for both isolated intersections and network-wide scenarios, along with Supplementary CSV Files, which contain complete matrices to enhance reproducibility and facilitate further analysis.

2. Related Works

The evolution of traffic signal control (TSC) systems has been driven by the need to manage increasing urban congestion effectively. From isolated intersections to large-scale

networks, all strategies require an understanding of the relationships between traffic signals. While most strategies concentrate on isolated intersections due to scalability and lower computational costs, research on optimizing complete networks has grown significantly, particularly with the use of computational intelligence (CI) models [13].

CI tools are well suited for identifying spatial correlations and time-based events within networks [14], often utilizing graphs where nodes represent study objectives and edges indicate their relationships [15,16]. However, many publications overlook the relationships between signals or provide non-specific suggestions on obtaining these relationships. Some studies consider phase information [17], while others require additional data about intersections [18] or rely on network topology [9].

In our previous work [12], we proposed an algorithm that establishes relationships among traffic signals in a complete network using an adjacency list based on topology. This approach provides valuable information for generating green waves and prioritizing traffic signals for emergency vehicles.

Li et al. [19] developed a higher-order conflict graph approach for network-wide traffic signal control. Their method models traffic signal configurations using a conflict graph, where nodes represent movements permitted by specific signal phases and edges connect conflicting movements. While our approach shares similarities with Li's work in considering movements permitted by signal heads, there are key differences in methodology and focus.

We employed adjacency matrices derived from our previous work to calculate the *traffic light conflict matrix*, allowing for efficient computation and representation of conflicts between signal heads. Our algorithm automates the entire process of generating the conflict graph, eliminating manual intervention and improving scalability. Our current focus is on efficiently identifying and representing conflicts without addressing timing optimization, serving as a crucial filter for other optimization tools. The *traffic light conflict matrix* generated by our method is particularly valuable for emergency vehicle prioritization systems, providing a clear representation of potential conflicts for efficient emergency routing.

These distinctions highlight the different objectives of our respective methods for analyzing traffic signal networks, despite the shared basis in movement-based conflict graphs.

3. Definitions

3.1. Adjacency Matrix

Consider an undirected graph $G = (V, E)$ with vertex set V and edge set E . The adjacency matrix representation A of G constitutes a square matrix of order $n = |V|$. The entries A_{ij} of this matrix are determined by the following rule:

$$A_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E, \\ 0 & \text{if } (v_i, v_j) \notin E, \end{cases}$$

where vertices $v_i, v_j \in V$.

For directed graphs, we modify this formulation accordingly:

$$A_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \text{ forms a directed edge from } v_i \text{ to } v_j, \\ 0 & \text{otherwise.} \end{cases}$$

The adjacency matrix, thus, encodes the graph structure G into numerical form, where each non-zero entry A_{ij} signifies connectivity between vertices v_i and v_j . In weighted graphs, these entries may instead correspond to edge weights rather than binary values.

3.2. Traffic Light Conflict Matrix

The *traffic light conflict matrix* C is an $n \times n$ symmetric matrix that represents conflicts between traffic signal heads in a road network. Each element C_{ij} indicates whether signal heads i and j control conflicting movements.

$$C_{ij} = \begin{cases} 1, & \text{if signal heads } i \text{ and } j \text{ control conflicting movements} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Alternatively, the *traffic light conflict matrix* can be computed as follows:

$$C = A \cdot A^T, \quad (2)$$

where

- A is the adjacency matrix describing the connections between real signal heads (SH) and crossing signal heads (CH).
- A^T is the transpose of the adjacency matrix.
- The product $A \cdot A^T$ determines the conflicts between signal heads based on shared connections to conflict points.

This formulation is useful in traffic network analysis, allowing for efficient conflict detection using matrix operations.

3.3. Road

A road is a transportation route designed for vehicles and pedestrians. It is typically paved, publicly accessible, and connects towns, cities, regions, or countries.

3.4. Link

In this study, a link refers to a segment of road infrastructure, modeling roads, streets, lanes, or other transportation pathways. A link can be considered a segment or representation of a road within a traffic simulation or model. Some considerations when representing roads using links include the following:

- Unidirectionality and bidirectionality: In the case of a bidirectional road, two separate links are modeled, one for each direction.
- Direction of traffic: When simulating traffic with PTV Vissim, ensure that the direction of traffic flow on the links matches the actual road design.
- Connections: Connections between different links must occur at endpoints, not at intermediate points.

Figure 1 shows an example of a right-hand traffic road with two-way traffic that is analytically divided into two links.

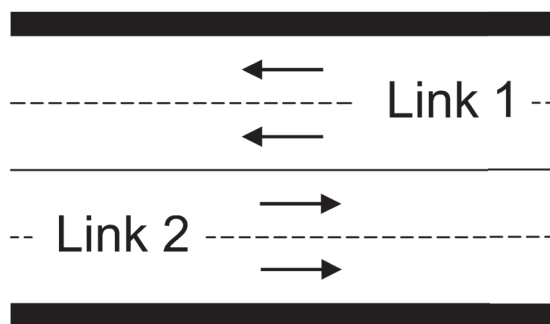


Figure 1. Two-way road represented by two links, each for a different traffic direction.

3.5. Traffic Light

A traffic light is a signaling device that regulates traffic flow at intersections, pedestrian crossings, and key points in a road network. Its main function is to ensure safe and efficient movement by controlling the right of way for vehicles and pedestrians.

3.6. Signal Head

A signal head is a simulated object that replicates the function of a real traffic light within a road network. It represents traffic signals, including control lights for different traffic directions and users, such as vehicles and pedestrians.

3.7. Endpoint Signal Head

An endpoint signal head is a fictitious traffic signal introduced at network boundaries to represent the endpoints of a road network. It helps define the system's limits and facilitates the analysis of traffic flow interactions beyond intersections.

3.8. Cut Signal Head

A crossing signal head is a virtual signal introduced at conflict points where vehicle paths intersect within an intersection or road network. It is used to model and identify potential traffic signal conflicts based on movement patterns.

4. Materials and Methods

4.1. Challenges in Network Analysis

The implementation of our methodology faced several significant challenges during development. When working with complex traffic networks, we identified three primary obstacles that required specific solutions.

First, establishing relationships between traffic lights belonging to intersecting roads proved particularly complex. Our solution introduced “fictional crossing signal heads” at intersection points, allowing us to model conflict areas without modifying existing physical infrastructure.

Second, developing an algorithm to accurately locate points where roads intersect presented considerable computational difficulties. Our current implementation uses a basic geometric approach that identifies potential crossing points based on the spatial coordinates of road segments. While functional, this method has limitations when dealing with complex road geometries or multi-level intersections. The algorithm searches for points where line segments representing road centerlines intersect in a two-dimensional plane, which works well for standard grid-like networks but may require refinement for more complex urban configurations.

Third, traffic network designers employ varying methodologies for connecting roads: some link the end of one road to the beginning of another, while others create connections at intermediate points. This inconsistency necessitated developing flexible approaches to conflict identification that could accommodate different network construction styles within simulation environments like Vissim.

We validated our algorithm through manual verification in smaller networks and strategic regions of larger, more complex networks. These tests confirmed that our approach correctly identifies structural conflicts in the vast majority of cases, though we acknowledge the inherent limitations of purely topological analysis.

4.2. Conceptual Framework

The conflict relationships between signal heads can be represented as an undirected graph $G_c = (V_c, E_c)$, where $V_c = \{v_1, v_2, \dots, v_N\}$ is a set of N signal heads, and E_c is the set

of edges that represent conflict relationships between signal heads. The *traffic light conflict matrix* $C \in \mathbb{R}^{N \times N}$ represents these conflicts, where $c_{ij} = c_{ji} = 1$ if signal heads i and j control conflicting states, and 0 otherwise.

This representation captures the bidirectional nature of conflicts between signal heads.

4.3. Methodology Overview

Our methodology for constructing the *traffic light conflict matrix* involves a systematic process that can be applied to both isolated intersections and large-scale road networks. The key steps include the following:

1. **Network Representation:** Representing the road network as a graph, identifying real signal heads (SH), and introducing fictitious endpoint signal heads (EP) and crossing signal heads (CH) to model network boundaries and conflict points, respectively.
2. **Adjacency Matrix Generation:** Constructing an adjacency matrix A that captures the relationships between real signal heads and both endpoint and crossing signal heads.
3. **Traffic Light Conflict Matrix Computation:** Using the adjacency relationships represented in matrix A to compute the *traffic light conflict matrix* C , where each element indicates whether two signal heads control conflicting movements.

This process allows for the efficient identification of potential signal conflicts based on network topology, providing a foundation for optimizing traffic signal control strategies.

Figure 2 summarizes our methodology.

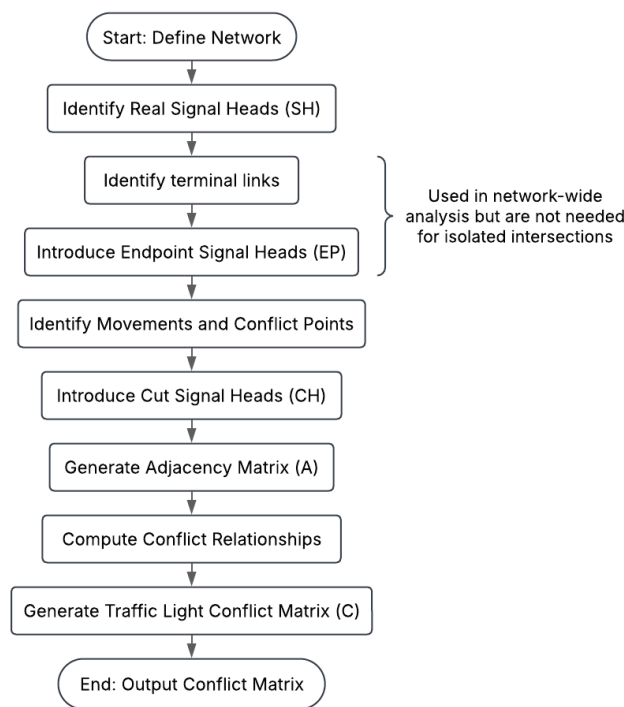


Figure 2. Overview of the methodology.

4.3.1. Isolated Intersection Analysis

For an isolated intersection, conflicts arise when two or more traffic movements intersect within the same space. Each movement is controlled by a real signal head (SH). To systematically detect conflicts, we introduce crossing signal heads (CHs), fictitious elements placed at movement intersection points. Each CH represents a conflict involving multiple SHs. These CHs allow us to construct an adjacency matrix, where each 1 indicates a conflict between an SH and a CH.

Then, we generated the intersection's adjacency matrix, which includes both real signal heads and fictional crossing signal heads. Each crossing signal head is introduced at a movement intersection point to represent a conflict. These conflict points serve as the input data for our algorithm, allowing for systematic identification of signal conflicts.

Figure 3 illustrates an example of an isolated signalized intersection with four roads, its corresponding conflict graph, and the associated *traffic light conflict matrix*.

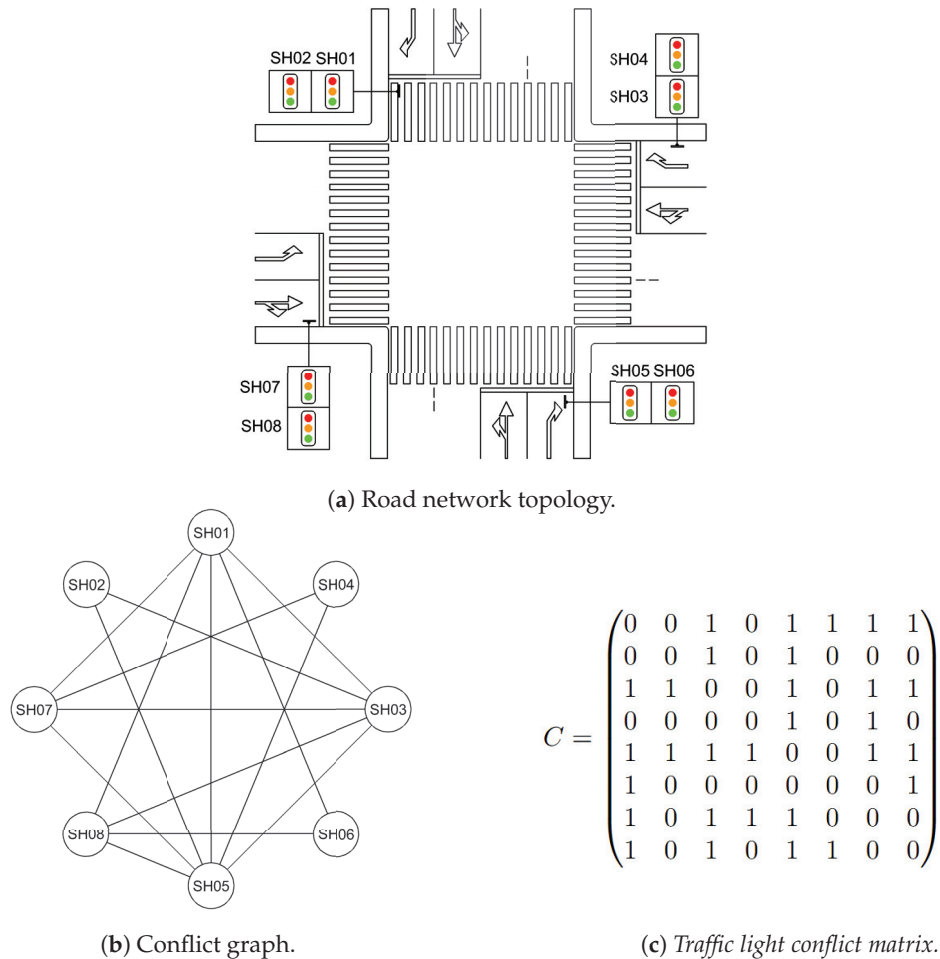


Figure 3. Example of an isolated intersection.

In this representation, we have the following:

- The intersection scheme displays the physical layout, including lanes, permitted movements, and the signal heads controlling these movements.
- In the conflict graph, each node represents a signal head controlling a specific movement through the intersection.
- Edges in the conflict graph connect nodes (signal heads) that control conflicting movements, directly representing the conflict relationships.
- The *traffic light conflict matrix* provides a numerical representation of the conflict graph, where each element c_{ij} indicates whether signal heads i and j control conflicting movements (1) or not (0).

In our example intersection, 8 real signal heads control various movements. We identified 34 conflict points, each assigned a fictional crossing signal head (CH01–CH34) based on its associated real signal heads.

This information is used to generate the corresponding adjacency matrix. A simplified version is shown in Figure 4, displaying only the most relevant adjacencies for clarity. The full matrix includes all 34 columns for the crossing signal heads, but, for brevity, only a

subset is shown. A value of 1 in the matrix indicates an adjacency between a real signal head (SH) and a crossing signal head (CH), while 0 represents no adjacency.

$$A = \begin{pmatrix} & \text{CH01} & \text{CH02} & \text{CH03} & \text{CH04} & \text{CH05} & \text{CH06} & \text{CH07} & \text{CH08} & \text{CH09} & \text{CH10} & \text{CH11} & \text{CH12} & \text{CH13} & \text{CH14} & \dots & \text{CH34} \\ \text{SH01} & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \text{SH02} & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \text{SH03} & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & \dots & 0 \\ \text{SH04} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & \dots & 0 \\ \text{SH05} & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ \text{SH06} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \text{SH07} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ \text{SH08} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Figure 4. Adjacency matrix for an isolated intersection with real and fictional signal heads.

Each CH represents a specific conflict involving multiple real signal heads. In the adjacency matrix, all SHs associated with that conflict will have a 1 in the corresponding CH column.

Building on this, the adjacency matrix provides the foundation for constructing the *traffic light conflict matrix*. By analyzing its columns, we can systematically identify conflicts: signal heads that share a 1 in the same column are considered to be in conflict. This relationship can be expressed mathematically as follows:

$$C_{ij} = \sum_{k=1}^m A_{ik} \cdot A_{jk}, \quad (3)$$

where

- A is the adjacency matrix of dimensions $n \times m$, where n is the number of real signal heads (SH) and m is the number of crossing signal heads (CH).
- C is the *traffic light conflict matrix* of dimensions $n \times n$.
- C_{ij} is the element in row i and column j of the *traffic light conflict matrix* C .
- A_{ik} is the element in row i and column k of the adjacency matrix A .
- A_{jk} is the element in row j and column k of the adjacency matrix A .

This formula captures the idea that two real signal heads i and j are in conflict if and only if there exists at least one crossing signal head k for which both A_{ik} and A_{jk} are 1.

The process of transforming the adjacency matrix into the *traffic light conflict matrix* involves the following steps:

1. For each real signal head (SH), we identify its adjacent crossing signal heads (CH) by looking for 1s in the corresponding row of the adjacency matrix.
2. For each adjacent CH identified, we then search for other SHs that are also adjacent to this CH by looking for 1s in the same column.
3. These other SHs are in conflict with the original SH, and we mark this conflict in the *traffic light conflict matrix* with a 1.

Figure 5 shows the resulting *traffic light conflict matrix*.

$$C = \begin{pmatrix} & \text{SH01} & \text{SH02} & \text{SH03} & \text{SH04} & \text{SH05} & \text{SH06} & \text{SH07} & \text{SH08} \\ \text{SH01} & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ \text{SH02} & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ \text{SH03} & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ \text{SH04} & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ \text{SH05} & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ \text{SH06} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \text{SH07} & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ \text{SH08} & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

Figure 5. A *traffic light conflict matrix* for an isolated intersection.

This process can be simplified to identifying the conflict points between movements controlled by signal heads, as this directly provides the conflicting signal heads. It is important to note that Figures 3c and 5 present the same *traffic light conflict matrix* but with different purposes. Figure 3c is part of an illustrative set (Figure 3) that conceptually demonstrates the expected relationships in an isolated intersection, while Figure 5 represents the actual computational result obtained by following the systematic algorithmic process described above. This dual presentation reinforces how our theoretical framework translates into practical implementation.

The methodology we have described above, involving the introduction of fictional signal heads at these conflict points, justifies this approach when dealing with large-scale networks. Following on from this, we will describe the general methodology and the algorithm applicable to any road network.

4.3.2. Network-Wide Analysis

We will now extend the methodology to entire road networks presents additional challenges, such as the following:

- Managing conflicts across multiple intersections.
- Defining network boundaries for accurate traffic modeling.

To address these challenges, as shown in the methodology indicated above, we introduce another class of fictitious signal heads: endpoint signal heads (EP). These act as network boundary markers, ensuring that the traffic flow at the edges of the analyzed network is appropriately modeled.

To better visualize this concept, Figure 6 presents a network with four intersections, where real signal heads are displayed alongside their permitted movements, and endpoint signal heads are also indicated.

Examining Figure 6, we can identify key adjacency relationships between signal heads (SH) and endpoint signal heads (EP) in the network. These relationships form the basis for understanding traffic flow and potential conflicts within the system.

Based on the adjacency relationships observed in the figure, we generated the adjacency matrix shown below (Figure 7).

The complete adjacency matrix contains 32 real signal heads and 8 endpoint signal heads, resulting in a 32×40 matrix. Endpoint signal heads, by definition, do not have adjacencies to other signal heads, so their adjacencies are not represented in the matrix. Here, we present a representative subset to facilitate visibility and understanding of the network's connectivity structure.

These adjacency relationships illustrate how different signal heads interconnect and how endpoint signal heads represent the boundaries of the analyzed network.

After introducing endpoint signal heads to define the network boundaries, the next step is to identify and introduce the crossing signal heads. The specific methodology for identifying these crossing points is beyond the scope of this analysis. However, Vissim users can utilize a MATLAB function, which is provided in the accompanying documentation, to locate these crossing points within their network models. In complex network configurations, it is crucial to recognize that the conflict points resulting from link cuts may be either signalized or unsignalized. Consequently, the signal heads nearest to an conflict point might not always be situated on the links directly involved in that intersection. Instead, they may be positioned on more distant links within the network structure. This spatial distribution of signal heads adds an extra layer of complexity to the analysis and must be carefully considered when developing traffic control strategies.

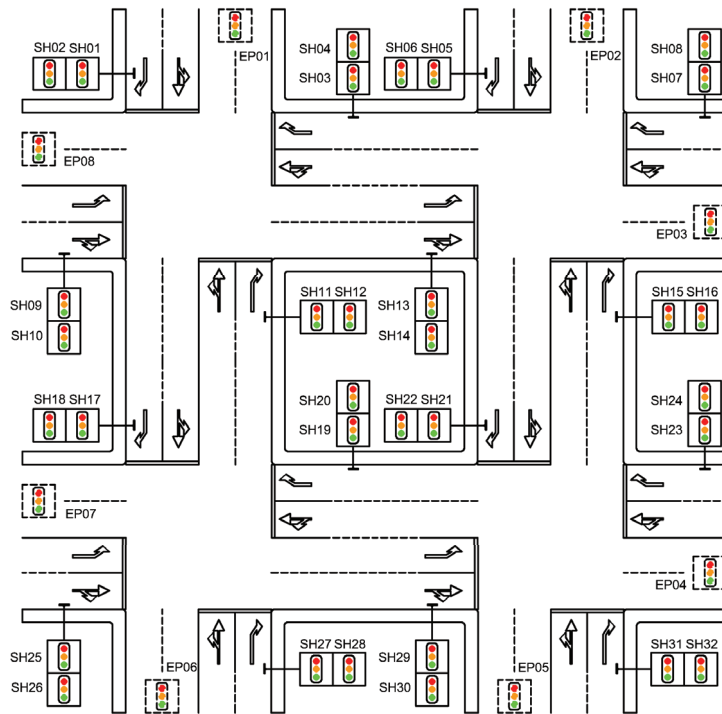


Figure 6. The real and endpoint signal heads in the network.

$$A_1 = \begin{pmatrix} & SH01 & SH02 & SH03 & \dots & SH13 & SH14 & SH15 & SH16 & SH17 & SH18 & \dots & EP06 & EP07 & EP08 \\ SH01 & 0 & 0 & 0 & \dots & 1 & 1 & 0 & 0 & 1 & 1 & \dots & 0 & 0 & 0 \\ SH02 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ SH03 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 1 & 1 & \dots & 0 & 0 & 1 \\ SH04 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ SH05 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ SH31 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ SH32 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

Figure 7. Adjacency matrix for a network with real and endpoint signal heads.

In our example case, we identified the crossing links and added crossing signal heads at each conflict point. Based on the crossing signal heads, we can generate the adjacency matrix relating to real signal heads (SH) to crossing signal heads (CH). Figure 8 shows this matrix. Due to the large size of the matrix, we present a subset focusing on a few signal heads.

$$A_2 = \begin{pmatrix} & CH01 & CH02 & CH03 & CH04 & \dots & CH020 & CH021 & CH022 & CH023 & \dots & CH57 & CH58 & CH59 \\ SH01 & 1 & 1 & 1 & 1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ SH02 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ SH03 & 1 & 0 & 0 & 0 & \dots & 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \end{pmatrix}$$

Figure 8. Adjacency matrix for a network with real and crossing signal heads.

The complete adjacency matrix A for the network-wide analysis was constructed by horizontally concatenating A_1 and A_2 .

$$A = (A_1 | A_2).$$

This resulted in a matrix that represents the adjacency relationships between the real signal heads and both endpoint signal heads and crossing signal heads. The rows of A correspond

to the real signal heads, while the columns represent the endpoint signal heads (from A_1) and the crossing signal heads (from A_2).

Once the complete adjacency matrix A was obtained, the calculation of the *traffic light conflict matrix* (C) followed the same procedure as for an isolated intersection. This matrix was derived through the following matrix operation:

$$C = A \cdot A^T, \quad (4)$$

where A^T is the transpose of matrix A . After this matrix multiplication, two additional steps were necessary to obtain the final *traffic light conflict matrix*:

1. Binarization: Any non-zero value in the resulting matrix was converted to 1, indicating the presence of a conflict.
2. Diagonal nullification: The diagonal elements of the matrix were set to zero as a signal head cannot be in conflict with itself.

The complete process can be expressed mathematically as follows:

$$C_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \text{ and } \sum_k A_{ik} \cdot A_{jk} > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

This process ensures that the matrix multiplication captures all of the conflict relationships: two real signal heads will be in conflict if and only if both share at least one connection with the same fictitious signal head (either an endpoint signal head or a crossing signal head).

For our example network, the resulting conflict matrix had the dimensions of 32×32 , corresponding to the 32 real signal heads. Due to its size, we present below a representative submatrix showing the conflict relationships for a subset of signal heads.

$$C = \begin{pmatrix} & SH01 & SH02 & SH03 & SH04 & SH05 & SH06 & SH07 & SH08 & \dots & SH31 & SH32 \\ SH01 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ SH02 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ SH03 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & \dots & 0 & 0 \\ SH04 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & \dots & 0 & 0 \\ SH05 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ SH06 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ SH07 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ SH08 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ SH31 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ SH32 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

In this matrix, each element C_{ij} indicates whether a conflict exists between the real signal heads i and j . A value of 1 indicates that both signal heads control movements that are in conflict, while a value of 0 indicates no conflict between them. For example, we can observe that SH01 is in conflict with SH03, while it has no conflicts with SH02, SH04, etc.

The complete matrix provides a comprehensive representation of all possible conflicts between signal heads across the entire network, serving as a foundation for traffic control strategy optimization. This representation facilitates the identification of signal head groups that can have the same state (green) simultaneously, as well as those that must necessarily have different states to avoid conflicts.

It is important to note that the conflict matrix is symmetric ($C_{ij} = C_{ji}$), reflecting the bidirectional nature of conflicts: if signal head i is in conflict with signal head j , then signal head j is also in conflict with signal head i .

4.3.3. Algorithm for Generating the Traffic Light Conflict Matrix

The methodology for creating the *traffic light conflict matrix* (C) in general networks involves identifying signal heads, their relationships, and potential conflicts. Algorithm 1 outlines the steps to generate the conflict matrix.

Algorithm 1 Algorithm for generating the *traffic light conflict matrix*

Input: Adjacency matrix A between real signal heads and fictional signal heads

Output: Traffic light conflict matrix C

```

1:  $C_{raw} \leftarrow A \cdot A^T$  ▷ Matrix multiplication
2: Initialize  $C$  with zeros of size  $n \times n$  ▷  $n$  is the number of real signal heads
3: for each pair  $(i, j)$  where  $i, j \in \{1, 2, \dots, n\}$  do
4:   if  $i \neq j$  AND  $C_{raw}(i, j) > 0$  then
5:      $C(i, j) \leftarrow 1$  ▷ Binarization and diagonal nullification
6:   end if
7: end for
8: return  $C$ 

```

The algorithm involves several key computational steps:

1. Construction of the adjacency matrix A : This matrix represents the connections between real signal heads and fictional signal heads (both endpoint and crossing signal heads).
2. Matrix multiplication: The product $A \cdot A^T$ calculates a raw conflict matrix, where the value at position (i, j) indicates the number of conflict points shared between signal heads i and j .
3. Binarization: Since we are only interested in whether a conflict exists (not how many conflict points are shared), we convert all positive values to 1.
4. Diagonal nullification: A signal head cannot be in conflict with itself, so all diagonal elements are set to zero.

This process is implemented in our algorithm by applying the following rule:

$$C_{ij} = \begin{cases} 0, & \text{if } i = j \text{ (no self-conflicts)} \\ \text{binarize}\left(\sum_k A_{ik} \cdot A_{jk}\right), & \text{if } i \neq j \end{cases}, \quad (6)$$

where the binarize function converts any positive value to 1, representing the existence of at least one conflict point.

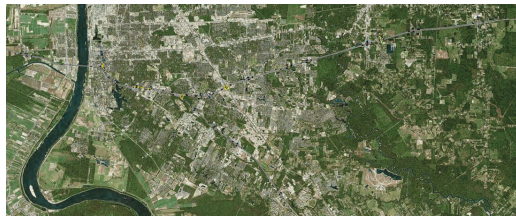
While the *traffic light conflict matrix* is ultimately represented as an adjacency matrix, the internal computations leverage adjacency lists for efficiency, particularly given the sparse nature of typical road network graphs. This matrix-based approach allows for efficient computational implementation while ensuring all of the conflict relationships are accurately captured.

This implementation ensures a comprehensive identification of all the potential signal conflicts based on the network's topological structure.

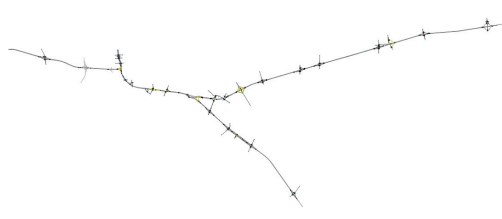
5. Experiments and Results

To validate our conflict matrix algorithm, we selected four diverse urban networks from PTV Vissim [20]: the Baton Rouge corridor [21], a central London district, the Boise

metropolitan area, and a Luxembourg urban sector. Satellite imagery and corresponding road networks for these test cases appear in Figures 9–12.

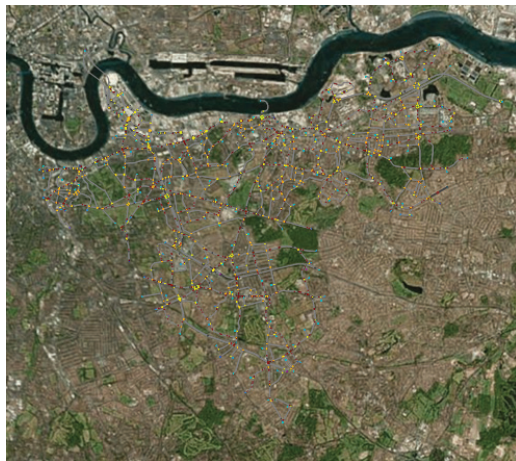


(a) Road map.



(b) Road lines scheme.

Figure 9. Map and road network of Baton Rouge (US).



(a) Road map.



(b) Road lines scheme.

Figure 10. Map and road network of London (GB).



(a) Road map.



(b) Road lines scheme.

Figure 11. Map and road network of Boise (US).

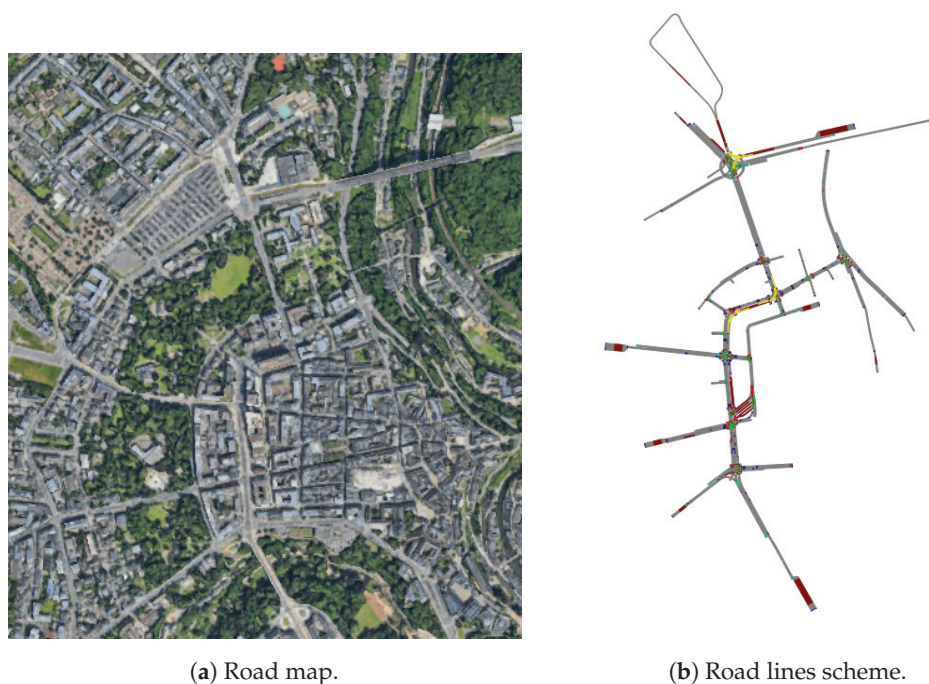


Figure 12. Map and road network of Luxembourg (LU).

We conducted testing on modest hardware: an Intel i5-4310U processor (2.00 GHz base, 2.60 GHz turbo), 8 GB RAM, and Windows 10 Enterprise 64-bit (version 22H2). All implementations used MATLAB R2024a. To establish reliable performance metrics, we processed each network ten consecutive times and calculated average execution times.

The results obtained are summarized in Table 1. For each network, the number of links, the number of intersections, the number of signal heads, the number of conflicts, and the average execution time for 10 iterations of the function are presented. These results demonstrate the efficiency of the algorithm in networks of varying sizes and configurations.

It should be noted that the execution times reported in the table correspond specifically to the conflict matrix generation once the intersection points between links have been identified. Our current implementation for detecting these intersection points (the `Crossing_links` function) is relatively basic and could be improved for complex geometries. However, we have included it as a subfunction within `Adjacency_list_of_signal_heads_extended` in the Supplementary Material section to ensure full reproducibility of our results. The processing time for this intersection detection step varies depending on network complexity, and is not included in the reported execution times as it represents a preprocessing step that can be optimized independently.

Table 1. Network information and *traffic light conflict matrix* computation results.

Network	Number of Links	Number of Signalized Intersections	Number of Signal Heads	Number of Conflicts	Execution Time (Seconds)
Baton Rouge (US)	2038	139	345	320	7.375842
London (GB)	8372	240	265	692	4.968577
Boise (US)	720	72	547	101,971	16.344132
Luxembourg (LU)	416	25	126	310	1.1717

The Supplementary Materials section contains our complete implementation, providing researchers with everything needed to reproduce our work. These mate-

rials include our full MATLAB source code along with the specific datasets used in our analysis: `Baton_Rouge_dataset.mat`, `London_dataset.mat`, `Boise_dataset.mat`, and `Luxembourg_dataset.mat`. By sharing these resources, we enable independent verification and extension of our findings by interested researchers.

6. Discussion

Our algorithm for generating the *traffic light conflict matrix* introduces an innovative approach to traffic management by systematically identifying potential signal head conflicts through network structural relationships. By leveraging topological data instead of traditional phase diagrams or real-time traffic information, we provide a universal method for conflict detection that transcends the limitations of existing intersection-level analyses.

A key advantage of our method is its ability to fully automate the conflict detection process. The most relevant prior work, Li et al., introduced a graph-based conflict representation for traffic signals but did not provide a scalable automation framework. Instead, their conflict graphs were manually derived from isolated intersections and networks, making their approach impractical for large-scale implementation. Our algorithm eliminates this limitation by automating the entire process, making it feasible for real-world traffic management at an urban scale.

An alternative approach to traffic signal conflict detection could involve machine learning or artificial intelligence (AI) methods. However, such solutions would present several significant drawbacks:

1. **Computational Cost.** AI models require significant training and inference time, making them far more expensive to run compared to our deterministic matrix-based approach.
2. **Data Dependency.** AI would require large, labeled datasets of real-world traffic conflicts, which are not readily available across different cities.
3. **Generalization Issues.** AI models trained on one urban area may fail to generalize to different city layouts without retraining.
4. **Lack of Transparency.** AI-based solutions, particularly deep learning models, function as black boxes, making it difficult to interpret why a certain conflict was identified. Our matrix-based method, in contrast, provides a clear, interpretable representation of conflicts, ensuring that traffic engineers can validate and trust the results.

Thus, our method provides an optimal balance between accuracy, efficiency, and scalability. It requires only readily available topological data, processes large-scale networks in milliseconds, and integrates seamlessly with existing traffic optimization frameworks. While future research may explore AI-based enhancements, our approach presents the most practical, cost-effective, and interpretable solution for large-scale urban traffic management.

One consideration for the applicability of our method is its validation with real-world traffic conditions. Our experiments were conducted on networks ported from Visum and Vissim, which are widely used in urban mobility research and transportation planning. These networks accurately model real-world road layouts and traffic infrastructures, ensuring a realistic representation of network topology and signal configurations.

Since our *traffic light conflict matrix* is designed as a modular input for traffic light analysis and optimization, it can be integrated with other data-driven approaches, including real-time traffic data, AI-based analysis, or other optimization techniques. The strength of our method lies in its flexibility—it provides a stable, structural foundation for traffic signal studies without being constrained by specific data sources.

An important consideration in our implementation is the classification of conflicts by severity. Recognizing that not all conflicts pose the same level of hazard in real-world traffic scenarios, our method distinguishes between two types of conflicts:

1. **Convergent conflicts:** These occur when two real signal heads converge directly, typically representing merging or diverging movements. These conflicts generally pose lower safety risks.
2. **Crossing conflicts:** These occur at road intersection points where real signal heads connect through a fictional signal head at a crossing point. These conflicts typically represent more severe hazards, such as those arising from left-turn movements crossing opposing traffic.

This classification aligns with established traffic engineering principles that consider crossing conflicts more hazardous than merging conflicts. Our implementation not only identifies conflicts, but also categorizes them by type, providing traffic engineers with critical information to prioritize the resolution of higher-risk conflicts when designing signal timing plans. This enhancement improves both the practical utility and safety implications of our conflict detection approach.

Our topological approach is designed to serve as a structural foundation that can be integrated with real-time applications through a layered architecture. This architecture consists of three primary components: (1) our topological conflict matrix as a static base layer that remains valid regardless of traffic conditions, (2) a dynamic layer incorporating real-time traffic volume data from sensors or cameras, and (3) an optimization layer that combines both structural constraints and current conditions to make adaptive decisions. This layered approach significantly reduces computational complexity by pre-identifying which signals must be considered together during real-time optimizations. For example, in emergency vehicle prioritization, our conflict matrix immediately identifies which signals must be coordinated when an emergency vehicle approaches, allowing the system to focus only on relevant conflict relationships rather than analyzing the entire network. Similarly, for adaptive signal timing, optimization algorithms can be applied only to the specific sets of conflicting signals identified by our matrix, rather than attempting global optimization across all signals. This strategic reduction in the search space is particularly valuable in time-sensitive applications, where computational efficiency directly impacts response time and system performance. By providing these essential structural constraints, our approach enables more efficient and effective real-time traffic management solutions while maintaining flexibility to incorporate changing traffic conditions.

However, we acknowledge several limitations and opportunities for future research:

- **Real-time Adaptation:** Integrating dynamic traffic data could enhance the algorithm's responsiveness.
- **Validation:** Extensive real-world testing is necessary to fully validate the method's effectiveness.
- **System Integration:** Further research is needed to incorporate the conflict detection method into existing traffic management systems.
- **Comprehensive Considerations:** Future iterations could incorporate pedestrian and cyclist interactions for a more holistic approach.

In conclusion, our study offers a practical and scalable method for identifying potential conflicts across entire networks. This tool allows us to focus calculation efforts on critical traffic lights when applying timing optimization techniques, thereby decreasing computational costs and processing time. The implications of this work extend beyond immediate traffic control. It enables city planners and traffic engineers to make data-driven decisions that can reduce congestion, lower emissions, and improve overall urban mobility. As cities continue to grow and evolve, tools like ours will be crucial in creating more efficient, safer, and more livable urban environments.

For users looking to implement our approach, we designed it to minimize adoption barriers. Vissim users can easily incorporate our solution using the MATLAB scripts, which extract all necessary network information and generate adjacency and conflict matrices automatically, included in the Supplementary Materials section. Detailed step-by-step instructions are provided in the included Readme.txt file. For municipalities without Vissim, our approach can still be applied using the network datasets provided in our Supplementary Materials section or can be adapted to other topological data sources. Traffic engineers with basic MATLAB knowledge can implement our solution with minimal training, as the process is largely automated through our documented scripts. We are currently developing a web platform where researchers and municipal engineers can upload their network data (initially in Vissim format, with plans to expand to OpenStreetMap, GeoFabrik, and Google Maps) and automatically receive complete conflict analysis results. This platform will further reduce implementation barriers, especially for smaller municipalities with limited technical resources. While this platform is under development, we offer direct assistance to any municipality wishing to test our approach with their networks. Our method represents a cost-effective enhancement to existing traffic management systems rather than requiring significant infrastructure changes or specialized training.

7. Conclusions

This study presents several significant contributions to the field of urban traffic management.

1. We developed an innovative methodology for identifying conflicts between traffic signals using only topological data from road networks. This approach offers a robust solution applicable to urban networks of any size, overcoming limitations of previous methods that focus on isolated intersections.
2. Our algorithm leverages readily available network infrastructure data, such as GIS datasets from platforms like OpenStreetMap and Google Maps. This eliminates the need for specialized hardware or proprietary devices, making our solution widely accessible and implementable.
3. The research significantly advances traffic management system optimization, particularly in detecting conflicts between traffic signals across entire networks. This comprehensive approach is crucial for improving traffic flow in complex urban environments.
4. Our method can enhance emergency vehicle prioritization strategies through immediate spatial coordination, potentially reducing response times in critical situations.
5. By focusing on the structural relationships within road networks, our approach provides a stable foundation for traffic management strategies that is adaptable to changes in traffic patterns and signal timing adjustments.
6. This work establishes a solid groundwork for future research in traffic network optimization, especially in understanding the intricate relationships between traffic signals in large-scale urban environments.

These contributions not only address critical gaps in current traffic optimization strategies, but also provide practical, actionable solutions for improving coordination and efficiency in urban transportation networks.

Future improvements to our approach could include developing more sophisticated algorithms for crossing point detection, especially in complex urban environments with multi-level intersections. The current implementation relies on basic geometric intersection calculations that work effectively in planar networks but have limitations in three-dimensional contexts. Advanced spatial algorithms incorporating elevation data and non-linear road geometries could enhance detection accuracy. Additionally, machine learning techniques could potentially identify crossing patterns from visual or spatial data,

reducing the reliance on perfect geometric representations. These enhancements would extend the applicability of our conflict matrix approach to increasingly complex urban environments, including those with tunnels, overpasses, and elaborate interchange designs that characterize modern metropolitan areas.

In future works, we plan to leverage the traffic light conflict matrix and the adjacency matrix from our previous research ([12]) to develop a novel emergency vehicle priority approach.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/systems13040290/s1>, Code S1: Dataset_network.m; Code S1: Adjacency_list_of_signal_heads.m; Code S1: Adjacency_list_of_signal_heads_extended.m; Code S1: Adjacency_list_to_matrix.m; Code S1: Conflict_list_of_signal_heads_type.m; Dataset S1: Baton_Rouge_dataset.mat; Dataset S1: Boise_dataset.mat; Dataset S1: London_dataset.mat; Dataset S1: Luxembourg_dataset.mat; Text S1: LICENSE.txt; Text S1: README.txt; Dataset S1: Traffic Light Adjacency Matrix.csv; Dataset S1: Traffic Light Conflict Matrix.csv.

Author Contributions: Conceptualization, methodology, development of algorithms and MATLAB programs, manuscript writing, and data publication management were conducted by corresponding author S.R.-B.; A.C.-N. and M.O.-M. reviewed the conceptualization, bibliographic references, content planning, and scheduling, and also reviewed and validated the drafts until the final document was obtained; S.S.-M. reviewed the mathematical content and contributed to the structuring of the algorithms. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad de Cádiz grant number RYC-2017-22222.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: We would like to thank the entire PTV Group team for their support, especially Marilo Martin-Gasulla, Product Manager Academia and Research.

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

Appendix A. Detailed Examples of Traffic Light Conflict Matrix Generation

This appendix provides detailed step-by-step examples illustrating the methodology for generating the *traffic light conflict matrix* described in Section 4. These examples cover both an isolated intersection and a network-wide scenario, demonstrating the algorithm's application in different contexts. The examples will focus on how road networks and signal heads are represented, how conflict points are identified, how the adjacency matrix is built, how crossing signal heads are used, and how network boundaries are handled. To enhance reproducibility and facilitate further analysis, CSV files containing the complete matrices for these examples are included as supplementary material.

Appendix A.1. Isolated Intersection

Appendix A.1.1. Identifying Real Signal Heads (SH)

Figure 3a illustrates the road network and its corresponding signal heads for an isolated intersection. The signal heads are labeled SH01 through SH08.

Appendix A.1.2. Identifying Conflict Points and Introducing Crossing Signal Heads (CH)

Conflict points are identified based on overlapping movements. Figure A1 illustrates key conflict points (red circles) between specific movements in the intersection. The figure highlights two primary conflicts: the movement controlled by SH01 (magenta lines) intersecting with the movement controlled by SH03 (blue lines), and the movement controlled

by SH02 (green lines) also conflicting with SH03's movement. For clarity, other potential conflicts have been intentionally omitted from this visualization.

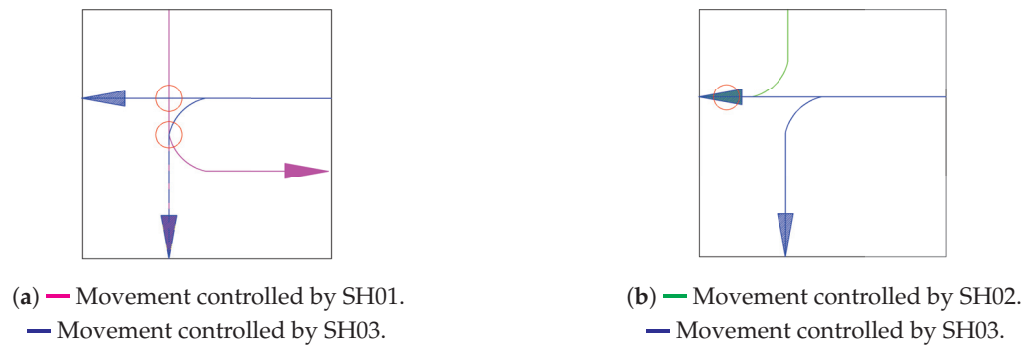


Figure A1. Example of conflict movements.

Appendix A.1.3. Generating the Adjacency Matrix (A)

Based on the detected conflicts, we constructed the adjacency matrix A . To generate the *traffic light conflict matrix*, we began with the intersection's adjacency matrix, incorporating both real signal heads and fictional crossing signal heads. One fictional crossing signal head was added for each pair of intersecting movements. This information serves as input for our algorithm.

In this example intersection, there are eight real signal heads. Using a MATLAB function, we identified the following movement conflicts:

- SH01 conflicts with SH03, SH05, SH06, SH07, and SH08.
- SH02 conflicts with SH03 and SH05.
- SH03 conflicts with SH01, SH02, SH05, SH07, and SH08.
- SH04 conflicts with SH05 and SH07.
- SH05 conflicts with SH01, SH02, SH03, SH04, SH07, and SH08.
- SH06 conflicts with SH01 and SH08.
- SH07 conflicts with SH01, SH03, SH04, and SH05.
- SH08 conflicts with SH01, SH03, SH05, and SH06.

Consequently, 34 conflict points between movements were identified. For each conflict point, we introduced a fictional crossing signal head, denoted as CH01 to CH34. The distribution of these crossing signal heads (CH) among the real signal heads (SH) is as follows:

- CH01 to CH05: Conflicts involving SH01.
- CH06 to CH07: Conflicts involving SH02.
- CH08 to CH12: Conflicts involving SH03.
- CH13 to CH14: Conflicts involving SH04.
- CH15 to CH20: Conflicts involving SH05.
- CH21 to CH22: Conflicts involving SH06.
- CH23 to CH26: Conflicts involving SH07.
- CH27 to CH30: Conflicts involving SH08.

Considering SH01 as an example, we have the following:

- CH01, CH02, CH03, CH04, and CH05 are identified as adjacent CHs.
- For CH01, SH03 is also adjacent.
- For CH02, SH05 is also adjacent.
- For CH03, SH06 is also adjacent.
- For CH04, SH07 is also adjacent.
- For CH05, SH08 is also adjacent.

The complete adjacency matrix A represents the relationships between the real signal heads (SH) and crossing signal heads (CH). The rows of A correspond to the real signal heads, and the columns represent the crossing signal heads. A value of 1 in the matrix indicates the presence of a conflict between a real signal head and a crossing signal head, while 0 indicates no conflict.

Appendix A.1.4. Computing the *Traffic Light Conflict Matrix*

The final *traffic light conflict matrix* C was obtained using the adjacency relationships. Therefore, in the *traffic light conflict matrix*, SH01 will have conflicts (marked by 1) with SH03, SH05, SH06, SH07, and SH08.

The adjacency matrix for this intersection is provided in the Supplementary Materials CSV Files, with the rows and columns corresponding to the signal heads (SH01–SH08) and crossing signal heads (CH01–CH34). The cells indicate the presence (1) or absence (0) of a conflict between the corresponding signal heads.

Appendix A.2. Network-Wide Scenario

Appendix A.2.1. Identifying Real Signal Heads (SH)

Figure 6 illustrates a sample network. Table A1 details the adjacency relationships between the real and endpoint signal heads for the example network.

Table A1. Adjacency relationships.

Signal Head	Adjacent Signal Heads
SH01	SH10, SH13, SH14, SH17, SH18
SH02	EP08
SH03	SH17, SH18, EP08
SH04	EP01
SH05	EP03, SH21, SH22
SH06	SH03, SH04
SH07	SH03, SH04, SH21, SH22
SH08	EP02
SH09	EP01
SH10	SH13, SH14, SH17, SH18
SH11	EP01, EP08
SH12	SH13, SH14
SH14	EP03, SH21, SH22
SH15	EP02, SH03, SH04
SH16	EP03
SH17	EP06, SH29, SH30
SH18	EP07
SH19	EP06, EP07
SH20	SH11, SH12

Table A1. Cont.

Signal Head	Adjacent Signal Heads
SH21	EP04, EP05
SH22	SH19, SH20
SH23	SH19, SH20, EP05
SH24	SH15, SH16
SH25	SH11, SH12
SH26	EP06, SH29, SH30
SH27	EP07, SH11, SH12
SH28	SH29, SH30
SH29	SH15, SH16
SH30	EP04, EP05
SH31	SH15, SH16, SH19, SH20
SH32	EP04

Appendix A.2.2. Identifying Terminal Links and Introducing Endpoint Signal Heads (EP)

To define the network boundaries and ensure comprehensive analysis, the following steps were implemented:

- Identify the edges of the network.
- Introduce fictitious signal heads (EP01, EP02, etc.) at these boundary points to represent the endpoints of the network.
- Connect these fictitious signal heads to the corresponding real signal heads in the adjacency matrix.
- Create a closed system that allows for comprehensive analysis of traffic flow and signal interactions within the defined network boundaries.

Appendix A.2.3. Identifying Conflict Points and Introducing Crossing Signal Heads (CH)

For the network-wide scenario, the complete adjacency matrix A was constructed to represent the relationships between the real signal heads (SH) and both endpoint signal heads (EP) and crossing signal heads (CH). The rows of A correspond to the real signal heads, while the columns represent the endpoint signal heads and the crossing signal heads. The adjacency relationships were determined based on network topology and traffic flow considerations.

The adjacency matrix for this network is provided in the Supplementary Materials CSV Files, with the rows and columns corresponding to the real signal heads (SH01–SH32) and endpoint signal heads (EP01–EP08). The cells indicate the presence (1) or absence (0) of an adjacency relationship between the corresponding signal heads.

References

1. Blincoe, L.; Miller, T.R.; Wang, J.S.; Swedler, D.; Coughlin, T.; Lawrence, B.; Guo, F.; Klauer, S.; Dingus, T. The Economic and Societal Impact of Motor Vehicle Crashes. Report by U.S. Department of Transportation, National Highway Traffic Safety Administration. 2019. Available online: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813403.pdf> (accessed on 1 April 2025).
2. Schrank, D. 2023 *Urban Mobility Report*; Texas A&M Transportation Institute: College Station, TX, USA, 2024.
3. Jima, D.; Sipos, T. The Impact of Road Geometric Formation on Traffic Crash and Its Severity Level. *Sustainability* **2022**, *14*, 8475. [CrossRef]

4. Park, B.B.; Chen, Y. *Quantifying the Benefits of Coordinated Actuated Traffic Signal Systems: A Case Study*; Vtrc 11-Cr2; The National Academies of Sciences, Engineering, and Medicine: Washington, DC, USA, 2010.
5. Haddad, J.; Tur, N.; Veikherman, D.; Buchnik, E.; Ferster, S.; Kalvari, T.; Karliner, D.; Litov, O.; Zagoury, A.; Rottenstreich, O.; et al. Quantitative Approach for Coordination, at Scale, of Signalized Intersection Pairs. In Proceedings of the 2024 European Control Conference (ECC), Stockholm, Sweden, 25–28 June 2024; pp. 1881–1888. [CrossRef]
6. Xing, J.; Zhang, Q.; Cheng, Q.; Zu, Z. A Geographical and Temporal Risk Evaluation Method for Red-Light Violations by Pedestrians at Signalized Intersections: Analysis and Results of Suzhou, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 14420. [CrossRef] [PubMed]
7. Jin, Q. Automatic Control of Traffic Lights at Multiple Intersections Based on Artificial Intelligence and ABST Light. *IEEE Access* **2024**, *12*, 103004–103017. [CrossRef]
8. Chen, W.; Yang, S.; Li, W.; Hu, Y.; Liu, X.; Gao, Y. Learning Multi-Intersection Traffic Signal Control via Coevolutionary Multi-Agent Reinforcement Learning. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 15947–15963. [CrossRef]
9. Wang, Y.; Xu, T.; Niu, X.; Tan, C.; Chen, E.; Xiong, H. STMARL: A Spatio-Temporal Multi-Agent Reinforcement Learning Approach for Cooperative Traffic Light Control. *IEEE Trans. Mob. Comput.* **2022**, *21*, 2228–2242. [CrossRef]
10. Ding, L.; Zhao, D.; Zhu, B.; Wang, Z.; Tan, C.; Tong, J.; Ma, H. SpeedAdv: Enabling Green Light Optimized Speed Advisory for Diverse Traffic Lights. *IEEE Trans. Mob. Comput.* **2024**, *23*, 6258–6271. [CrossRef]
11. Sheriff, F. ELMOPP: An application of graph theory and machine learning to traffic light coordination. *Appl. Comput. Informatics* **2024**, *20*, 217–230. [CrossRef]
12. Rojas-Blanco, S.; Cerezo-Narváez, A.; Otero-Mateo, M.; Sáez-Martínez, S. Adjacency List Algorithm for Traffic Light Control Systems in Urban Networks. *Systems* **2024**, *12*, 539. [CrossRef]
13. Qadri, S.S.S.M.; Gökçe, M.A.; Öner, E. State-of-art review of traffic signal control methods: Challenges and opportunities. *Eur. Transp. Res. Rev.* **2020**, *12*, 1–23.
14. Shaygan, M.; Meese, C.; Li, W.; Zhao, X.G.; Nejad, M. Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transp. Res. Part C Emerg. Technol.* **2022**, *145*, 103921. [CrossRef]
15. Cui, H.; Meng, Q.; Teng, T.H.; Yang, X. Spatiotemporal correlation modelling for machine learning-based traffic state predictions: State-of-the-art and beyond. *Transp. Rev.* **2023**, *43*, 780–804. [CrossRef]
16. Jiang, W.; Luo, J. Graph neural network for traffic forecasting: A survey. *Expert Syst. Appl.* **2022**, *207*, 117921. [CrossRef]
17. Zhu, R.; Li, L.; Wu, S.; Lv, P.; Li, Y.; Xu, M. Multi-agent broad reinforcement learning for intelligent traffic light control. *Inf. Sci.* **2023**, *619*, 509–525. [CrossRef]
18. Chalaki, B.; Malikopoulos, A.A. Optimal Control of Connected and Automated Vehicles at Multiple Adjacent Intersections. *IEEE Trans. Control Syst. Technol.* **2022**, *30*, 972–984. [CrossRef]
19. Li, W.; Wang, B.; Khattak, Z.H.; Deng, X. Network-Level Traffic Signal Cooperation: A Higher-Order Conflict Graph Approach. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 990–999. [CrossRef]
20. Fellendorf, M.; Vortisch, P. Microscopic Traffic Flow Simulator VISSIM. In *Fundamentals of Traffic Simulation*; Springer: New York, NY, USA, 2010; pp. 63–93. [CrossRef]
21. Ahmed, S.; Osman, O.; Codjoe, J. Mitigating Traffic Congestion on I-10 Mississippi River Bridge in Baton Rouge, LA. *MATEC Web Conf.* **2019**, *271*, 06002. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Hybrid Model for Motorway EV Fast-Charging Demand Analysis Based on Traffic Volume

Bojan Rupnik ^{1,*}, Yuhong Wang ² and Tomaž Kramberger ¹¹ Faculty of Logistics, University of Maribor, 3000 Celje, Slovenia; tomaz.kramberger@um.si² Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China; wangyuhong@nbu.edu.cn

* Correspondence: bojan.rupnik@um.si

Abstract: The expected growth of electric vehicle (EV) usage will not only increase the energy demand but also bring the requirement to provide the necessary electrical infrastructure to handle the load. While charging infrastructure is becoming increasingly present in urban areas, special attention is required for transit traffic, not just for passengers but also for freight transport. Differences in the nature of battery charging compared to that of classical refueling require careful planning in order to provide a resilient electrical infrastructure that will supply enough energy at critical locations during peak hours. This paper presents a hybrid simulation model for analyzing fast-charging demand based on traffic flow, projected EV adoption, battery characteristics, and environmental conditions. The model integrates a probabilistic model for evaluating the charging requirements based on traffic flows with a discrete-event simulation (DES) framework to analyze charger utilization, waiting queues, and energy demand. The presented case of traffic flow on Slovenian motorways explored the expected power demands at various seasonal traffic intensities. The findings provide valuable insight for planning the charging infrastructure, the electrical grid, and also the layout by anticipating the number of vehicles seeking charging services. The modular design of the model allowed replacing key parameters with different traffic projections, supporting a robust scenario analysis and adaptive infrastructure planning. Replacing the parameters with real-time data opens the path for integration into a digital twin framework of individual EV charging hubs, providing the basis for development of an EV charging hub network digital twin.

Keywords: electric vehicles; EV charging; traffic flow; EV charging hub

1. Introduction

The increasing number of electric vehicles (EVs) is already exposing issues caused by traffic structure changes, which include shifts in local and long-haul freight transport, seasonal passenger travel patterns, as well as gradual adoption of alternatively fueled vehicles. Given this development, the lack of charging infrastructure presents an obstacle for EV adoption and also requires a different understanding of charging behavior as opposed to refueling. Various technological and organizational solutions may mitigate these issues to a degree; however, the still-developing infrastructure can present a serious issue to future traffic flows. Furthermore, the lack of available infrastructure may hinder the EV adoption rate. With most research focused on urban planning and pricing policies to accommodate the increasing EV requirements, the field of infrastructure planning along transit routes also requires a deeper insight.

Selecting proper locations for the charging infrastructure can depend on various factors, such as accessible geography, land property rights, flooding areas or other hazards,

environmental protection zones, etc. For example, while available land may initially appear suitable, its utility may be limited if grid access is lacking or if it lies within an environmental protection zone. Similarly, locations near existing substations may reduce infrastructure costs but may not align with traffic flow needs. The requirement for charging hubs to be available along motorways at every 60 km for cars and every 120 km for trucks and buses as per EU regulations should also be taken into consideration [1].

While the current EV adoption may be relatively low, the expected development requires careful planning of future requirements [2]. Setting up electrical and other infrastructure can require substantial investments, which is why providing sufficient capacities for future demand can prevent the need for costly adaptations later. Motorway charging in particular requires understanding different charging behavior as opposed to urban charging. Highly directional and seasonally variable traffic patterns on motorways do not allow the same approach as in urban planning. As the aforementioned traffic structure on motorways includes a significant number of commercial vehicles, infrastructure planning needs to anticipate their distinct energy needs and schedules in order to ensure smooth traffic flows.

This paper addresses the issue of motorway charging by proposing a hybrid simulation model for planning and analyzing EV fast-charging on motorways. The hybrid model combines a decision model to predict the charging demand based on traffic flow and a discrete-event model to simulate the charging operations. While the main parameter was traffic volume, which was estimated as a forecast based on historical traffic development, the model also incorporated battery and charger characteristics, temperature effects on battery efficiency, and estimated EV adoption rate. In this research, the presented model was used to analyze the operations of a charging hub in order to anticipate its power demands in peak hours, explore utilization rates of the charging hub under different traffic loads, and predict and or prevent bottlenecks that influence local traffic flow and user satisfaction.

The model was used with data from a traffic counter on Slovenian motorways [3], which, due to its geographical location, provided a mixed traffic structure of predominantly not local but international commercial vehicles hauling freight [4]. On the other hand, cars presented mostly domestic traffic, with the biggest volume being work trips between the largest Slovenian cities, with the exception of the tourist season and certain holidays, when foreign (as well as domestic) car volume drastically increases on the motorways. The main premise here is that transit traffic will rely on motorway AC fast charging capacities [5], and alternatives will be available to domestic users. Although pricing policies and scheduling options can affect the charging service selection and can also influence the traffic flow [6], the bulk of international transport will have no other option than to charge at motorway services.

While some individual charging stations are available along the Slovenian motorways, the expected shift to EV in the coming years requires adaptation to greater demand. While sparsely positioned chargers may distribute the demand over a larger area, providing access through charging hubs capable of serving a multitude of vehicles also addresses the problem of providing electricity to distant locations, as the location of a charging hub also requires the construction of a suitable substation. The results of the presented model will provide the basis for planning the required substation output power levels, charging hub configuration, and layout requirements.

2. Related Work

2.1. EV Charging Behavior and User Preferences

Understanding EV user behavior is essential for estimating demand profiles and infrastructure performance. In [7], the state of charge (SoC) and energy storage distribution are modeled based on real-world data from Denmark, providing insight into battery

utilization and the potential for vehicle-to-grid (V2G) services. Similarly, ref. [8] analyzes plug-in EV charging behavior and finds that over half of users charge exclusively at home, with only 5% relying on public charging—primarily to reduce range anxiety. Charging preferences are also explored in [9], where the authors identify influencing factors across transport, vehicle, and power system domains. The study notes that most users prefer overnight charging and avoid deep discharges, contributing to evening peak loads and battery preservation behaviors.

Workplace and flexible charging patterns are examined in [10], which presents an algorithm for optimizing charger flexibility and peak demand reduction. The study considers charger types, SoC variability, and historical consumption, and while focused on workplace environments, its insights are transferable to motorway settings, particularly for managing high-traffic periods.

A comprehensive review of factors influencing electric vehicle (EV) range prediction, categorizing them into vehicle design, driver behavior, and environmental conditions is given in [11]. They highlight the complexity of accurately forecasting EV range due to the interplay of these diverse factors. While their work offers a broad overview, it underscores the necessity for models that integrate real-time traffic dynamics, charging behavior, and infrastructure constraints, particularly in the context of long-distance highway travel.

In [12], a novel adaptive stated choice experiment to investigate range anxiety and charging decisions during long-distance EV travel is presented. This study captures how drivers dynamically weigh the SoC, detours, waiting times, and charger availability over the course of a journey. A mixed logit model reveals that the decision to charge is influenced not only by the remaining range but also by sociodemographic factors, including age and vehicle brand. The authors observe that some segments, such as Tesla drivers and younger users, are less susceptible to range anxiety.

2.2. Commercial EV Charging and Freight Transport

Commercial EVs, especially heavy-duty trucks, introduce distinct charging demands and infrastructure requirements. In [13], the authors compare on-route and return-to-base charging strategies for medium- and heavy-duty trucks, highlighting challenges such as regulatory constraints and charger location planning. This work also addresses commercial V2G integration and the specific design needs of freight vehicles.

Long-haul electric trucking is explored in [5], comparing two fast-charging network scenarios: one with 100 km spacing and another with 50 km. Using 2030 projections and a 15% truck electrification rate, the study shows that denser networks reduce wait times but require significantly higher investment.

The regulatory framework surrounding truck charging infrastructure is discussed in [14], which outlines EU targets for fast-charging stations on the TEN-T network. The paper emphasizes gaps in public charging for freight, particularly in countries like Poland, and argues for a structured, law-based methodology to guide charger deployment at megawatt-scale power levels.

Several studies address the spatial distribution and sizing of EV charging stations. In [2], the authors identify infrastructure gaps along EU highways, particularly the TEN-T network, where fast chargers are concentrated in Western Europe. The EU's plan for 3 million chargers by 2030, including requirements for high-power hubs at fuel stations, underscores the urgency of strategic planning—an issue also tackled by our model.

Traffic-based demand modeling is used in [15], where different timeframes and charger configurations are evaluated. The study highlights the importance of range anxiety as a design constraint and promotes adaptive charger placement along major corridors. In [16], a

multi-criteria methodology avoids co-locating chargers at petrol stations, favoring locations near population centers, logistics hubs, and high-traffic routes.

Urban optimization strategies also contribute useful insights. Reference [17] applies particle swarm optimization to balance station loads and reduce bottlenecks caused by an uneven distribution. Although focused on urban settings, such approaches inform macro-level infrastructure layout for regional networks.

2.3. Simulation and Load Modeling Approaches

Simulation methods are commonly used to estimate power demand and charging dynamics. In [18], a microscopic traffic simulation shows that high-demand charging stations can create local congestion and trigger detours, increasing pressure on nearby routes. Similarly, ref. [19] compares random and coordinated charging strategies using Monte Carlo methods, concluding that coordinated charging shifts loads away from peak hours and improves grid efficiency.

A discrete event simulation (DES) model with survey data to simulate EV charging behavior under varying conditions of charger availability and user preferences is presented in [20]. The findings support strategies like charging frequency reduction and promoting home charging to spread demand more evenly. A broad review of infrastructure modeling strategies is provided in [21], which classifies node-based, path-based, and tour-based methods. The authors note that hybrid approaches are often more effective and stress the need for incrementally scalable, strategically located infrastructure along transit corridors.

An agent-based simulation framework to estimate EV charging loads based on aggregated user behavior, spatial constraints, and temporal usage patterns is presented in [22]. Their model incorporates driver preferences, SoC thresholds, charger availability, and probabilistic arrival times to simulate realistic, urban charging dynamics. While their focus is on urban environments and behavior-driven modeling, the study shares common ground with our approach in simulating charging demand using system-level variables.

Another agent-based simulation is proposed in [23] with a high-resolution agent-based simulation (ABS) framework to evaluate the layout of electric vehicle charging infrastructure (EVCI) in urban environments. Their model integrates GPS trajectory data of electric taxis in Shenzhen with a geographic information system (GIS) road network, enabling dynamic routing and real-time decision making by agents based on queuing conditions and traffic states. Key contributions include direct behavior extraction from GPS data, rule-based charger selection, and a multi-objective Pareto optimization to balance station utilization and driver satisfaction.

A recent contribution is provided by [24], where the authors propose a hybrid method combining p-median location optimization, a genetic algorithm, and discrete-event simulation to identify optimal locations and determine capacities for EV charging stations. Their work emphasizes demand-driven infrastructure development, incorporating market penetration forecasts, station queuing performance, and charging module type differentiation (fast vs. normal).

A comprehensive review of EV charging infrastructure planning methodologies is presented in [25], which examines integrated approaches that consider both transportation and power systems. Their work classifies existing models based on spatial, temporal, and technological scope, and highlights the growing importance of **co-optimization** techniques that simultaneously address vehicle routing, grid constraints, and charging station operations. They emphasize that realistic planning requires not only demand estimation but also grid-aware simulation and iterative design methods.

3. Materials and Methods

With the currently underdeveloped infrastructure for future requirements, this paper does not focus on exploring the EV charging network along TEN-T but instead focuses on finding the requirements for setting up an EV charging hub capable of serving a multitude of vehicles at any location along the motorway which fulfil basic conditions of having suitable geographical features, resolved land ownership and also proximity to the existing electrical grid. Without an existing developed grid, this first location can serve as a pivot point for setting up surrounding charging hubs along the motorway and plan the required infrastructure accordingly.

As planning infrastructure is a long-term investment, it is necessary to consider future development of multiple factors with the focus on:

1. Future traffic volume;
2. EV adoption into the traffic structure;
3. Technological advances.

Future traffic volumes can be estimated based on historical data. There are multiple factors to consider, such as future economic development, policy changes, global crises such as pandemics, or even new technological advances that could thwart accurate prediction. Still, historical data can provide a reliable way of predicting the trend of future traffic.

EV adoption rate can be estimated in a similar way. With the expected turn to non-fossil engines in the EU until 2035 [1], alternative solutions are not only EVs, but also synthetic fuel and hydrogen, which may impact the structure of future fleets. It is also important to consider that these vehicles will only gradually replace existing internal combustion engine (ICE) vehicles.

Both EV battery efficiency as well as charging times have improved over time, and it can be assumed that this trend will continue. That means that the current charging patterns will probably change in the future. Whether charging times will come near those of classical car refueling or not, it will be necessary to provide the necessary power for charging EV masses or instead expect queues and delays.

The method presented in this paper focuses on a single charging hub consisting of multiple DC charging points for both cars and trucks. The traffic flow structure at any given time presents an input for a decision model, which then relays the EV load to a discrete event simulation (DES) model of the charging hub, resulting in a calculation of the power consumption at that time (Figure 1).

The inputs of both the decision model as well as the latter charging simulation are estimates that can be adjusted for exploring the requirements under different EV development scenarios. The parameters used in this method are listed in Table 1.

Table 1. Parameters used for modelling the traffic flow and EV charging.

Parameter	Description	Unit
t	Simulation clock	h
λ_t	Traffic flow at time t	Vehicles per hour
V	List of vehicles	Set
V_i	i_{th} vehicle in sequence	Indexed vehicle
EV_i	Engine type of V_i (EV or not)	Binary
BC_i	Battery charge of V_i	kWh
BC_{imax}	Battery capacity	kWh
SoC_i	Battery state of charge of V_i	[0, 1]
TV_i	Vehicle type (car, light/medium/heavy truck)	Set
r	Ratio of EV for each vehicle type	[0, 1]

Table 1. Cont.

Parameter	Description	Unit
L_i	Locality index	Binary
J	Charger count at charging hub	Integer
C_j	j_{th} charger at charging hub $j = 1, 2, \dots, J$	Integer
CP_j	Maximal charging output of C_j	kW
PB_i	Maximal charging power of V_i batter	kW
$Temp(t)$	Estimated temperature at t	Celsius
RC_i	Capacity-to-range ratio	kWh/km
$p(t)$	Time-based charging probability	$[0, 1]$

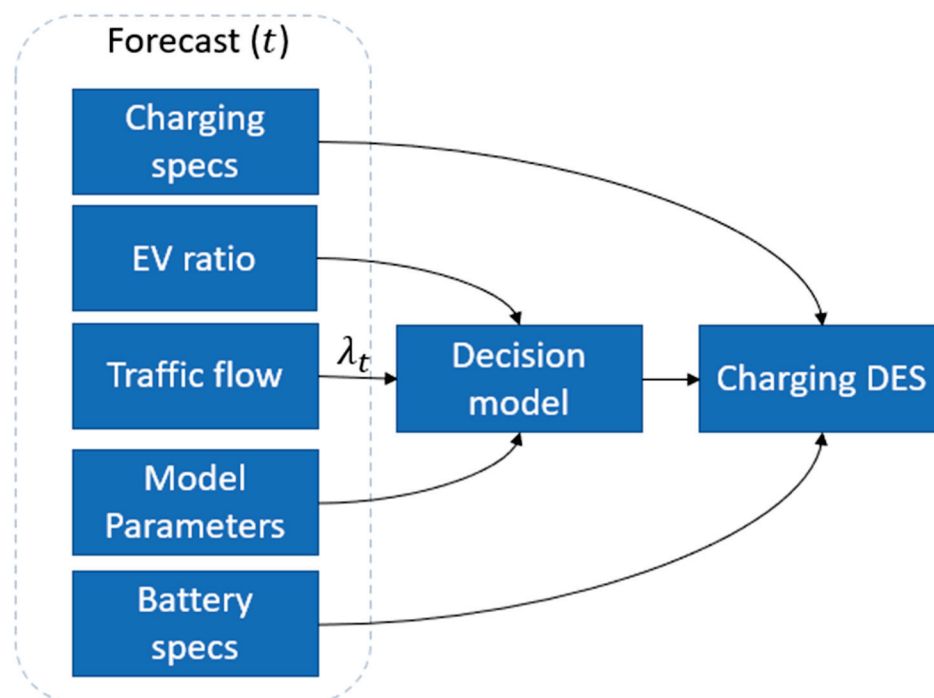


Figure 1. Schematic for the power consumption analysis.

3.1. Traffic Flow

The simulation clock t advances based on the occurrence of discrete events and does not progress in fixed time steps. The timestamp of each event—such as vehicle arrival or charger availability—is either precomputed at the start of the simulation or dynamically generated in response to prior events. Although the input data provides traffic volumes at hourly intervals, the simulation clock is not limited to hourly steps, but can take on any fractional value within that interval. This allows the analysis to be conducted at a high level of precision, thus accurately capturing event dynamics and enabling precise tracking of system state changes over time.

With λ_t representing the number of vehicles passing a traffic count within 1 h at t , this value was used to model the exact arrival of each individual vehicle using the Poisson process [26], which can be applied for modelling traffic flow distributions. While, according to [27], highway traffic does not necessarily follow the Poisson process, it can be applied in cases where vehicle arrival does not appear in bursts. As the available data was sampled at an hourly rate, which was not precise enough to reveal such bursts, we assumed that the Poisson process was fitting to describe the given arrival rate. Equation (1) was used

to generate exponentially distributed inter-arrival times on the assumption of a Poisson arrival process, where $\lambda(t)$ presents the average arrival intensity at a given time.

$$t_{V_i} = t_{V_{i-1}} - \frac{\ln(U(0, 1))}{\lambda(t_h)} \quad (1)$$

where t_{V_i} presents the arrival time of each vehicle from V passing the charging hub, $\lambda(t_h)$ is the target hourly vehicle count, and $U(0, 1)$ is a randomly generated value in $[0, 1]$, uniformly distributed. As $\lambda(t_h)$ was updated when the simulation clock passed an hour, we followed a non-homogenous Poisson process, which adapted to the available data.

3.2. EV Charging Decision Model

In order to estimate the number of vehicles that are likely to require charging, the fleet composition needed to be provided with further detail to augment each V_i with additional information. Along with vehicle parameters, it is also important to consider weather conditions that can influence the range of EV batteries and traffic patterns that depend on time such as the difference between workdays and weekdays or commute hours. Each of the parameters for the decision model (Table 1) was set stochastically based on forecasted estimates at time t .

Depending on the fleet composition, the vehicles passing the charging hub can be either cars, light trucks, medium or heavy trucks, or buses. The initial parameter defines the vehicle type-based ratio between each type forecasted at time t (Equation (2)):

$$types(t) = [0, r_1, \dots, r_5] \quad (2)$$

$$TV_i = I + 1, \text{ where } r_I \leq U(0, 1) < r_{I+1} \quad (3)$$

With the selection of the vehicle type, the next step is to set the engine type of V_i , which is either EV or not (Equation (4)). This requires a forecast of the number of EVs that also depends on vehicle types, as there is a higher adoption rate of cars compared to that of trucks, for example.

$$EV_i = \begin{cases} 1, & \text{if } U(0, 1) < EV_ratio(t, TV_i) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $EV_ratio(t, TV_i)$ presents the ratio of EVs of the whole traffic volume for each vehicle type. This part presents the first selection of potential candidates for charging, as only EVs are eligible for stopping at the charging hub, eliminating any others. For further analysis, the range of the vehicles was calculated based on battery capacity, which was retrieved from the forecast of battery development.

$$BC_i = U(C_{min}(TV_i), C_{max}(TV_i)) \quad (5)$$

where C_{min} is the minimal expected battery capacity for a specific vehicle type at time t , C_{max} is the upper bound, and rng is a function generating a random value between the lower and upper bound. As vehicles of different manufactures have different battery properties, their capacities should adhere to that as well.

Assessing the SoC requires information about traffic flows that cannot be retrieved solely from traffic count data. Further research is planned in order to augment the data with origin–destination analysis in order to provide a more accurate representation of

charging requirements across the charging grid. For the purpose of this research, the SoC was modelled using as a Beta distribution (Equation (6)).

$$\begin{aligned} X &\sim \text{Beta}(\alpha, \beta) \\ \text{SoC} &= 0.1 + (-0.95 - 0.1)X \\ \alpha &= 2, \beta = 5 \end{aligned} \quad (6)$$

the Beta distribution was set assuming that vehicles needed to travel a certain distance in order to reach the charging hub, irrelevant on the origin, therefore, a SoC above 0.9 was unlikely. As the projected charging hub was located along the motorway, cars originating from nearby motorway entrances near populated areas was a probable source of such SoCs, otherwise, the highest expected SoC should be from cars that have been charging at previous charging hubs along the route that are expected to be 60 km apart.

The battery capacity and SoC were used to estimate the range of the vehicle, which also depended on the vehicle type and temperature. The battery efficiency based on temperature used in this model is displayed in Figure 2, which depicts an approximation based on [28].

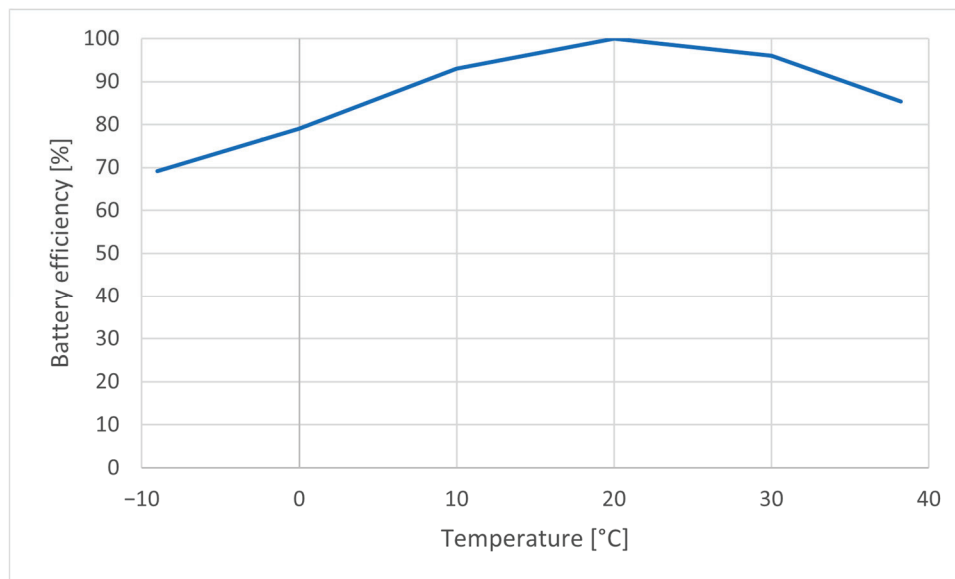


Figure 2. Battery efficiency depending on temperature.

The remaining range of each V_i was then calculated using Equation (7), depending on the remaining charge, RC_i is presented as distance/charge ratio for each individual vehicle type, and battery efficiency at the temperature forecast at t .

$$\begin{aligned} \text{charge} &= C_i * \text{SoC}_i \\ \text{range}_i &= \text{charge} * \text{efficiency}(\text{Temp}(t)) * RC_i \end{aligned} \quad (7)$$

Calculating the range provided the base for the first decision option. Any vehicle with a range lower than that required to reach the next charging option would be redirected to the current charging hub; however, there are also other traffic patterns that need to be included in this consideration.

Local and transit traffic differ in their charging behavior, with local traffic relying on local charging options at their destination. International transport is that which will predominately be using the motorway charging hubs. The charging decision for local

traffic was thus further reduced, while no such penalties were applied to transit traffic. The locality flag for the vehicles was set as follows (Equation (8)):

$$L_i = \begin{cases} 1, & \text{if } U(0, 1) < local(t) \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $local(t)$ presents the ratio of local to transit traffic. Local traffic will not seek service at the charging hub unless the battery charge is critically low.

The final parameters influencing the charging decision were time of day and time of year. Cars and trucks alike are unlikely to engage in charging during morning hours, as this time usually represents morning commute where cars have probably been already charged at home. Similarly, transit traffic trucks will have been charged at truck stops on slow DC chargers during the night. It can also be assumed that trucks appearing during the night will seek slow DC charging options that have smaller power outputs. Furthermore, truck drivers are obligated to take 45 min breaks during their travel after 4.5 h, during which they may choose to use the breaks for AC charging.

A probability weight w was defined for each vehicle type that reduces the number of vehicles requiring charging at improbable hours (Figure 3).

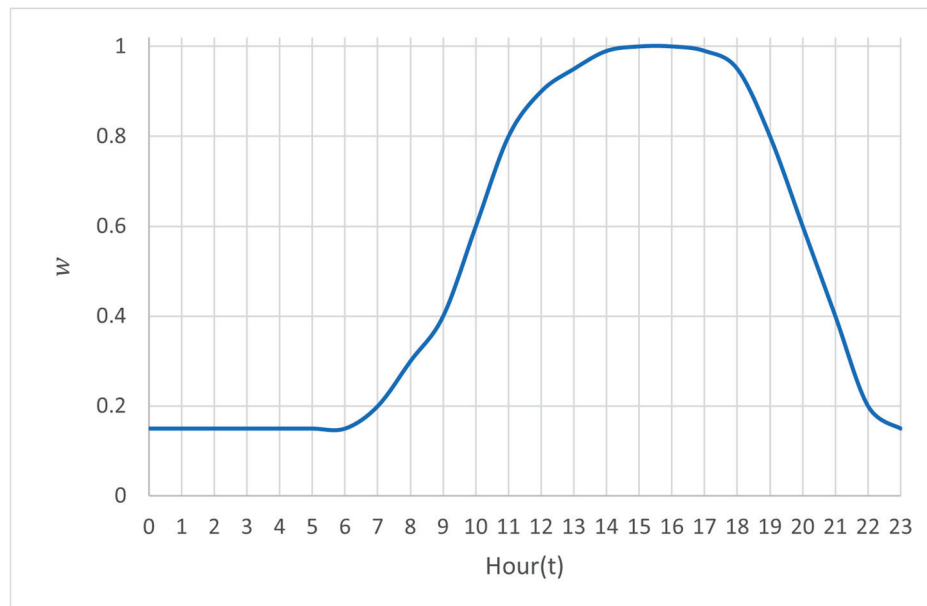


Figure 3. Charging probability weight.

While hauling presents a significant portion of traffic throughout the year, it is also necessary to give attention to passenger transport, where tourist seasons see a drastic increase in both domestic as well as foreign cars. The latter will especially be prone to seeking charging services along the motorways, thus representing an atypical power demand for cars compared to that in the rest of the year.

The decision model calculated the charging probability for each individual vehicle from the inbound vehicle stream. As the simulation was intended for the initial charging hub, it was implicit that the next charging hub along the motorway will be no further than 60 km away. The calculated range based on the SoC and battery capacity allowed calculating the requirement for each vehicle to charge at the current center or continue their travel. If the vehicle's estimated range falls below 90 km—comprising 60 km of remaining distance and a 30 km safety buffer to account for range anxiety—or if the state of charge (SoC) is expected to drop below 0.2 before reaching the next charging hub, the charging

probability p_c is set at 0.95, leaving a small probability for exceptions. A current SoC below 0.2 forced the car to charge regardless of other parameters, assuming that no other charging options were reachable. The locality index reduced the charging probability further by a factor of 4, which was a conjectured value based on the assumption that local vehicles would avoid charging on motorways. The charging probability was then multiplied by the time-of-day weight w (see pseudocode in Table 2).

Table 2. Pseudocode for the decision model.

Charging Decision Function
<pre> function ChargingDecision (V_i, SoC, range, w, t) $p = 0$ if ($EV_i = 0$) return FALSE else if (range < 90 km or SoC at next center < 0.2) $p = 0.95$ else $p = 0.05$ if ($L_i = 1$) $p = p/4$ if (SoC < 0.2) return TRUE $p = w(t) * p$ if (rng() < p) return TRUE else return FALSE end </pre>

3.3. Charging Hub DES Model

The charging hub model consisted of an inbound source of EVs, a waiting queue, and two separate arrays of servers (chargers), one for cars and light trucks, and the other for medium and heavy trucks and buses (Figure 4). The charging process itself requires information about the current each vehicle's battery capacity, SoC, and maximal charging power.

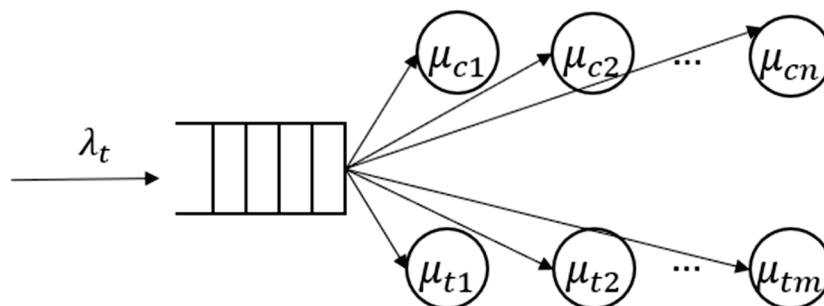


Figure 4. Charging DES queuing model for a charging center.

The queue presents a parking space or waiting line for the chargers for service using a FIFO strategy. As chargers provide different charging output powers, the queue is responsible for relaying the vehicles to the most appropriate charger that fits the maximal allowed charging power of the EVs battery. In order to find the maximum expected power load of the charging hub, the number of chargers was unlimited for the initial simulation run. However, in order to find a feasible configuration of the charging hub, the number of chargers should adapt to arrival intensity. For this purpose, the queue also needed limited capacity and maximal waiting times of vehicles. The EVs were waiting in the queue until a

suitable charger was available. Vehicles would not enter the queue if its capacity was full or the average waiting time exceeded a tolerance value.

The service rate of the chargers was calculated based on the vehicles' maximal battery charging power BC_{imax} , current charge BC_i (which equals $BC_{imax} * SoC_i$) and the charger's output power CP_j , and the randomly generated target charge $TC_i \in [0.8, 1.0]$, as some batteries allow for charging to full capacity, while others' recommended capacity is at 80%. As battery charging is not linear, the times were calculated separately depending on the initial SoC_i for each of the charging phases where the charging power adapted to randomly selected values from ranges in Table 3. The values were loosely estimated based on [29,30], with most charging curves behaving in a similar manner.

$$\mu_j = (TC_i * BC_{imax} - BC_i) / \min(E(SoC_i) * BP_i, CP_j) \quad (9)$$

where E is the charging efficiency at the given SoC and started from its initial value to the SoC ranges in Table 3. The total charging time was calculated by summing all partial times. The charging power adapted to either to the maximal battery power or the charger output power. Table 4 shows the pseudocode for the presented charging simulation.

Table 3. Charging power utilization depending on the SoC.

SoC	Charging Efficiency
<0.2	50–80%
0.2–0.6	90–100%
0.6–0.8	60–80%
0.8–1.0	10–30%

Table 4. Pseudocode for the charging simulation.

Charging DES
Procedure Charging ($V_i, SoC, range, t, queue$) For each V_i in V if (ChargingDecision($V_i, SoC, range, t$) == TRUE) then if (available(queue)) then assign($V_i, queue$) for each C_j in C if (available(C_j)) then release($V_i, queue$) assign(V_i, C_j) <tr(<math>C_j) = release time of C_j by Equation (9) for each C_j in C if tr(C_j) >= t then release(C_j) clock_increment(t) end </tr(<math>

3.4. Input Data and Parameters

The focus of this research was on the rapid DC charging simulation, omitting the slow charging requirements that were assumed to take place outside the motorway grid. Planning infrastructure requires anticipating the EV charging demands of future development, and the projected year provides a turning point on EV adoption for all newly sold vehicles. To demonstrate the charging DES, 2040 was the selected forecasted period, when a significant portion of EVs of all types can be expected.

The traffic data was retrieved from the traffic count records provided by the Slovenian Motorway Company [3]. These provided details about vehicle type and count passing at monitored motorway sections at an hourly rate. The forecast for future traffic flows, EV

ratio and battery specifications were done using a rudimentary approximation of future traffic flows, battery and charging specifications, temperature, and EV ratio. The forecast was done based available traffic counter data from 2021–2022, excluding previous periods due to the influence of the pandemic. To provide a sample forecast for 2040, the trend in the last 4 years was estimated and projected onto the available data from 2022.

Figure 5 displays traffic flow changes throughout the year, where (a) shows a peak in the tourist season for passenger cars and a relatively low flow of trucks, while (b) provides a detailed flow of vehicles during a busy workday. The ratio of EVs (c) provides the expected number of EVs as candidates for charging from (a), while the projected battery efficiency (d) was assumed based on the hourly temperature forecast for 2040.

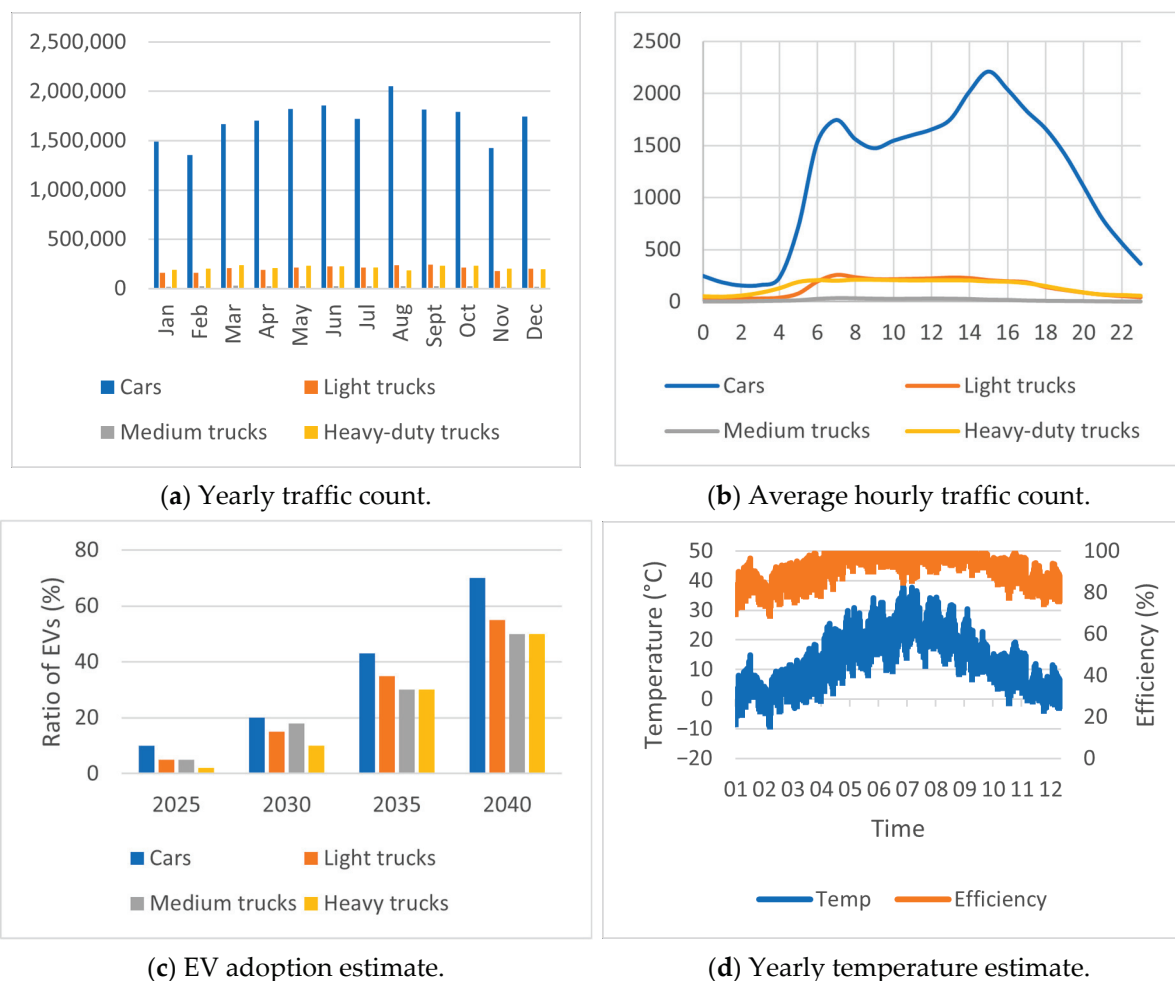


Figure 5. Traffic structure input data for the simulation model consisting of traffic flow throughout the year (a), an hourly detail on a workday (b), ratio of EVs in the total fleet (c), and battery efficiency at the temperature in (d).

These inputs present reference values, which can be adapted to explore any scenario of traffic flow, EV adoption (adapted based on [31]), climate, or technological development. As such, the simulation could observe the influence of any parameter on the power demand, charging times, and waiting queues. While we have performed multiple simulations, we have selected two sample scenarios with different charging hub configurations presented in Table 5 in order to demonstrate the application of the method.

Table 5. Charging hub configuration for testing scenarios.

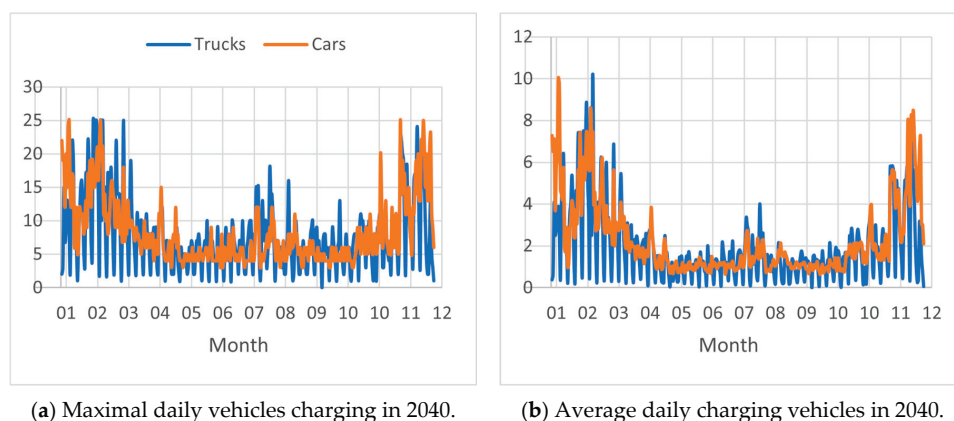
Scenario	Vehicle Type	Number of DC Chargers	Charging Power [kW]
S1	Cars and Light trucks	25	100–350
	Medium- and heavy-duty trucks	25	350–1000
S2	Cars and Light trucks	15	100–350
	Medium- and heavy-duty trucks	15	350–1000

4. Results

The simulation results are provided by sampling the waiting queue, individual charger utilization, and output at a 1 min interval. The discrete-event simulation provided exact times of any event that included arrival at the queue, start of charging, and end of charging. As such, the number of events depended mostly on the number of available chargers, when arrival intensity neared the charging hub service rate. The 1 min sampling interval was short enough to cover most state changes and provide the main performance indicators. A shorter interval would not have provided any significant precision improvements but would have required more computation.

The simulation was performed for the duration of one year, to observe seasonal trends and overall infrastructure demands. Along with the year-round overview, the presented results also displayed more detailed charging operations during a busy winter week. The simulation results for a weekly overview of the charging hub utilization are displayed in Figure 6. The presented scenarios displayed the behavior of the charging hub with a configuration of sufficient chargers to avoid any queues (S1) and a scenario that allowed vehicle build-ups during peak hours (S2).

The yearly overview revealed the influence of temperature on battery efficiency and charging, where the frequency of charging was drastically increased in the cold months compared to that during warm ones. The effect was even visible during summer months where a drastic increase in car traffic had a much lesser effect on charging demand than the reduced battery range during the winter period. In scenario S1, there were only occasional minimal queues and waiting times, so fully available service at practically any time was provided for the projected traffic volume. Occasionally, either cars or trucks filled the capacity of the hub, but at this configuration, they never fully occupied every charger simultaneously. Total power output exceeded 20 MW, where the majority was required for charging commercial vehicles. Substation planning should consider this as a reference value for the necessary output. Lesser output capabilities would reduce charger efficiency, lengthen the charging times, and induce waiting queues for charging services during the peak hours and season.

**Figure 6.** Cont.

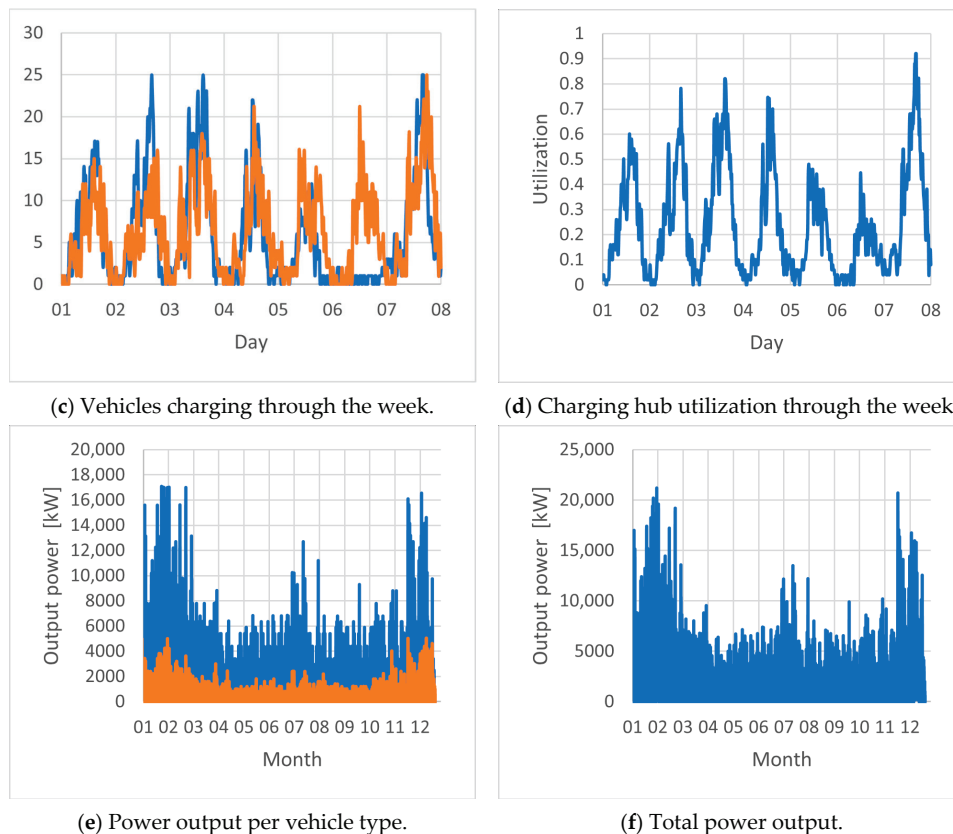


Figure 6. Results of S1 analysis for the charging hub operations with the maximal (a) and average (b) daily charging vehicles, a detailed overview of charging during the week (c) and overall utilization of the charging hub (d), and used charging power per vehicle type (e) and total output (f).

S1 presents a smoothly running scenario with negligible influence on traffic and waiting times; however, the utilization of the charging hub was relatively low in the off-season. Given that most of the time the hub was not fully utilized, the number of chargers could be reduced; however, these warrants exploring the effects that a lower number of chargers has on traffic. Scenario S2 (Figure 7) simulates the same traffic volume as that in S1, with only 30 available charging stations.

The reduced number of chargers in S2 naturally reduced the maximal output of the charging hub. Without the necessary capacities, waiting queues for the chargers appeared for as much as 20 vehicles, where trucks dominated. This value provides insight into designing the parking area from which the charging area is accessed. The maximum waiting times for this configuration got near 1 h, while the same traffic volume produced no queues in S1. While it may appear that this time should be higher because of queue size, the high number of chargers ensured a high frequency of releasing them for the next customers. The substation output requirements for this scenario were lower, and the overall utilization of the hub reached 100% on several occasions. While it is possible to reduce the infrastructure costs with a less powerful substation and fewer chargers with higher utilization, it is necessary to consider the effects on traffic delays as well as customer satisfaction. While providing additional charging stations does not require difficult modifications, upgrading a substation to higher output levels, however, might present a more substantial investment, which should be avoided by planning with a buffer of expected power demands.

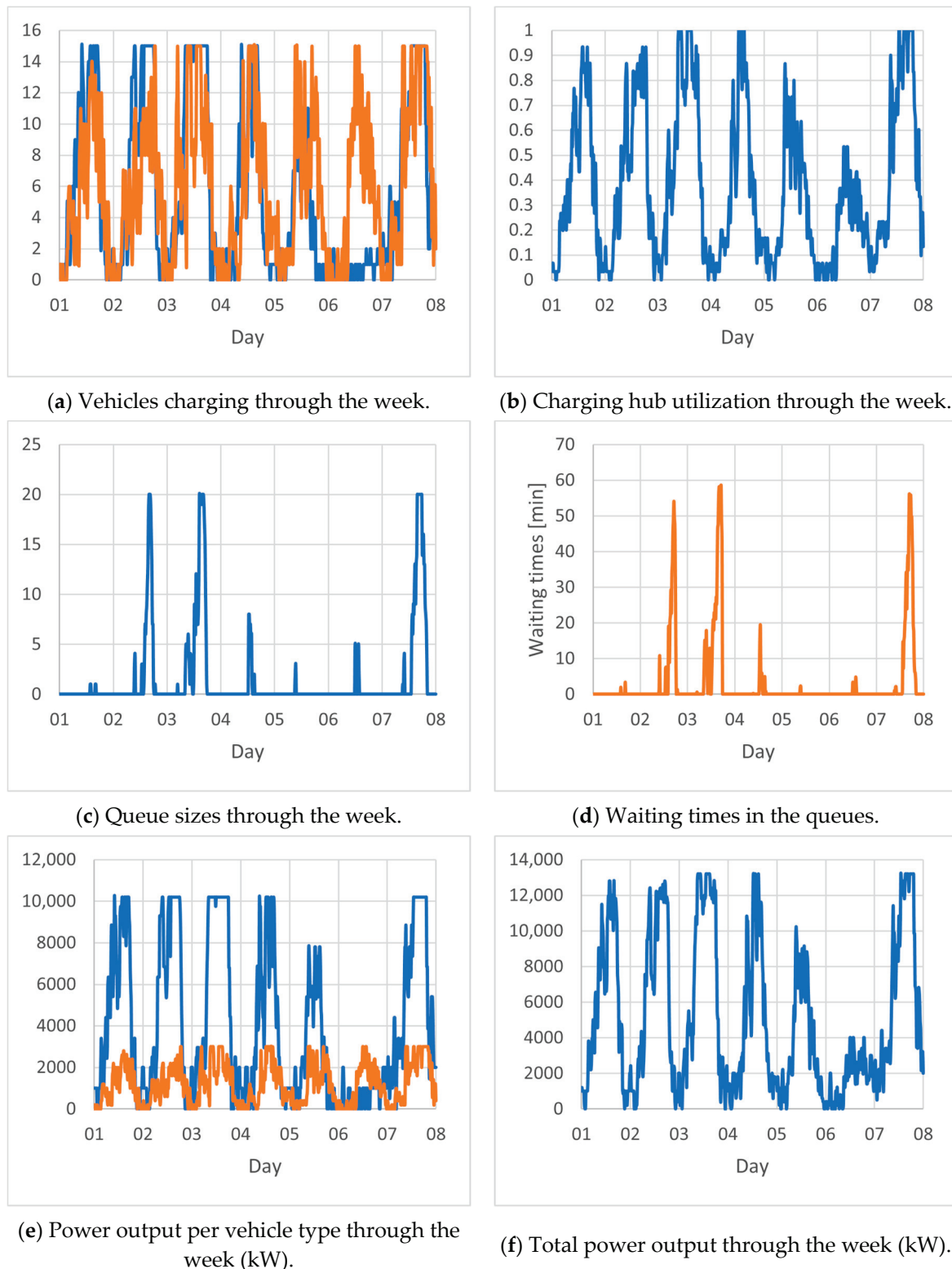


Figure 7. Cont.

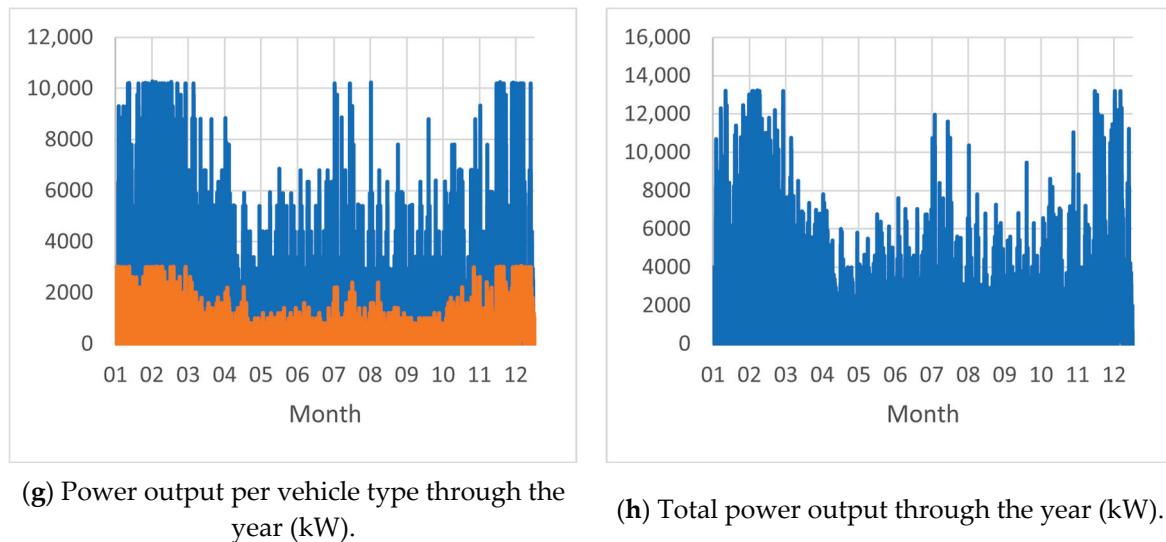


Figure 7. Results of S2 analysis for the charging hub operations with the number of vehicles charging (a), charging hub utilization (b), queues (c), waiting times (d), weekly power outputs per vehicle type (e), total output (f) and yearly power outputs per vehicle type (g), and total output (h).

5. Discussion

Providing proper electrical infrastructure is a necessity in order to ensure that EV traffic flow is fluent. While it may be expected that due to technological developments EV charging may eventually resemble classical fueling from the customers point of view, it is necessary to consider other stakeholders, which include traffic management, charging service providers, and transmission and distribution system operators. From the individual stakeholder view, they may have different objectives that are not necessarily aligned.

As charging generates revenue, charging providers should seek the maximal utilization of the individual fast chargers, which varies throughout the day (and night) as well as through different seasonal traffic intensities. The cost of installing chargers warrants suitable utilization through their life cycle. Too many underutilized chargers can reduce the return on the investment, implying that predicting future EV traffic requirements is essential for planning.

On the other hand, maximizing the utilization of chargers can reduce the throughput of the charging hub, creating queues and increasing waiting times, displeasing drivers as well as increasing chances of congestions. While installing transmission and distribution lines is mostly limited by geographical issues mentioned in the introduction, substation capacities need to be planned with reference to the expected power loads. As substation costs increase with higher outputs, the number of charging stations and simultaneous utilization provides that estimate for power requirements for fast charging, and it is additionally necessary to consider other requirements such as for other commercial facilities at the charging hub. Thoughtful planning is needed here to consider not only the requirements in the near future but anticipate the demands in the long term.

It is necessary to consider the implications that both location as well as configuration of the charging hub will have on traffic flows. If the charging and grid infrastructure will not adjust to the adoption of EVs, this may lead to increased traffic on secondary roads along the highway. On the other hand, dynamic pricing or time-slot reservations might influence the behavior of the traffic flow itself, leading to a potentially lesser traffic peaks, although this topic requires further research. Possible extensions of the hybrid model could incorporate motorway-specific pricing strategies, including time-of-use rates, dynamic pricing, and reservation-based fees. These policies can influence not only charging behavior

but also traffic flows by redirecting demand to less congested routes or other time windows. The effectiveness of such strategies may vary across different user profiles: for instance, private drivers may respond more strongly to pricing incentives, while commercial vehicles are more sensitive to time constraints and may prioritize minimal downtime over cost savings. By segmenting vehicle types and adjusting charging decision logic accordingly, the model could simulate differentiated behavioral responses and support the design of pricing schemes that balance infrastructure use and operational efficiency across user groups.

Analyzing future charging demands for electrical grid and charging infrastructure planning that can be achieved by the presented method can be straightforwardly upgraded for analysis of charging hub grid operations, which can provide increased precision for predicting the operation of each individual charging hub and also the impact on traffic flows. Further, the method can be adapted to consider different pricing schemes and competition from different charging service providers.

The presented method can be used as a part of a digital twin that integrates real-time data from traffic monitors, data from other charging hubs, weather forecasts, and specific events that can influence the expected traffic or conditions. The presented hybrid model was designed flexibly so that any input can be replaced with different loads. As such, the traffic forecast could be replaced with publicly accessible traffic monitoring data, and temperature conditions—with data from weather forecast services. Estimations for battery parameters and EV charging outputs can be replaced by real specification once the charging hubs are constructed. Given that many of the parameters presented in this case are modelled based on either estimates from other studies or assumptions, replacing those inputs with those from existing services can provide both real-time and more precise short-term prediction of charging operations. This also provides both a calibration tool for the presented model as well as a validation option by comparing real-world operations to modelled results.

6. Conclusions

This study presents a hybrid model for evaluating motorway fast-charging infrastructure, combining traffic volume-based demand estimation with discrete-event simulation. The approach accounts for dynamic factors such as vehicle state-of-charge (SoC), battery efficiency influenced by temperature, and probabilistic charging behavior across diverse EV types. Applied to a realistic national motorway corridor scenario, the model demonstrates significant seasonal variation in energy demand and queueing behavior, particularly under winter conditions. It also revealed that the charger hub configuration—especially the mix of power levels and spatial spacing—had a direct impact on wait times and load distribution.

The model advances the existing literature by integrating behavioral decision logic with operational simulations in a transit setting, where infrastructure must serve directional, long-distance flows. Unlike urban-focused models, it explicitly captures constraints and patterns unique to highway environments, including limited detour flexibility and route-fixed demand. The simulation's modular design allows key parameters to be easily updated or replaced with real-time inputs, offering potential for deployment in digital twin systems and interactive planning tools.

However, the study also has limitations. Charging behavior probabilities and efficiency adjustments were based on simplified assumptions and aggregate estimates rather than empirical trip-level data. Additionally, the model did not include spatial origin–destination (OD) matrices or grid-side constraints, which may affect the infrastructure layout and power availability.

Future work will focus on integrating the data of traffic generators and OD data to improve traffic flow accuracy and extend the model toward network-wide coordination of

charging infrastructure. While the current model focused on the physical layout, traffic flow, and environmental conditions affecting motorway charging demand, behavioral and economic factors such as pricing strategies and user segmentation are areas for future investigation. The impact of dynamic pricing, time-slot reservations, and user-specific constraints could significantly influence charging patterns and the traffic distribution. As these mechanisms are increasingly adopted in real-world networks, incorporating them into simulation frameworks is essential for providing more realistic demand scenarios and supporting policy design.

Furthermore, integration with grid simulation models or co-simulation platforms would enable full-system evaluations that reflect both traffic and energy levels. These developments would support real-time decision making and robust infrastructure planning in line with evolving EV policy targets across the TEN-T network.

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

Funding: This research was supported by the “Innovation Yongjiang 2035” Key R&D Programme (Grant No. 2024H032) and by the DigiEles project funded by the European Recovery and Resilience Facility.

Data Availability Statement: The data that 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. Directive-2014/94-EN-EUR-Lex. Directive 2014/94/EU. Available online: <https://eur-lex.europa.eu/eli/dir/2014/94/oj/eng> (accessed on 20 February 2025).
2. Szumska, E.M. Electric Vehicle Charging Infrastructure along Highways in the EU. *Energies* **2023**, *16*, 895. [CrossRef]
3. DARS d.d. Available online: <https://www.dars.si/> (accessed on 20 February 2025).
4. Road Goods Transport. 2023. Available online: <https://www.stat.si/StatWeb/en/News/Index/12824> (accessed on 30 March 2025).
5. Shoman, W.; Yeh, S.; Sprei, F.; Plötz, P.; Speth, D. *Public Charging Requirements for Battery Electric Long-Haul Trucks in Europe: A Trip Chain Approach*; Fraunhofer-Gesellschaft: Munich, Germany, 2023. [CrossRef]
6. Morrissey, P.; Weldon, P.; O'Mahony, M. Future standard and fast charging infrastructure planning: An analysis of electric vehicle charging behaviour. *Energy Policy* **2016**, *89*, 257–270. [CrossRef]
7. Hipolito, F.; Vandet, C.A.; Rich, J. Charging, steady-state SoC and energy storage distributions for EV fleets. *Appl. Energy* **2022**, *317*, 119065. [CrossRef]
8. Lee, J.H.; Chakraborty, D.; Hardman, S.J.; Tal, G. Exploring electric vehicle charging patterns: Mixed usage of charging infrastructure. *Transp. Res. Part Transp. Environ.* **2020**, *79*, 102249. [CrossRef]
9. Azadfar, E.; Sreeram, V.; Harries, D. The investigation of the major factors influencing plug-in electric vehicle driving patterns and charging behaviour. *Renew. Sustain. Energy Rev.* **2015**, *42*, 1065–1076. [CrossRef]
10. Diaz-Londono, C.; Fambri, G.; Maffezzoni, P.; Gruosso, G. Enhanced EV charging algorithm considering data-driven workplace chargers categorization with multiple vehicle types. *eTransportation* **2024**, *20*, 100326. [CrossRef]
11. Varga, B.O.; Sagoian, A.; Mariasiu, F. Prediction of Electric Vehicle Range: A Comprehensive Review of Current Issues and Challenges. *Energies* **2019**, *12*, 946. [CrossRef]
12. Thorhaug, M.; Rich, J.; Mabit, S.E. Charging behaviour and range anxiety in long-distance EV travel: An adaptive choice design study. *Transportation* **2024**, 1–23. [CrossRef]
13. Al-Hanahi, B.; Ahmad, I.; Habibi, D.; Masoum, M.A.S. Charging Infrastructure for Commercial Electric Vehicles: Challenges and Future Works. *IEEE Access* **2021**, *9*, 121476–121492. [CrossRef]
14. Mazur, M.; Dybała, J.; Kluczek, A. Suitable law-based location selection of high-power electric vehicles charging stations on the TEN-T core network for sustainability: A case of Poland. *Arch. Transp.* **2024**, *69*, 75–90. [CrossRef]

15. Colombo, C.G.; Borghetti, F.; Longo, M.; Foiadelli, F. Electrification of Motorway Network: A Methodological Approach to Define Location of Charging Infrastructure for EV. *Sustainability* **2023**, *15*, 16429. [CrossRef]
16. Schmidt, M.; Zmuda-Trzebiatowski, P.; Kiciński, M.; Sawicki, P. Lasak, “Multiple-Criteria-Based Electric Vehicle Charging Infrastructure Design Problem. *Energies* **2021**, *14*, 3214. [CrossRef]
17. Liu, J.-Y.; Liu, S.-F.; Gong, D.-Q. Electric Vehicle Charging Station Layout Based on Particle Swarm Simulation. *Int. J. Simul. Model.* **2021**, *20*, 754–765. [CrossRef]
18. Bian, H.; Ren, Q.; Guo, Z.; Zhou, C.; Zhang, Z.; Wang, X. Predictive Model for EV Charging Load Incorporating Multimodal Travel Behavior and Microscopic Traffic Simulation. *Energies* **2024**, *17*, 2606. [CrossRef]
19. Guerrero, J.D.A.; Bhattarai, B.; Shrestha, R.; Acker, T.L.; Castro, R. Integrating Electric Vehicles into Power System Operation Production Cost Models. *World Electr. Veh. J.* **2021**, *12*, 263. [CrossRef]
20. Lopez, N.S.; Allana, A.; Biona, J.B.M. Modeling Electric Vehicle Charging Demand with the Effect of Increasing EVSEs: A Discrete Event Simulation-Based Model. *Energies* **2021**, *14*, 3734. [CrossRef]
21. Metais, M.O.; Jouini, O.; Perez, Y.; Berrada, J.; Suomalainen, E. Too much or not enough? Planning electric vehicle charging infrastructure: A review of modeling options. *Renew. Sustain. Energy Rev.* **2022**, *153*, 111719. [CrossRef]
22. Chaudhari, K.; Kandasamy, N.K.; Krishnan, A.; Ukil, A.; Gooi, H.B. Agent-Based Aggregated Behavior Modeling for Electric Vehicle Charging Load. *IEEE Trans. Ind. Inform.* **2019**, *15*, 856–868. [CrossRef]
23. Zhang, Y.; Tan, J. A data-driven approach of layout evaluation for electric vehicle charging infrastructure using agent-based simulation and GIS. *Simulation* **2024**, *100*, 299–319. [CrossRef]
24. Çelik, S.; Ok, Ş. Electric vehicle charging stations: Model, algorithm, simulation, location, and capacity planning. *Heliyon* **2024**, *10*, e29153. [CrossRef]
25. Unterluggauer, T.; Rich, J.; Andersen, P.B.; Hashemi, S. Electric vehicle charging infrastructure planning for integrated transportation and power distribution networks: A review. *eTransportation* **2022**, *12*, 100163. [CrossRef]
26. Lartey, J.D. Modelling of Daily Long-Term Urban Road Traffic Flow Distribution: A Poisson Process Approach. *Open J. Model. Simul.* **2025**, *13*, 89–105. [CrossRef]
27. Gramaglia, M.; Serrano, P.; Hernandez, J.A.; Calderon, M.; Bernardos, C.J. New insights from the analysis of free flow vehicular traffic in highways. In Proceedings of the 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Lucca, Italy, 20–24 June 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–9. [CrossRef]
28. Lindgren, J.; Lund, P.D. Effect of extreme temperatures on battery charging and performance of electric vehicles. *J. Power Sources* **2016**, *328*, 37–45. [CrossRef]
29. Wu, Z.; Bhat, P.K.; Chen, B. Optimal Configuration of Extreme Fast Charging Stations Integrated with Energy Storage System and Photovoltaic Panels in Distribution Networks. *Energies* **2023**, *16*, 2385. [CrossRef]
30. Etengoff, A. What is an EV Battery State of Charge (SOC)? EV Engineering & Infrastructure. Available online: <https://www.evengineeringonline.com/what-is-an-ev-battery-state-of-charge-soc/> (accessed on 7 March 2025).
31. Electric Vehicles in Europe: Accelerating Sales Trends and Regulatory Impacts. Available online: <https://www.enerdata.net/publications/executive-briefing/electric-vehicles-trends.html> (accessed on 20 February 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Assessing the Influence of Equipment Reliability over the Activity Inside Maritime Container Terminals Through Discrete-Event Simulation

Eugen Rosca ¹, Florin Rusca ¹, Valentin Carlan ², Ovidiu Stefanov ¹, Oana Dinu ¹ and Aura Rusca ^{1,*}

¹ Faculty of Transport, National University of Science and Technology Politehnica Bucharest, 060042 Bucharest, Romania; eugen.rosca@upb.ro (E.R.); florin.rusca@upb.ro (F.R.); ovidiu.stefanov2019@gmail.com (O.S.); oana.dinu@upb.ro (O.D.)

² Faculty of Business and Economics, University of Antwerp, 2000 Antwerpen, Belgium; valentin.carlan@uantwerpen.be

* Correspondence: aura.rusca@upb.ro; Tel.: +40-72-639-7851

Abstract: (1) Background: The reliability of port equipment is of significant interest to industry stakeholders due to the economic and logistical factors governing the operation of maritime container terminals. Failures of key equipment like quay cranes can halt operations or cause economically significant delays. (2) Methods: The impact assessment of these disruptive events is conducted through terminal activity modeling and discrete-event simulation of internal processes. The system's steady-state or transient condition, induced by disruptive events, is statistically assessed within a set of scenarios proposed by the authors. (3) Results: The Heidelberg–Welch and Geweke tests enabled the evaluation of steady-state and transient conditions within the modeled system, which was affected by the reduced reliability of container-handling equipment. (4) Conclusions: The research findings confirmed the usefulness of modeling and simulation in assessing the impact of equipment reliability on maritime container terminal operations. If the magnitude of the disruptive event exceeds the terminal's absorption capacity, the system may become blocked or remain in a transient state without the ability to recover. This underscores the necessity of analyzing the reliability of critical handling equipment and implementing corrective maintenance actions when required.

Keywords: maritime container terminal; modeling; discrete-event simulation; stationarity convergence; handling equipment reliability

1. Introduction

Maritime container terminals are a crucial component of global logistics chains, serving as key hubs connecting maritime transport with land-based networks. The movement of goods across countries and continents is facilitated by high-capacity container ships. By using maritime container terminals, economic connections between geographically distant nations are strengthened, contributing to global economic growth and development. According to UNCTAD, containerized transport remains essential in 2023, with an estimated volume of transported goods nearing 2000 million tons. The number of loading units has also remained stable at approximately 164 million twenty-foot equivalent units (TEUs) [1]. These figures underscore the operational pressure faced by maritime container terminals providing uninterrupted service while mitigating the impact of disruptive events. Such events include natural phenomena (e.g., storms, strong winds, snow) as well as failures in handling, storage, and transport equipment within the terminal. The research presented in

this study focuses on evaluating the impact of the latter category on terminal operations. The terminal is modeled as a dynamic system characterized by multiple operational states, which arise from either full-capacity or under-capacity operation resulting from disruptive events, such as the failure of a quay crane. Thus, the system could be in a stationary or non-stationary state.

Efficient container terminal operations are a critical goal for terminal management. The main objective of the research is achieved by designing a layout that meets the terminal's specific operational requirements. The layout must effectively integrate various modes of transport (maritime vessels, trucks, trains), handling and storage equipment, and designated storage areas within the terminal [2]. The proper functioning of handling, storage, and transport equipment is analyzed using reliability functions and performance indicators. The occurrence of a disruptive event can significantly impact the terminal's economic efficiency [3]. Equipment failures, particularly those affecting container-handling machinery, can lead to substantial operational disruptions. While terminal operational reserves may help absorb and mitigate these disruptions within an acceptable timeframe, in certain cases, the magnitude of the disturbance results in severe degradation of service quality parameters, ultimately affecting the terminal's economic productivity. Research in this field primarily focuses on modeling the activities of the terminal and evaluating key service quality parameters, such as waiting times, the number of handling operations and storage area optimization. However, a notable gap exists in analyzing the impact of handling equipment failures resulting from reduced reliability. The practical implications of handling equipment reliability for maritime container terminal administrators can be highly diverse. The strategic and operational analysis conducted by these administrators must consider equipment reliability, its utilization level within terminal processes, and the costs associated with its acquisition, operation, and potential failures when making investment decisions. In many cases, maritime container terminals operate within multinational corporations that strictly manage acquisition and operational costs. The occurrence of an unforeseen disruptive event is often difficult for the management of such companies to accept. However, the impact of such an event can be minimized if it primarily affects equipment with a low influence on overall terminal operations due to their availability—such as a terminal truck breakdown. Conversely, it can be a major influence if critical equipment, such as a quay crane, fails. A reduction in the handling and storage capacity at the terminal level is not easily accepted by its management. For this reason, research on system state evaluation—whether stable or transient—is of interest both from an organizational perspective and in terms of economic implications.

In this study, discrete-event modeling was employed as a methodological approach, with a practical application to a real-world container terminal in Europe. The input data for the model was collected from terminal administration and processed to ensure accurate integration into the developed simulation framework. The primary data collected included information on the arrival times of vessels at the container terminal, the technological processes occurring within the terminal, the time durations required for these processes, as well as topological aspects necessary for simulation related to terminal design. Since data collection was conducted during summer, disruptions caused by meteorological phenomena that could lead to terminal activity interruptions or restrictions on maritime vessel access to berths are not considered.

A set of scenarios was designed to assess the impact of quay crane failures over varying time periods. The data analysis enabled the identification of periods in which the modeled system operated in either a transient or steady state. Additionally, scenarios in which a return to the steady-state condition in due time is no longer possible were identified.

Building upon the previously introduced concepts and methodology, this research aims to identify the system states in response to a disruptive event. In the specific case of maritime container terminals, discrete-event modeling provides a feasible tool for evaluating operational parameters that define a system's stable or transient state. To comprehensively present the research conducted, the article is structured as follows:

- **Introduction:** This section provides a descriptive overview of the challenges encountered in container terminal operations and identifies the key activities that may be disrupted by external perturbations.
- **Literature Review:** A critical review of existing studies in the field is conducted to identify research gaps and establish the baseline from which this study originates.
- **Modeling the Equipment Reliability:** This section presents the mathematical framework underlying equipment reliability assessment, system state analysis, and stationarity vectors.
- **Simulation Model and System States:** It details the methodology and structural components of the simulation model used as a tool for generating operation scenarios.
- **Case Study: Constantza South Container Terminal:** This section describes the primary data collected from a real maritime container terminal and presents the scenarios tested in the simulation model to evaluate system state transitions.
- **Results and Discussion:** The findings obtained through simulation are presented, with a focus on waiting times measured for each simulation entity across the seven proposed scenarios.
- **Conclusions:** This section summarizes the authors' conclusions regarding the necessity and innovative aspects of the research, as well as future research directions derived from the obtained results. The identified limitations of the model are also discussed, along with potential methods to address them.

2. Literature Review

Due to its significant economic importance, the maritime container terminal has been the subject of extensive research, which has addressed various aspects, including strategic development, marine pollution, modeling of both terminal operations and related activities, as well as the evaluation of terminal reliability as a complex system. According to Steenken et al., three decision-making levels can be identified (strategic, operative planning, and real-time control) [4].

From a strategic perspective, research has been conducted both at the level of logistics chain [5,6] and through comparative analyses of land-use and terminal localization [7]. Some studies focus on the specific aspects of developing new terminals [8], while others assess the factors influencing the competitiveness of maritime container ports [9]. A specialized area of research within this strategic framework examines the impact of mooring of ultra-large container vessels on port infrastructure. The increasing deployment of high-capacity vessels has significant implications for port infrastructure development [10]. Simulation-based research can also be used to identify strategies for port infrastructure development. This approach enables the identification of opportunities for port expansion to become local or regional hubs [11].

Another key research direction addresses the environmental impact of maritime container terminal operations. Advanced methodologies, such as the Automatic Identification System (AIS), are employed to decode vessel data and estimate ship emissions. These analyses incorporate information on ship technical specifications and operational characteristics [12]. Additionally, research on pollution examines the technical aspects of data transmission from monitoring systems [13,14]. Indicators such as the Ship Carbon Intensity, proposed by the International Maritime Organization, play a crucial role in assessing the

environmental impact of vessel scheduling for cargo handling activities [14]. In this context, optimization processes in the terminal may inadvertently lead to negative environmental consequences. However, maritime terminals are also evaluated in terms of sustainability strategies. For instance, implementing automated mooring and unmooring technologies can reduce greenhouse gas emissions in port operations [15]. Furthermore, reviews of environmentally friendly practices, technologies, and equipment contribute to define a “model” port concept, which could be theoretically adopted by port administrations worldwide [16,17]. In relation to port terminals, the maritime transport component and voyage attributes, such as travel speed, can affect pollutant emissions, cargo lead time, and supply chain costs [18]. Likewise, in terms of the environmental component, the choice of a specific mode of transport, such as container transport, ensures the integration of maritime terminals into the sustainable development strategy of ports [19]. The connection with terrestrial transport networks can also be evaluated from an environmental perspective, particularly in terms of emissions. Urban congestion in port cities can lead to pollution hotspots, which is why research in this area is also of significant importance [20]. Scenarios involving new technologies (electric trains, hydrogen trains) with a positive environmental impact have been studied using simulation tools. The results obtained thus enable the development of optimal strategies for their control to maximize benefits [21].

At the operational planning level, research has focused on mathematical modeling of interactions between transport modes, handling equipment, and even between multiple container terminals. Studies in this area explore ways to enhance port performance by improving knowledge of terminal operating systems [22–34]. One key aspect of this research involves evaluating the interaction between land transport networks and terminal operations. Metrics such as truck queue length and waiting times at terminal gates are critical for assessing service efficiency [23–30]. The event-oriented simulation method can be used to simulate road traffic in the port area for a comprehensive analysis of port logistics systems. Discrete-event modeling thus proves to be a valuable tool for evaluating these systems [27]. Discrete-event simulation (DES) models are highly versatile tools, with applications in cost–benefit analysis, predicting container arrival times at terminals and berth allocation within port terminals [28–30]. Findings indicate that implementing a structured truck appointment system can significantly influence drayage costs and overall efficiency [31]. Similarly, optimizing rail transport and synchronizing terminal activities with adjacent railway stations have been shown to reduce delays and increase transit capacity [32,33].

At the same operational level, research has addressed berth allocation and container relocation problems. Heuristic methods and genetic algorithms have proven to be useful in solving these challenges [34,35]. Additionally, simulation models offer advantages in terms of adaptability and time compression for scenario-based analyses [36,37].

A notable gap in existing research at this level is the analysis of the impact of disruptive events on terminal operations. Few studies have explored the diverse nature of these disruptions. One such study [38] examines the impact upon business of power outages at container terminals. In another study [39], the authors focus on resilience measurement and dynamic optimization of the container logistics supply chain under adverse events. However, the study is largely theoretical. A clear gap emerges between the well-developed theoretical concepts presented by the authors and the application of discrete-event simulation, which is essential for accurately modeling operations in a maritime container terminal. The present research aims to address this gap by analyzing the reliability of handling equipment and the consequences of equipment failures. This area of study has significant practical implications, providing valuable insights for maritime container terminal administrators in enhancing operational resilience.

3. Modeling the Equipment Reliability

3.1. The Equipment Reliability Indicators

The reliability function $R(t)$ is calculated as a probability that the equipment will perform its intended function without failure under specified conditions for a certain period. A set of indicators is used to evaluate the dependability and consistent performance of the equipment:

- Mean Time Between Failures (MTBF): the average time between successive failures for repairable equipment.
- Mean Time To Failure (MTTF): the average time for a piece of equipment to operate before failing, applicable to non-repairable equipment.
- Failure Rate ($\lambda(t)$): the frequency at which equipment fails, expressed in failures per unit of time ($t, t + \Delta t$).
- Availability: the proportion of time that the equipment is in a functioning state. This metric is used for repairable equipment.
- Maintainability (Mean Time To Repair): the average time required to repair a piece of equipment.

Other mathematical indicators used can be:

- Failure function ($F(t)$): the probability that an equipment will fail during the time interval $(0, t)$
- The breakdown frequency of failures ($f(t)$): the limit of the ratio between the probability of failure in the interval $(t, t + \Delta t)$ and the size of the interval

The primary relationships between these indicators are summarized in Table 1.

Table 1. Dependences among reliability indicators [40].

Indicators	Indicator According to Which It Is Expressed			
	$F(t)$	$f(t)$	$R(t)$	$\lambda(t)$
$F(t)$	-	$\int_0^t f(u)du$	$1 - R(t)$	$1 - e^{-\int_0^t \lambda(u)du}$
$f(t)$	$\frac{dF(t)}{dt}$	-	$-\frac{dR(t)}{dt}$	$\lambda(t)e^{-\int_0^t \lambda(u)du}$
$R(t)$	$1 - F(t)$	$\int_t^\infty f(u)du$	-	$e^{-\int_0^t \lambda(u)du}$
$\lambda(t)$	$\frac{1}{1-F(t)} \frac{dF}{dt}$	$\frac{f(t)}{\int_t^\infty f(u)du}$	$-\frac{1}{R(t)} \frac{dR}{dt}$	-
MTBF	$\int_0^\infty [1 - F(t)]dt$	$\int_0^\infty tf(t)dt$	$\int_0^\infty R(t)dt$	$\int_0^t e^{-\int_0^t \lambda(u)du} dt$

For different distributions of the failure function $F(t)$, the reliability indicators are depicted in Table 2.

Table 2. Reliability indicators [40].

Indicators	Distribution Function			
	Exponential	Normal	Weibull	Log-Normal
$F(t)$	$1 - e^{-\lambda t}$	$\Phi\left(\frac{t-m}{\sigma}\right)$	$1 - e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}$	$\Phi\left(\frac{\ln t - m}{\sigma}\right)$
$f(t)$	$e^{-\lambda t}$	$\Phi\left(\frac{m-t}{\sigma}\right)$	$e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}$	$\Phi\left(\frac{m - \ln t}{\sigma}\right)$
$R(t)$	$\lambda e^{-\lambda t}$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}}$	$\frac{\beta(t-\gamma)^{\beta-1}}{\eta^\beta} e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}$	$\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - m)^2}{2\sigma^2}}$
$\lambda(t)$	λ	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}}$	$\frac{\beta(t-\gamma)^{\beta-1}}{\eta^\beta}$	$\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - m)^2}{2\sigma^2}}$
MTBF	$\frac{1}{\lambda}$	m	$\gamma + \eta\gamma\left(\frac{1}{\beta} + 1\right)$	$e^{m + \frac{\sigma^2}{2}}$

where λ , σ , β and m are specific parameters of different distribution functions presented in table.

3.2. The Transient Period Toward Stationarity

Following a disruption of the container handling process (e.g., equipment breakdown, accidents), the system enters a non-stationary phase, where ships and containers are queuing, and increased waiting times are counted. Once the disruption ends, if the arrival process is stationary, the system moves asymptotically toward a steady state characterized by three conditions: (i) constant mean (the mean of the time series is constant over time), (ii) constant variance (the variance of the time series is constant over time), and (iii) constant autocovariance (the covariance between observations at any two points in time depends only on the time lag between them). Heavily loaded systems tend more slowly to a stationary state. Burn-in data refers to the portion of discarded observations so that the effect of initial values on the posterior sequence is minimized. A variety of methods have been proposed to circumvent the non-stationarity and the autocorrelation among collected observations and to evaluate the transient phase duration to steady state (burn-in data) [41,42].

Based on the remark that removing the initial transient data from a given sequence of observations has the remaining sequence behaving as a stochastic stationary process, Schruben [43] and Schruben et al. [34] developed a test to identify the transient phase. Given the output series of observations $x_1, x_2, \dots$, the test is based on the high sensitivity of the sequence of partial sums:

$$S_k = \bar{X}(n) - \bar{X}(k), \quad (1)$$

where $k = 0, 1, \dots, S_0 = S_n = 0$, $\bar{X}(n)$ and $\bar{X}(k)$ are means over n and k first observations. Assuming under null hypothesis that the output series is stationary, the standardized sequence of partial sums, namely,

$$B(t) = \frac{\lfloor nt \rfloor S_{\lfloor nt \rfloor}}{\hat{\sigma}[\bar{X}(n)]\sqrt{n}}, \quad 0 < t \leq 1, \text{ and } \lfloor \dots \rfloor \text{ is the integer part}, \quad (2)$$

converges to a Brownian bridge process with zero mean and variance equal to 1 as $n \rightarrow \infty$.

Using spectral analysis of variance and the findings of Schruben [33] and Schruben et al. [44] on detecting initial transients using Brownian bridge theory, Heidelberg and Welch [45] proposed a convergence test consisting of two parts:

1. Calculation of the Cramér–von Mises (CVM) statistic on the whole sequence and test the null hypothesis that it comes from a stationary distribution. If the null hypothesis is rejected, then discard the first 10% of the sequence and repeat the stationarity test. If the null hypothesis is rejected, then discard the next 10% and calculate the statistical test. Repeat until the null hypothesis is accepted or 50% of the initial sequence is discarded. If the test still rejects the null hypothesis, then the observation sequence fails the test and needs to be run longer.
2. If the sequence passes the first part of the diagnostic, then the part of the sequence that was not discarded from the initial one is used to compute the half-width of the mean. Half-width of this interval is compared with the estimate of the mean. If the ratio between the half-width and the mean is lower than ϵ , the half-width test is passed. Otherwise, the length of the sample is deemed not long enough to estimate the mean with sufficient accuracy and more observations must be added.

Heidelberg and Welch [45] suggest increasing the run length by a factor $I > 1.5$, each time, so that the estimate has the same, reasonably large proportion of new data.

Geweke [46] proposed a convergence diagnostic based on a test for equality of the means of the first and last part of a chain. Two sequences of the observation data are used x_1^t ($t = 1, \dots, n_1$) and x_2^t ($t = n_a, \dots, n$), where $1 < n_1 < n_a < n$, being let $n_2 = n - n_a + 1$

the size of the second sequence. The two sequences do not overlap, so $n_1 + n_2 < n$. The means of the two sequences are $\bar{X}_1 = \frac{1}{n_1} \sum_{t=1}^{n_1} x_1^t$ and $\bar{X}_2 = \frac{1}{n_2} \sum_{t=n_a}^n x_2^t$. Being let $\hat{s}_1(0)$ and $\hat{s}_2(0)$ denote the consistent spectral density at zero frequency for the two sequences. If the initial sequence is stationary, then the statistics

$$Z = \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{\frac{\hat{s}_1(0)}{n_1} + \frac{\hat{s}_2(0)}{n_2}}} \quad (3)$$

converges to standard normal distribution $N(0,1)$ as $n \rightarrow \infty$. The Geweke test is a two-tailed test with the null hypothesis of equal means of the two sequences. A large absolute value of Z ($Z > 1.96$) indicates rejection of the null hypothesis.

To test the stationarity of the ship's waiting time and to compute the burn-in numbers of observed data, the following steps are taken:

1. To use the Heidelberg–Welch test to prove that the initial ship's waiting time sequence is stationary and compute the confidence interval of the ship's waiting time mean. If the test fails, a larger amount of data are required, and the first step is repeated.
2. To use the Geweke test for $n_1 = 0.1 \times n$ and $n_2 = 0.5 \times n$ as the author suggested. The first half of the initial ship's data sequence is divided into $n_{bins} - 1$ segments, then Geweke's Z-score is repeatedly calculated. The first Z-score is calculated with all iterations in the chain, the second after discarding the first ship's wait time data segment, the third after discarding the first two segments, and so on. The last Z-score is calculated using only the samples in the second half of the ship's wait time data sequence. The analysis of the sequence of Geweke's Z-score indicates the burn-in ship's data, where the Z-statistic is outside the 95% confidence interval. Following the burn-in point, the stationarity of the ship's waiting time is achieved.

4. Simulation Model and System States

To assess the impact of the equipment reliability on operations within a maritime container terminal, the first step is to model terminal activities using a simulation model. At this stage, key activities that significantly influence the quality of services provided by the terminal are identified. In the case of a maritime terminal, quay cranes are among the most critical components due to their essential role in loading and unloading containers from maritime vessels. The level of detail in each individual simulation can lead to the omission of critical components of terminal operations or introduce evaluation errors by generating inconsistencies that are difficult to detect. In the case of the simulation tool used in the paper's case study, the authors have identified the significant impact of quay crane operations on overall terminal performance. Specifically, the handling capacity of containers both on the vessel and at the quay can act as a limiting factor for the entire maritime terminal's throughput.

The second stage of the research involves identifying the various states in which quay cranes can operate and modeling their reliability using a suitable distribution function. The system's stationary state may transition into a transient state and vice versa, depending on equipment reliability and operational conditions.

Finally, specific computational procedures are applied to evaluate the duration for which the system remains in a transient state, during which the operating parameters of the container terminal are affected. The research framework is illustrated in Figure 1.

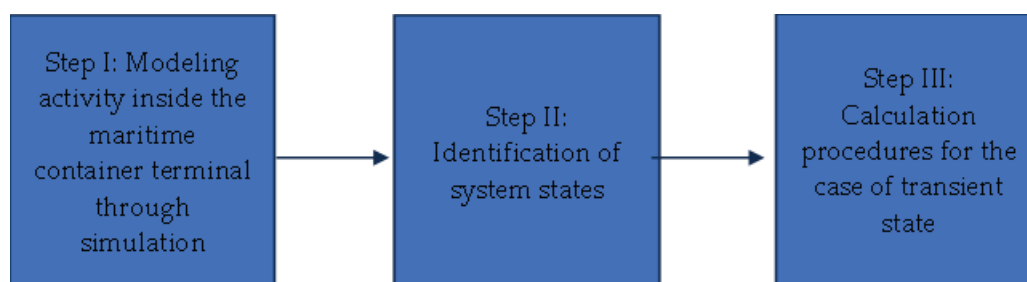


Figure 1. The logical flow of the research—source: authors.

The topology of maritime terminals and the interdependencies between container handling and storage processes facilitate their modeling through queuing theory. The discrete-event simulation modeling is preferred due to its suitability for capturing complex system dynamics. The effectiveness of this modeling approach has been demonstrated in previous research within the same or related fields [47–49]. The simulation model was developed using ARENA 12 software. Through dedicated modules, simulation entities are created, and logical processes are employed to replicate real-world operations within a maritime container terminal. Each simulation block serves a specific function, as outlined in Table 3.

Table 3. The simulation modules for the maritime container terminal model.

Module Type	ARENA Module	Role and Parameters
Create/divide/clone simulation entities	Vessel arrival	The arrival time of a new vessel is stochastically determined using a distribution function
	Terminal activity	The simulation entity of vessel is cloned to be able to separate the simulation thread into two paths
	Generating containers for loading process	The vessel simulation entity is cloned. The number of clones is equal to the number of containers allocated to be loaded
	Generating containers for upload	The vessel simulation entity is cloned. The number of clones is equal to the number of containers allocated to be uploaded
Seizing a resource	Berth seizing	A berth-type resource is seized and cannot be occupied by another simulation entity
	Unloading process	For the simulation entities associated with the containers that are unloaded from the vessels, a berth crane-type resource is used
	Loading process	For the simulation entities associated with the containers that are loaded from the vessels, a berth crane-type resource is used
	Truck loading	After unloading from the vessel, an internal truck from the terminal has the role of moving the container to the storage area
Delay time for another process	Berthing	Berthing is the process of safely navigating a vessel into its designated position
	Waiting for logistic activity	The vessel is waiting for the completion of the container handling activity
	Vessel exit	The vessel is leaving the terminal area

Table 3. Cont.

Module Type	ARENA Module	Role and Parameters
Setting simulation attributes	Generating the number of containers	Assigning the number of containers to be loaded and unloaded from every vessel
	Initialize control variables	Initialization of control variables that will be used in tracking the loading/unloading process of the containers
	Update control variable-uploading process	The control variable for uploading process is updated after every upload action of container from vessel to berth
	Update control variable-loading process	The control variable for loading process is updated after every load action of container from vessel to berth
Decisional modules	Unloading process is finish?	It is tested if the control variable has reached the set value for the number of unloaded containers
	Loading process is finish?	It is tested if the control variable has reached the set value for the number of loaded containers

To simulate the occurrence of a failure of the container handling equipment, a set of new modules are introduced into the simulation model (Table 4).

Table 4. The simulation modules for controlling equipment failure.

Module Type	ARENA Module	Role and Parameters
Creation of an event associated with equipment failure	Reliability QC1	Constant intervals of one hour are generated in which the state of quay crane 1 is checked
	Reliability QC2	Constant intervals of one hour are generated in which the state of quay crane 1 is checked
	Reliability QC3	Constant intervals of one hour are generated in which the state of quay crane 1 is checked
Setting the state of system	Test malfunction of QC1	The decision-making mode with the help of which the operating status of the QC1 equipment is decided with the support of a Failure function $F(t)$
	Test malfunction of QC2	The decision-making mode with the help of which the operating status of the QC2 equipment is decided with the support of a Failure function $F(t)$
	Test malfunction of QC3	The decision-making mode with the help of which the operating status of the QC3 equipment is decided with the support of a Failure function $F(t)$
Setting the good function attribute of quay crane	Malfunction QC1	Resource quay crane 1 is seized by a malfunction event
	Malfunction QC2	Resource quay crane 2 is seized by a malfunction event
	Malfunction QC3	Resource quay crane 3 is seized by a malfunction event
	Good function QC1	Resource quay crane 1 can be used again
	Good function QC2	Resource quay crane 2 can be used again
	Good function QC3	Resource quay crane 3 can be used again
The required time to repair an equipment	Repair QC1	Maintainability (Mean Time To Repair)—the waiting time before repairing quay crane 1
	Repair QC1	Maintainability (Mean Time To Repair)—the waiting time before repairing quay crane 2
	Repair QC1	Maintainability (Mean Time To Repair)—the waiting time before repairing quay crane 3

The modeled system states and disruptive events generated by the failure of a container handling equipment on the quays with the help of the simulation model are drawn in Figure 2.

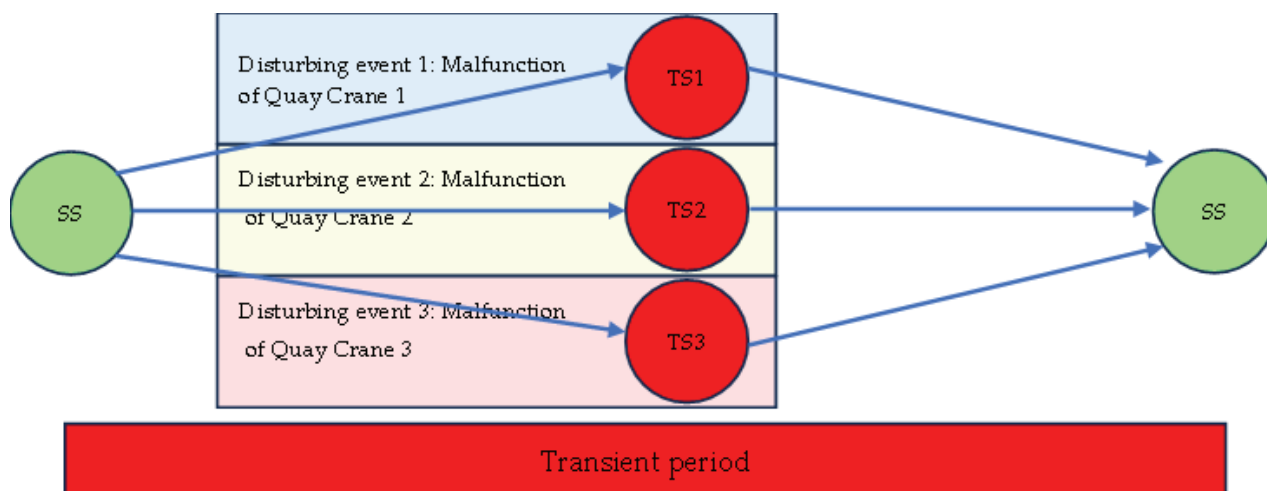


Figure 2. The system state (SS—stationary state, TS1—transient state 1, TS2—transient state 2, TS3—transient state 3)—source: authors.

5. Study Case of Constantza South Container Terminal

The Constantza South Container Terminal (CSCT) is the most significant container terminal in the European Union along the Black Sea. This maritime terminal is part of a global network operated by DP World. Its current capacity is approximately 1,200,000 TEU per year. Achieving and maintaining this throughput requires a precise evaluation of terminal flows and ensuring that the handling equipment has sufficient capacity to accommodate the operational demands. The most critical component of container handling—both for transferring containers from maritime vessels to land transport networks and for loading containers arriving from the port's hinterland onto ships—is represented by quay cranes (Figure 3).

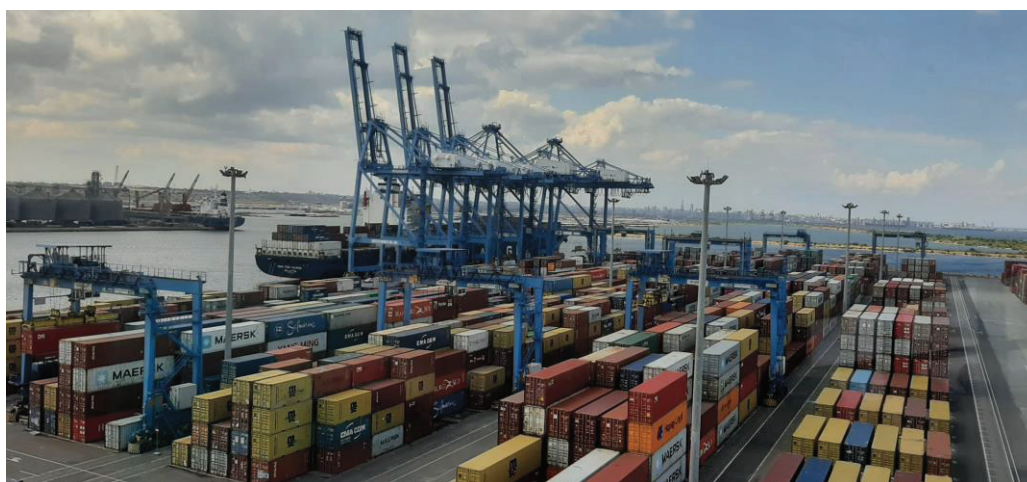


Figure 3. The quay cranes used inside CSCT—source: authors.

The occurrence of a disruptive event in quay crane operations results in alterations to the operating parameters of the entire port. Analyzing the impact of quay crane failures is of particular interest due to their significant economic implications for port activities.

The terminal is equipped with three types of quay cranes, which are arranged in groups—five positioned on one side of the terminal and one separately. The main characteristics of these cranes are detailed in Table 5. The number of cranes utilized typically ranges from one to three, depending on factors such as the vessel's specifications, crane availability, the number of containers, and their positioning on the vessel.

Table 5. Main characteristics of quay cranes [50].

Cranes	QGC Mitsubishi	QGC ZPMC	MHC Gottwald
Number	3	2	1
Lift capacity—under spreader single			
Lift/under spreader	40.6 To/n/a/	40.6 To/62.5 To/70	40.6 To/50
twin lift/under	50 To	To	To/100 To
hook beam			
Out reach/rows	45.00 m/16 rows	52.10 m/18 rows	50.00 m radius
across vessel	across	across	
Back reach	11 m	11 m	n/a
Spreader height above			
crane rail	34 m	36 m	n/a
Crane gauge	30.5 m	30.5 m	n/a

The simulation model developed in ARENA 12 is depicted in Figure 4. The simulation models the unloading and loading of containers onto ships using quay cranes. Within the terminal, these containers are further manipulated with the assistance of various types of equipment [4], including Rubber-Tired Gantries (RTG), Rail-Mounted Gantries (RMG), Empty Handlers (ECH), Reach Stackers (RS), Terminal Tractors (ITV), and Terminal Chassis (TC). However, these auxiliary equipment types were excluded from the simulation. Given their high numbers within the terminal, they provide sufficient reliability, leading the authors to deem their inclusion unnecessary for the scope of this study. The simulation model, developed using ARENA 12, is illustrated in Figure 4.

Data collection was conducted over the three summer months. As a result, the dataset does not capture periods of terminal shutdowns that could have been caused by extreme weather events. However, the absence of such disruptions does not significantly impact the simulation results, as maintenance activities would also be unfeasible during these periods with potential interruption due to the terminal's location within the maritime access zone of the port. The use of statistical tools in processing primary data is essential for developing simulation models [51]. The quality of the data input into the simulation model has a significant impact on the credibility of the results obtained. The time intervals between the loading and unloading processes of maritime vessels are represented in the graph from Figure 5.

To represent this flow within the simulation model, the normal distribution function was selected, where the parameter $\mu = 390$ min represents the mean (which is also the median and mode), and the parameter $\sigma^2 = 30$ min denotes the variance. The arrival of sea vessels follows a schedule established by the terminal administration, which considers the time required to complete container handling operations for the preceding vessel. To accurately reflect this schedule in the simulation, a time variable was integrated. The number of containers handled per vessel is modeled using a normal distribution, with parameters estimated based on the aforementioned variable.

The input data in simulation are presented in Table 6.

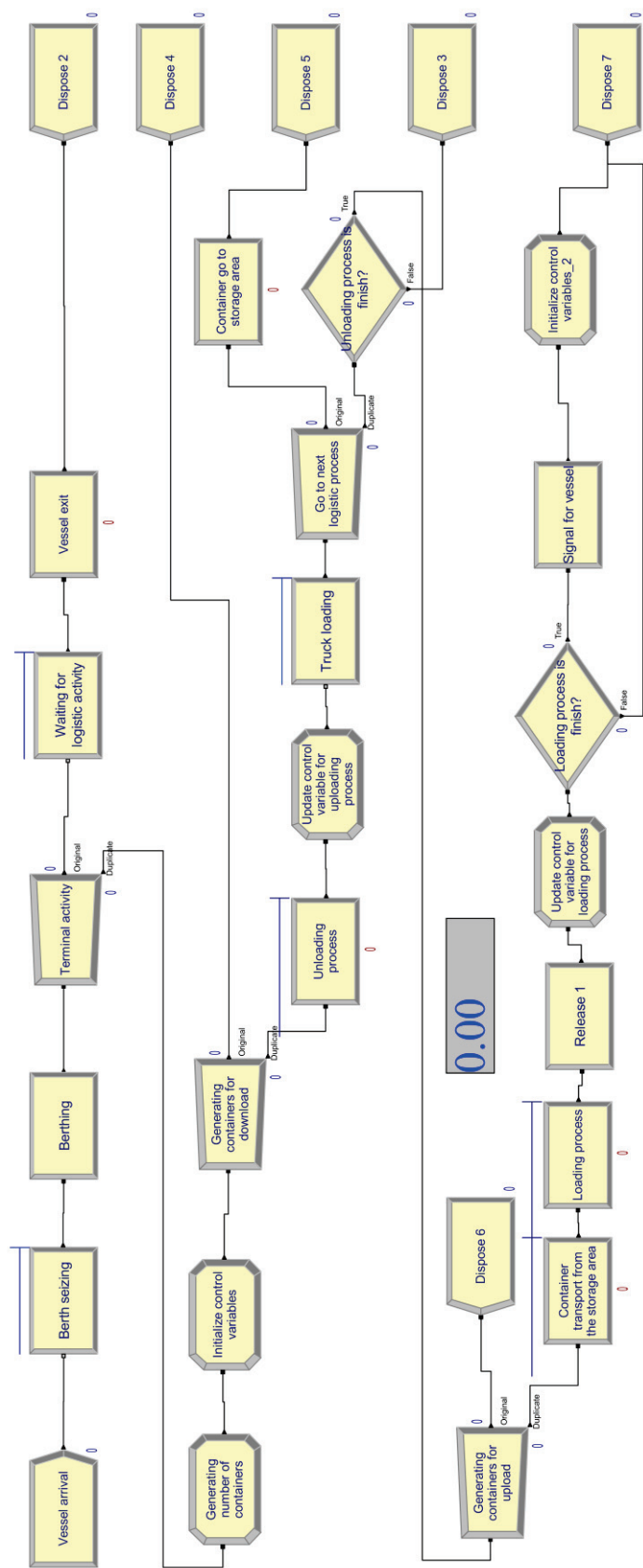


Figure 4. The simulation model—source: authors.

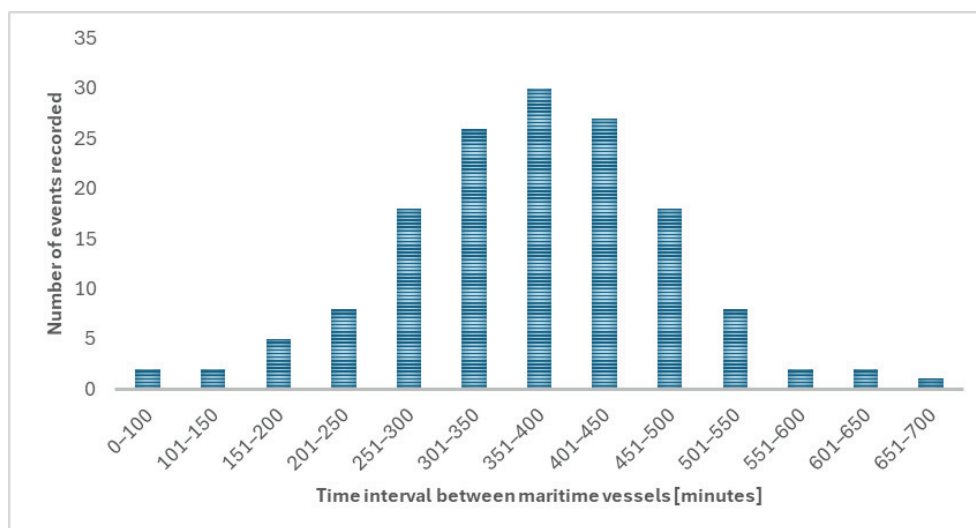


Figure 5. Time intervals between vessel arrivals.

Table 6. Main attributes of simulation input data.

Attribute	Value	Unit Measure
Number of quay cranes used for loading/unloading process at a single vessel	3	-
Number of trucks	20	-
Number of containers loaded/unloaded	Normal distribution	Containers
The required time interval for handling a container from/to maritime vessel	Normal distribution N (2, 0.1)	Minutes
The required time interval for handling a container from/to quay using a terminal's truck	Normal distribution N (10, 0.2)	Minutes
Berthing operation	3	Hours

To study the influence of equipment reliability over the activity inside the maritime container terminal, a series of scenarios were created. The simulation parameters for each scenario are presented in Table 7. The first scenario is a base situation with all three quay cranes working.

Table 7. The simulation scenario.

Scenario ID	Description
S00	All quay cranes are functional. The simulation duration is 365 days. The warm-up period during which no data are recorded is 5 days.
S11	Same as in S00. After 30 days, one quay crane is unavailable/broken for 24 h
S17	Same as in S00. After 30 days, one quay crane is unavailable/broken for 7 days
S114	Same as in S00. After 30 days, one quay crane is unavailable/broken for 14 days
S21	Same as in S00. After 30 days, two quay cranes are unavailable/broken for 24 h
S27	Same as in S00. After 30 days, two quay cranes are unavailable/broken for 7 days
S214	Same as in S00. After 30 days, two quay cranes are unavailable/broken for 14 days

6. Results and Discussions

The simulation model was used to obtain data in all seven scenarios. The stationarity of the ships' waiting time series for berth access is considered in analysis. For scenario S00, the ships' waiting time series and its density are depicted in Figure 6.

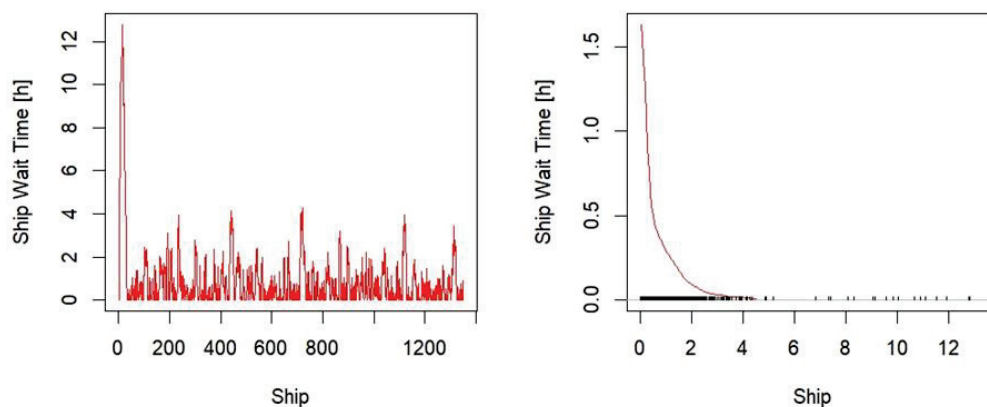


Figure 6. Ships' waiting time series (values and density)—scenario S00.

The Heidelberg–Welch test is passed ($p = 0.851$) and the ships' wait time series is stationary. The average wait time of the ships is 0.827 (h), and the quantiles are shown in Table 8.

Table 8. Ships' wait time quantiles.

Quantile	Wait Time (h)
2.5%	0
25%	0
50%	0.42
75%	1.11
97.5%	3.90

Comparing the first 10% of the series with the last half, the Geweke test statistics is 0.707, that confirms the stationarity. The plot of the z values ($n_{bins} = 50$) using the Geweke test for different fractions of the first half of the ships' waiting time series is presented in Figure 7. All the Z-scores move between ± 1.96 , so the differences between the mean waiting time in the first half bins and the mean of the last half of the series are not significant.

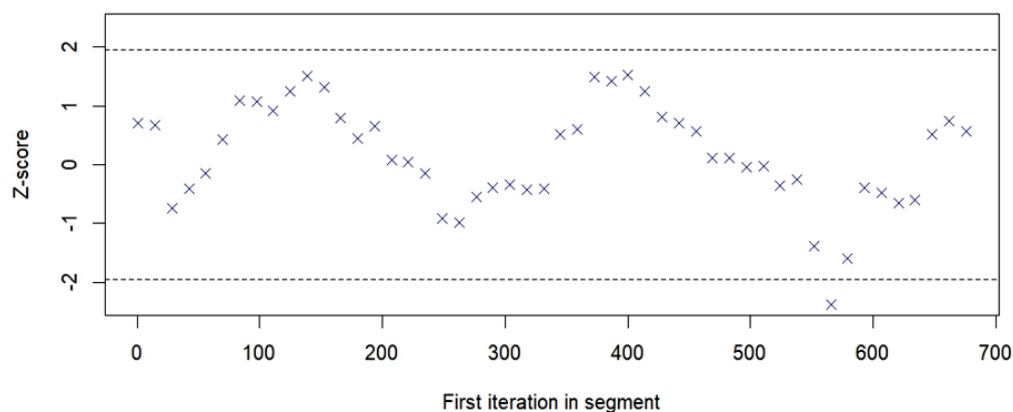


Figure 7. Z-scores in the first half of the ships' waiting time series (50 bins).

When the system experiences crane breakdowns, the ships' waiting time series following the moment of the breakdown are recorded and the stationarity of the series is tested.

For different numbers of cranes out of order and various durations of the breakdown, the results are depicted in Figures 8–13.

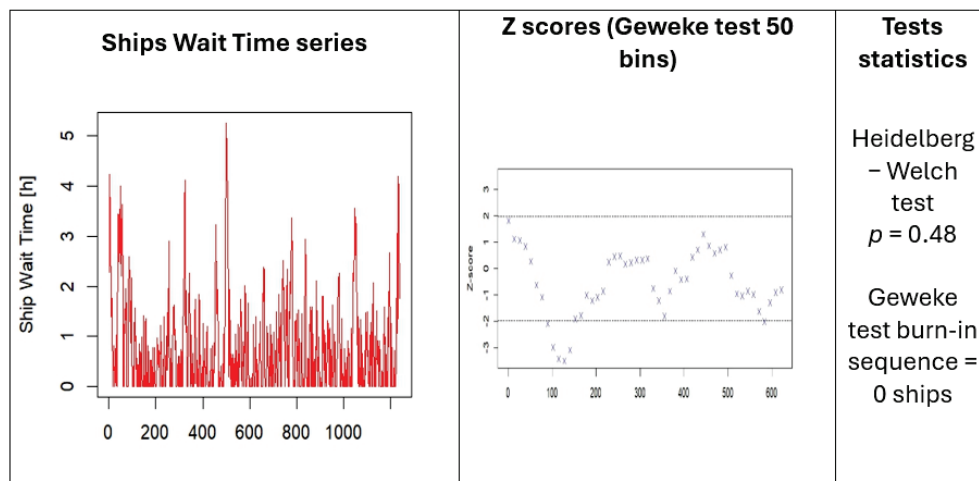


Figure 8. Ships' wait time series—scenario S11.

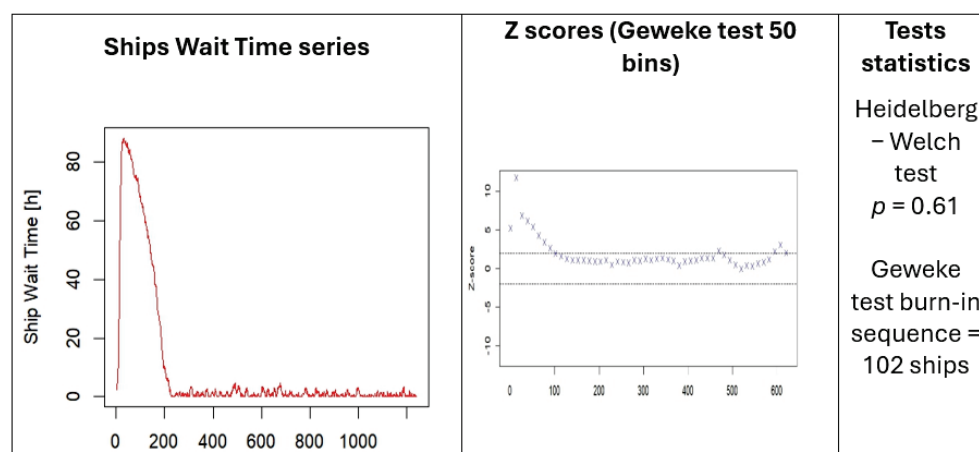


Figure 9. Ships' wait time series—scenario S17.

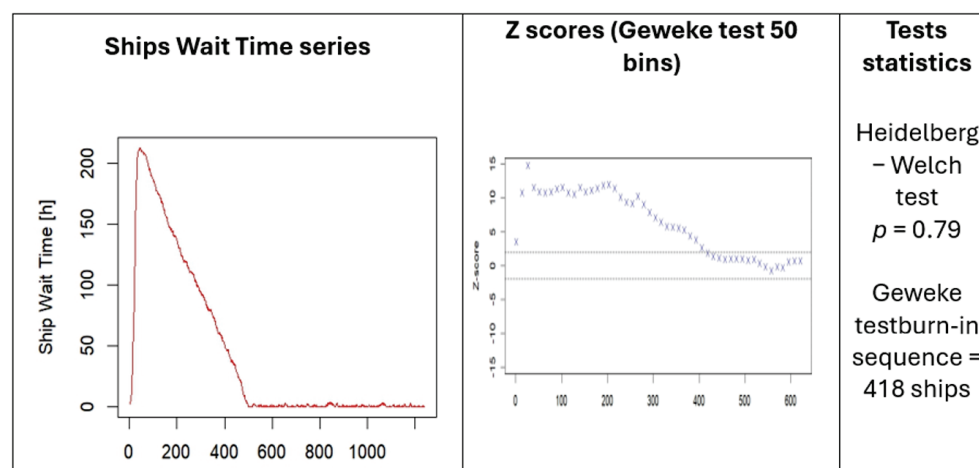


Figure 10. Ships' wait time series—scenario S114.

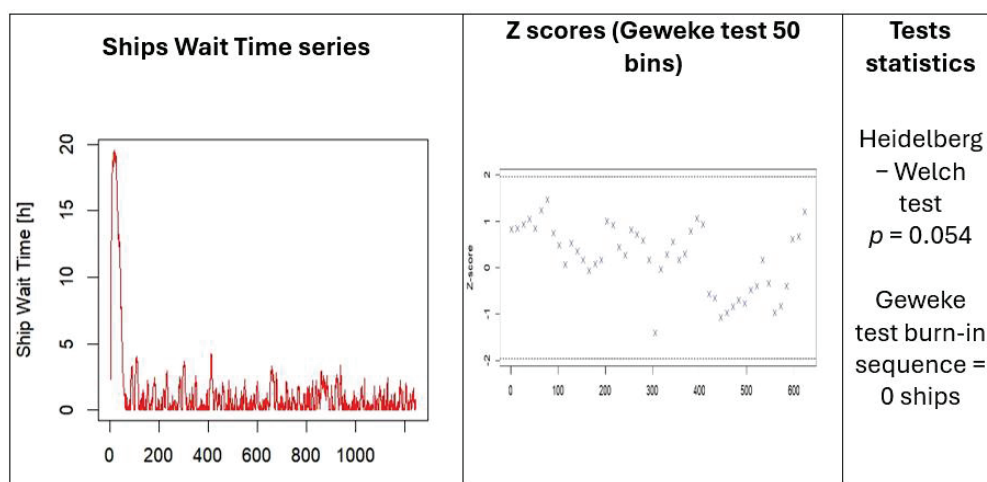


Figure 11. Ships' wait time series—scenario S21.

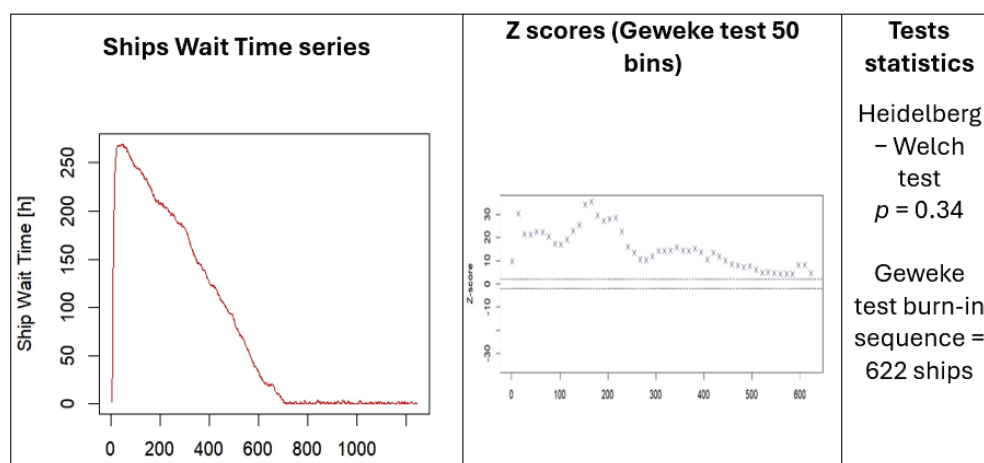


Figure 12. Ships' wait time series—scenario S27.

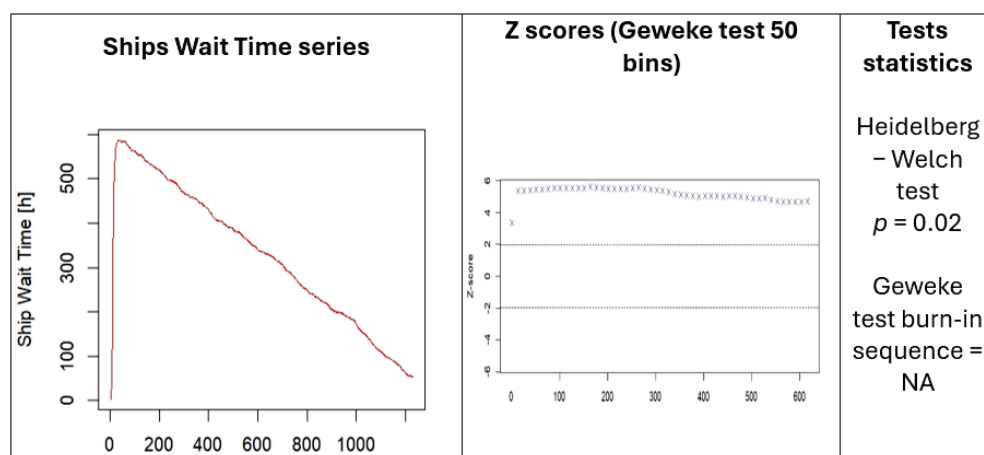


Figure 13. Ships' wait time series—scenario S214.

When one crane is out of order, the ships' waiting time series is going stationary, and the Heidelberg–Welch test provides $p > 0.05$ for scenarios S11, S17, and S114. The z-scores in the Geweke test follows the ± 1.96 interval after a burn-in sequence that depends on the breakdown duration. If the crane is stopped for 7 days, the stationarity is gained after a burn-in sequence of 102 ships, and after 418 ships when the crane is broken for 14 days.

If two cranes are out of order for one day (scenario S21), the ships' wait time series is stationary and the Geweke z-scores are in the limits of convergence. For 7 days of breakdown (scenario S27), the series becomes stationary (Heidelberg–Welch $p = 0.34$), but the Geweke z-scores tends to the convergence interval at the end of the series' first half (622 ships). If the interruption of cranes is 14 days, the series is not stationary (Heidelberg–Welch $p = 0.02$), and the z-scores are out of the convergence interval for the provided simulation length. A longer simulation run is needed to analyze the moment when the time series becomes stationary.

The findings highlight several key implications that container terminal administrators should consider enhancing operational efficiency and resilience:

- **Identification of Critical Equipment:** It is essential to identify the terminal's critical equipment, whose failure could severely disrupt operations and impact overall efficiency.
- **Significance of Maintenance Strategies:** Effective maintenance planning plays a crucial role in ensuring uninterrupted terminal operations, minimizing downtime, and preventing costly disruptions.
- **Mitigating the Impact of Unforeseen Events:** Unexpected disruptions can have significant consequences. Therefore, maintaining reserve capacity in the allocation of handling equipment is a strategic necessity to ensure operational continuity.
- **Leveraging Discrete-Event Simulation Models:** The use of discrete-event simulation enables cost-effective scenario testing, facilitating the identification of operational bottlenecks and critical vulnerabilities within terminal activities.

The conducted research presents a novel approach by combining discrete-event simulation with a mathematical model to evaluate the system's transition between states. In previous studies [26,27,30,33], simulation model results have been utilized in the strategic development process of the terminal. At the same time, in [39], the focus is on evaluating the resilience and dynamic optimization of terminal operations under the influence of disruptive events. However, the integration of discrete-event modeling with a mathematical model to determine system state transitions has not been previously explored. The results obtained in this study thus become a powerful tool for port administration in selecting the maintenance plan for critical equipment, specifically quay cranes.

7. Conclusions

The study of the impact of equipment failures in container terminal operations holds particular interest due to its significant economic implications. Even though the probability of failure for a critical piece of equipment—such as one or more quay cranes—is low, the impact of such an event can be substantial.

A simulation-based modeling approach to terminal operations enables the development of various scenarios for the evolution of activities at the desired level of details. This allow researchers to analyze system behavior and identify conditions under which operations may be obstructed or service parameters may deteriorate.

However, assessing the impact of critical equipment failures on operational performance remains relatively underexplored. In the specific case of maritime container terminals, quay cranes are complex pieces of equipment whose proper functioning is essential for terminal-wide operations. A failure of such equipment results in a reduced container throughput capacity, affecting the transfer of containers between maritime vessels and the hinterland, or vice versa. Whether the impact of such a failure can be absorbed by other quay cranes depends on various operational factors.

From a system state perspective, analyzing stable and transient states in maritime container terminals represents a novel approach. In conducting this analysis, the authors utilize a discrete-event simulation tool that they have developed and tailored to a real-

world scenario. This methodological framework underscores the originality of the research and its contribution to the field.

The simulation model should include only those processes and activities deemed relevant by the researcher for the specific study. In the case of a container terminal, quay cranes are functioning as service stations whose failure can either halt operations entirely or lead to significant delays within the system. When this occurs, the system transitions from a steady-state operation—corresponding to normal functionality—to a transient state, wherein qualitative and quantitative service parameters may deteriorate from their normal values. Statistical evaluation of these transitions can be conducted using the Heidelberg–Welch and Geweke tests, which assess whether the system converges to a steady-state condition.

The simulation model in this study was developed using the dedicated software tool ARENA 12. The model's structure allows for generalization to various container terminal topologies, whether maritime or inland. The level of detail in each individual simulation can lead to the omission of critical components of terminal operations or introduce evaluation errors by generating inconsistencies that are difficult to detect. In the specific case of this paper, the authors have identified the significant impact of quay crane operations on overall terminal performance. The quay cranes' handling capacity of containers both on the vessel and at the quay can act as a limiting factor for the entire maritime terminal's throughput. The authors acknowledge that this modeling approach is suitable only for achieving the objectives set in the research presented in the manuscript. If the simulation is to be used in the design phase or for defining technological processes, the model must be adapted to the specific process under analysis. Consequently, the level of detail should be selectively increased in the analyzed area while being reduced for processes that have minimal impact on the simulation's objectives. The methodology used to evaluate the system's transition from one state to another is highly generalized. It can be adapted for container terminals in different geographical regions or applied to various operational scenarios. The specific contribution of this research, as presented in this article, lies in the use of a simulation model as an analytical tool. However, this model can be adjusted to accommodate different terminal typologies. Additionally, the model's resource parameters and the time durations associated with technological processes must be recalibrated for each specific case. Nevertheless, as previously mentioned, once the simulation model is adapted, the analytical methodology outlined in Section 3.2 can be easily applied.

For the specific case study, seven potential scenarios were considered. These scenarios were designed to test the validity of the research model in a real container terminal setting. The results obtained highlighted multiple critical situations in which the system transitioned from a steady-state condition (represented by scenario S00) to transient states. Statistical tests were performed on these states, identifying scenarios in which the system returned to steady-state conditions and those in which the level of disruption prevented the system from stabilizing.

The results obtained from the six scenarios, characterized by the occurrence of a disruptive event—namely, the failure of one or two quay cranes for a period ranging from one to fourteen days—demonstrated the impact of these events on the state of the system. Specifically, for the first three scenarios, where only one crane is out of order, the ships' waiting time series tends to reach stationarity, as indicated by the Heidelberg–Welch test with $p > 0.05$. The findings show a return to stationarity after 102 and 480 ships in cases where the downtime of a quay crane lasts 7 days and 14 days, respectively.

In contrast, when two quay cranes become non-operational for 7 days, the system returns to a stationary state after 622 ships. These results highlight that both the amplitude

and duration of the disruptive event have a direct influence on the transition between system states.

The simulation model served as a tool for evaluating how a disruptive event leads the system from a stable state to a transient state. The construction of simulation scenarios was hypothetical but close to reality, designed to facilitate the testing of the reliability assessment methodology for handling equipment, as outlined in the article.

Considering these aspects, a series of limitations were identified separately for the simulation model and the mathematical model throughout the research process.

Identified limitations of the simulation model are:

- Due to the reduced level of detail in the simulation model, the interaction between human factors and technical equipment was evaluated using average values. Operators of handling equipment in the container terminal have varying levels of experience, which may lead to increased or decreased productivity depending on the individual operating the equipment. To address this limitation, average values provided by the terminal administration were used. However, in future research, these activities will be detailed according to the simulation's objectives.
- The reliability function could not be evaluated for secondary components of the simulation model, such as trucks used for container transport within the terminal. Due to their relatively high number and availability, as observed during on-site visits, this limitation was addressed by considering that these do not represent a restrictive component of the terminal's handling and transit capacity. Future research will test the influence of these pieces of equipment to enhance the model's sensitivity.
- Data collection during the summer period did not allow for the identification of the environmental impact on terminal operations. This limitation will be addressed in future research, where data recordings will facilitate the development of a predictive model for disruptions in the activity of the terminal caused by environmental hazards.

Identified limitations of the results analysis:

- To obtain credible and realistic results, the model requires input data from the simulation model that captures transitions between states. To overcome this limitation, the simulation duration was extended to 365 days, which was more than sufficient for obtaining the necessary data. Only for scenario S214 the simulation time horizon was not sufficient to test the stationarity convergence.

In the opinion of the authors, the most important limitation identified was in relation to environmental hazards. Due to regional-specific factors, maritime terminal operations are affected by adverse weather conditions such as storms, hurricanes, and strong winds. The authors recognize the significance of this aspect and for this reason are currently collecting data on terminal operations and the influence of meteorological conditions. The data will be incorporated into a future study once the primary data collection process is complete. Future research aims to incorporate these disruptive events into the vulnerability analysis of container-handling equipment failures.

Author Contributions: Conceptualization, F.R. and E.R.; methodology, E.R. and V.C.; software, F.R. and E.R.; validation, A.R. and O.D.; formal analysis, A.R., V.C. and O.D.; resources, O.S.; data curation, O.S. and O.D.; writing—original draft preparation, F.R. and V.C.; writing—review and editing, E.R.; supervision, A.R. and O.D.; project administration, A.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National University of Science and Technology Politehnica Bucharest, grant number 70/11.10.2023.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This work was supported by a grant from the National Program for Research of the National Association of Technical Universities—GNAC ARUT 2023.

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

References

1. United Nations Conference on Trade and Development. Available online: https://unctad.org/system/files/official-document/rmt2023_en.pdf (accessed on 5 April 2024).
2. Meisel, F. *Seaside Operations Planning in Container Terminals*; Physica: Heidelberg, Germany, 2009; ISBN 978-3-7908-2586-2.
3. Svanberg, M.; Holm, H.; Cullinane, K. Assessing the Impact of Disruptive Events on Port Performance and Choice: The Case of Gothenburg. *J. Mar. Sci. Eng.* **2021**, *9*, 145. [CrossRef]
4. Steenken, D.; Voß, S.; Stahlbock, R. *Container Terminals and Automated Transport Systems*; Springer: Berlin, Germany, 2005; pp. 3–49.
5. Ambrosino, D.; Sciomachen, A. Impact of Externalities on the Design and Management of Multimodal Logistic Networks. *Sustainability* **2021**, *13*, 5080. [CrossRef]
6. Wang, C.-N.; Nguyen, N.-A.-T.; Fu, H.-P.; Hsu, H.-P.; Dang, T.-T. Efficiency Assessment of Seaport Terminal Operators Using DEA Malmquist and Epsilon-Based Measure Models. *Axioms* **2021**, *10*, 48. [CrossRef]
7. Pujats, K.; Golias, M.; Konur, D. A Review of Game Theory Applications for Seaport Cooperation and Competition. *J. Mar. Sci. Eng.* **2020**, *8*, 100. [CrossRef]
8. Romero, V.M.; Fernandez, E.B. Towards a Reference Architecture for Cargo Ports. *Future Internet* **2023**, *15*, 139. [CrossRef]
9. Tijan, E.; Jović, M.; Žgaljić, D.; Aksentijević, S. Factors Affecting Container Seaport Competitiveness: Case Study on Port of Rijeka. *J. Mar. Sci. Eng.* **2022**, *10*, 1346. [CrossRef]
10. Sáenz, S.S.; Diaz-Hernandez, G.; Schweter, L.; Nordbeck, P. Analysis of the Mooring Effects of Future Ultra-Large Container Vessels (ULCV) on Port Infrastructures. *J. Mar. Sci. Eng.* **2023**, *11*, 856. [CrossRef]
11. Sugimura, Y.; Wakashima, H.; Liang, Z.; Shibasaki, R. Logistics strategy simulation of second-ranked ports on the basis of Japan's port reforms: A case study of Hakata Port. *Marit. Policy Manag.* **2022**, *50*, 707–723. [CrossRef]
12. Ribeiro da Silva, J.N.; Santos, T.A.; Teixeira, A.P. Methodology for Predicting Maritime Traffic Ship Emissions Using Automatic Identification System Data. *J. Mar. Sci. Eng.* **2024**, *12*, 320. [CrossRef]
13. Surugiu, M.C.; Gheorghiu, R.A.; Petrescu, I. Transmission of pollution data to traffic management systems using mobile sensors. *U.P.B. Sci. Bull. Ser. C* **2015**, *77*, 163–170.
14. Saracin, C.G.; Belibov, D. Real time monitoring of analog and digital sensors. *U.P.B. Sci. Bull. Ser. C* **2018**, *80*, 35–44.
15. Kim, S.-W.; Eom, J.-O. Ship Carbon Intensity Indicator Assessment via Just-in-Time Arrival Algorithm Based on Real-Time Data: Case Study of Pusan New International Port. *Sustainability* **2023**, *15*, 13875. [CrossRef]
16. Díaz-Ruiz-Navamuel, E.; Ortega Piris, A.; López-Díaz, A.-I.; Gutiérrez, M.A.; Roiz, M.A.; Chaveli, J.M.O. Influence of Ships Docking System in the Reduction of CO₂ Emissions in Container Ports. *Sustainability* **2021**, *13*, 5051. [CrossRef]
17. Clemente, D.; Cabral, T.; Rosa-Santos, P.; Taveira-Pinto, F. Blue Seaports: The Smart, Sustainable and Electrified Ports of the Future. *Smart Cities* **2023**, *6*, 1560–1588. [CrossRef]
18. Tran, N.K.; Lam, J.S.L. Effects of container ship speed on CO₂ emission, cargo lead time and supply chain costs. *Res. Transp. Bus. Manag.* **2022**, *43*, 100723. [CrossRef]
19. Li, J.; Jing, K.; Khimich, M.; Shen, L. Optimization of Green Containerized Grain Supply Chain Transportation Problem in Ukraine Considering Disruption Scenarios. *Sustainability* **2023**, *15*, 7620. [CrossRef]
20. Yıldırım, M.S. Investigating the Impact of Buffer Stacks with Truck Restriction Time Window Policy on Reducing Congestion and Emissions at Port of Izmir. *Int. J. Civ. Eng.* **2023**, *21*, 1107–1122. [CrossRef]
21. Altomonte, A.; Bracale, A.; Caramia, P.; De Falco, P.; Dillio, G.; Di Noia, L.P. Time-Domain Modeling and Simulation of a Fuel Cell Hybrid Truck Powertrain Operating in Port Logistics. In Proceedings of the 2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT), Male, Maldives, 11–12 March 2023; pp. 1–7. [CrossRef]
22. Hervás-Peralta, M.; Poveda-Reyes, S.; Molero, G.D.; Santarremigia, F.E.; Pastor-Ferrando, J.-P. Improving the Performance of Dry and Maritime Ports by Increasing Knowledge about the Most Relevant Functionalities of the Terminal Operating System (TOS). *Sustainability* **2019**, *11*, 1648. [CrossRef]
23. Abu Aisha, T.; Ouhimmou, M.; Paquet, M. Optimization of Container Terminal Layouts in the Seaport—Case of Port of Montreal. *Sustainability* **2020**, *12*, 1165. [CrossRef]
24. Guo, S.; Diao, C.; Li, G.; Takahashi, K. The Two-Echelon Dual-Channel Models for the Intermodal Container Terminals of the China Railway Express Considering Container Accumulation Modes. *Sustainability* **2021**, *13*, 2806. [CrossRef]

25. Qu, W.; Tao, T.; Xie, B.; Qi, Y. A State-Dependent Approximation Method for Estimating Truck Queue Length at Marine Terminals. *Sustainability* **2021**, *13*, 2917. [CrossRef]
26. Rusca, F.; Popa, M.; Rosca, E.; Rusca, A. Simulation model for maritime container terminal. *Transp. Probl.* **2018**, *13*, 47–54. [CrossRef]
27. Zhang, L. Research on Simulation Modeling of Container Terminal Logistics System Based on WITNESS. In Proceedings of the International Conference on Industrial IoT, Big Data and Supply Chain (IIoTBDSC), Beijing, China, 23–25 September 2022; pp. 173–178. [CrossRef]
28. Rotunno, G.; Lo Zupone, G.; Carnimeo, L.; Fanti, M.P. Discrete event simulation as a decision tool: A cost benefit analysis case study. *J. Simul.* **2023**, *18*, 378–394. [CrossRef]
29. Wang, R.; Li, J.; Bai, R. Prediction and Analysis of Container Terminal Logistics Arrival Time Based on Simulation Interactive Modeling: A Case Study of Ningbo Port. *Mathematics* **2023**, *11*, 3271. [CrossRef]
30. Siswanto, N.; Kusumawati, M.; Baihaqy, A.R.; Wiratno, S.E.; Sarker, R. New Approach for Evaluating Berth Allocation Procedures Using Discrete Event Simulation to Reduce Total Port Handling Costs. *Oper. Supply Chain Manag. Int. J.* **2023**, *16*, 535–553. [CrossRef]
31. Torkjazi, M.; Huynh, N.; Asadabadi, A. Modeling the Truck Appointment System as a Multi-Player Game. *Logistics* **2022**, *6*, 53. [CrossRef]
32. Muñuzuri, J.; Lorenzo-Espejo, A.; Pegado-Bardayo, A.; Escudero-Santana, A. Integrated Scheduling of Vessels, Cranes and Trains to Minimize Delays in a Seaport Container Terminal. *J. Mar. Sci. Eng.* **2022**, *10*, 1506. [CrossRef]
33. Rusca, F.; Popa, M.; Rosca, E.; Rusca, A.; Rosca, M.; Dinu, O. Assessing the transit capacity of port shunting yards through discrete simulation. *Transp. Probl.* **2019**, *14*, 101–111. [CrossRef]
34. Aslam, S.; Michaelides, M.P.; Herodotou, H. Berth Allocation Considering Multiple Quays: A Practical Approach Using Cuckoo Search Optimization. *J. Mar. Sci. Eng.* **2023**, *11*, 1280. [CrossRef]
35. Gulić, M.; Maglić, L.; Krljan, T.; Maglić, L. Solving the Container Relocation Problem by Using a Metaheuristic Genetic Algorithm. *Appl. Sci.* **2022**, *12*, 7397. [CrossRef]
36. Fedorko, G.; Molnár, V.; Mikušová, N.; Strohmandl, J.; Kižik, T. Simulation of Handling Operations in Marine Container Terminals for the Purposes of a Profession Simulator. *J. Mar. Sci. Eng.* **2023**, *11*, 2264. [CrossRef]
37. Doci, I.; Baxhuku, V. Forklift operations in the warehouse—Design and analysis with graphical modelling and simulations. *U.P.B. Sci. Bull. Ser. D* **2024**, *86*, 175–190.
38. Budiyo, E.H.; Gurning, R.O.S.; Pitana, T. The Application of Business Impact Analysis Due to Electricity Disruption in a Container Terminal. *Sustainability* **2021**, *13*, 12038. [CrossRef]
39. Xu, B.W.; Liu, W.T.; Li, J.J.; Yang, Y.S.; Wen, F.R.; Song, H.T. Resilience measurement and dynamic optimization of container logistics supply chain under adverse events. *Comput. Ind. Eng.* **2023**, *180*, 109202. [CrossRef]
40. Târcolea, C.; Filipoiu, A.; Bontaş, S. *Current Techniques in Reliability Theory-Current Applications in Reliability Theory*; Ştiinţifică şi Enciclopedică: Bucharest, Romania, 1989; pp. 87–137.
41. Pawlikowski, K. Steady-state simulation of queueing processes: Survey of problems and solutions. *ACM Comput. Surv.* **1990**, *22*, 123–170. [CrossRef]
42. Mahajan, P.S.; Ingalls, R.G. Evaluation of methods used to detect warm-up period in steady state simulation. In Proceedings of the 2004 Winter Simulation Conference, Washington, DC, USA, 5–8 December 2004; p. 671. [CrossRef]
43. Schruben, L.W. Detecting Initialization Bias in Simulation Experiments. *Oper. Res.* **1982**, *30*, 569–590. [CrossRef]
44. Schruben, L.W.; Singh, H.; Tierney, L. Optimal Tests for Initialization Bias in Simulation Output. *Oper. Res.* **1983**, *31*, 1167–1178. [CrossRef]
45. Heidelberger, P.; Welch, P.D. Simulation run length control in the presence of an initial transient. *Oper. Res.* **1983**, *31*, 1109–1144. [CrossRef]
46. Geweke, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*; Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., Eds.; Clarendon Press: Oxford, UK, 1992; pp. 169–193.
47. Rusca, F.; Rusca, A.; Rosca, E.; Coman, C.; Burciu, S.; Oprea, C. Evaluating the Influence of Data Entropy in the Use of a Smart Equipment for Traffic Management at Border Check Point. *Machines* **2022**, *10*, 937. [CrossRef]
48. Rusca, F.; Rosca, E.; Costescu, D.; Rusca, A.; Rosca, M.; Olteanu, S. The influence of storage area design on maritime container terminal capacity. In Proceedings of the Modtech International Conference—Modern Technologies in Industrial Engineering Vi (Modtech 2018), Constanta, Romania, 13–16 June 2018.
49. Rosca, E.; Rusca, F.; Rosca, M.A.; Rusca, A. Performance Analysis of Automated Parcel Lockers in Urban Delivery: Combined Agent-Based–Monte Carlo Simulation Approach. *Logistics* **2024**, *8*, 61. [CrossRef]

50. DPWorld Romania-Service-Terminal Equipment. Available online: <https://www.dpworld.com/romania/services/terminal-equipment> (accessed on 19 August 2024).
51. Wawrzyniak, J.; Drozdowski, M.; Sanlaville, É. A Container Ship Traffic Model for Simulation Studies. *Int. J. Appl. Math. Comput. Sci. Sciendo* **2022**, *32*, 537–552. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Riding Risk: Factors Shaping Helmet Use Among Two-Wheeled Electric Vehicle Riders in Fuzhou, China

Wenhan Lin ¹, Congying Li ², Weibin Zheng ³, Linwei Wang ¹ and Yanqun Yang ^{1,*}

¹ College of Civil Engineering, Fuzhou University, Fuzhou 350116, China

² Guangxi Communications Design Group Co., Ltd., Nanning 530022, China; licongying2024@163.com

³ Fuzhou Fuda Automation Technology Co., Ltd., Fuzhou 350028, China

* Correspondence: yangyanqun@fzu.edu.cn

Abstract: With the rapid increase in the number of two-wheeled electric vehicles, the number of accidents related to them has also greatly increased. However, despite facing a huge threat from accidents, the helmet, an efficient and legally required protection for riders, is not popular with Chinese two-wheeled electric vehicles riders. To study the factors affecting helmet use for these riders, this paper conducted an observational study to collect helmet use data for 16,207 two-wheeled electric vehicle riders in Fuzhou, China. With these data, this paper built a multivariate logistic regression model to study the main effects of various factors on helmet use, and analyze the interaction effects of these factors. Results showed that, on the one hand, area, weather, temperature, controller, separated non-motor-vehicle lanes, time, rider's age, and type of vehicle had significant effects on helmet use and the interaction between these factors is significant, especially the interaction between weather, temperature and other factors. On the other hand, level of service, gender and whether the riders are food delivery workers have no significant impact on helmet use, but show significant interaction effects with other factors.

Keywords: helmet use; multivariate logistic regression model; two-wheeled electric vehicles riders; observational study

1. Introduction

The widespread use of two-wheeled electric vehicles (TWEVs) has become a global phenomenon, with the total number exceeding 300 million worldwide [1]. This surge is attributed to TWEVs' affordability, convenience, and environmental benefits. However, the increase in TWEV usage has been accompanied by a corresponding rise in traffic-related accidents, significantly impacting public safety [2]. Globally, traffic crashes involving TWEVs are a major concern, with TWEV riders facing disproportionately higher risks compared to users of other vehicle types [3]. Electric bicycle accidents are more severe than traditional bicycle accidents due to higher speeds and greater momentum. This heightened risk presents new challenges for road safety, requiring updated measures to protect riders and pedestrians alike [4].

Effective interventions, including red-light cameras, wearing helmets, seat belts, and addressing safety incidents, can reduce road traffic accidents by 25–40% [5]. Helmet use, a critical factor in reducing the severity of injuries, has been identified as a primary means of mitigating these risk [6]. Studies have consistently demonstrated that helmets reduce the risk of severe head injuries by 69% and fatalities by 42% in the event of a crash [7,8]. Despite this evidence, helmet usage among TWEV riders remains suboptimal. However, progress in helmet use is not satisfactory, although many countries have enacted legislation

to force riders to wear helmets when riding [9]. A large number of studies have shown that riding without wearing a helmet is still common, which is proved by observational studies, like that of Felix et al. [10] in Myanmar, Várhegyi et al. [11] and Rusli R et al. [12] in Malaysia, and Xuequn [13] in China. Due to a significant number of motorcyclists not wearing helmets, ASEAN countries have implemented stricter enforcement of mandatory helmet laws for two-wheeler riders [14].

Understanding the reasons for this reluctance is essential to improving safety. Research indicates that helmet use is influenced by both subjective and objective factors. Subjective factors include individual attributes, such as education level, personal attitudes, and risk-taking behaviors [15–17]. However, these factors are challenging to address through short-term interventions. Consequently, researchers have turned to objective factors, such as secondary tasks during riding [18], law enforcement intensity [19], and demographic characteristics [13]. Despite this shift, there remains a lack of clarity regarding how these objective factors interact to influence helmet use, particularly among TWEV riders, whose behaviors may differ significantly from those of motorcyclists.

A review of existing studies reveals significant methodological gaps. Data collection often relies on direct observation or questionnaires, with analysis typically conducted using logistic or probit models to identify the relationships between variables and helmet use [20–22]. However, most studies have focused on isolated factors, neglecting the potential interaction effects that could provide deeper insights into rider behaviors. For example, while enforcement efforts are known to increase compliance, their interaction with other factors, such as rider demographics or traffic density, has not been adequately explored. Studies have shown, that for every 1% increase in police patrol time, the frequency of accidents during the same period decreases by 0.15% [23].

In response to these gaps, this study conducted a detailed observational analysis of TWEV riders in Fuzhou, China. A total of 16,206 TWEV riders were observed across eight locations, varying by traffic conditions, enforcement levels, and demographic composition. Using a multivariate logistic regression model, we analyzed the effects of factors such as gender, age, traffic density, and enforcement presence on helmet use, along with the interactions between these variables. This approach aims to provide policymakers and professionals with actionable insights into the factors driving helmet use and to support the development of targeted interventions. Ultimately, the findings are expected to enhance the effectiveness of helmet regulations and improve TWEV safety management.

2. Literature Review

2.1. Review of Research on TWEV Safety

In the field of traffic safety, research related to TWEV safety focuses mainly on the investigation of riding behavior and the analysis of crash data. Wang et al. conducted a questionnaire survey of TWEV riders in Guilin City and used structural equation modeling to analyze the relationship between safety knowledge, safety attitudes, and risk perception and risky driving behaviors, and used multiple regression modeling to test the effect of safety knowledge on various types of risky driving behaviors. The results showed that safety knowledge was significantly associated with risky driving behaviors in Chinese TWEV riders [24]. Chai et al. compared the behaviors of Chinese e-bike riders and bicycle riders at unsignalized intersections. They found that e-bike riders were less sensitive to inclement weather and intersection risk, but more sensitive to infrastructure quality [25]. Qian et al. developed a two-wheeled vehicle rider behavior questionnaire and conducted a survey in the city of Xi'an, finding that the characteristics of the abnormal behaviors of TWEV riders were more similar to those of motorcycle riders. They both showed a higher frequency of abnormal riding behavior compared to bicycle riders [26]. Gioldasis et al.

conducted a face-to-face road survey of TWEV riders to explore the relationship between risk factors and user attributes and travel behavior [27].

In addition to surveys of TWEV riding behaviors, a number of scholars have analyzed crash data. Shah et al. used crash data from the Tennessee Integrated Transportation Analytics Network (TITAN) database from April 2018 to April 2020 and analyzed this using the Pedestrian and Bicycle Crash Analysis Tool (PBCAT), which showed that every 10 e-scooter and bicycle motorized vehicle collisions resulted in injury or death to a TWEV rider or bicyclist. In addition, there were statistically significant differences between TWEV and bicycle collisions in terms of spatial and temporal distribution, demographics, lighting conditions, and distance from home [28]. Sadeghi et al. combined the XGBoost algorithm with a stochastic parametric binary logit model and used likelihood ratio tests to validate the temporal instability of motorized vehicle collisions in the UK over different periods of the COVID epidemic [29]. Gao et al. categorized EV crashes into two-vehicle crashes and single-vehicle crashes, and used a random parameter approach with mean and variance heterogeneity to examine the factors affecting injury severity [30]. In addition, some studies have explored the factors affecting the frequency of TWEV crashes from a macroscopic perspective, combining crash data and environmental characteristics using crash counting models [31,32].

2.2. Review of Research on Helmets

Numerous studies have shown that failure to properly use seat belts and helmets is one of the leading causes of road traffic accidents [33,34]. Wearing a helmet can be effective in reducing the risk of head injuries [35]. Thompson et al. investigated the effectiveness of bicycle helmets in preventing and protecting against head injuries. The results showed that bicycle helmets, regardless of type, can provide protection for cyclists of all ages involved in collisions, including those involving motor vehicles [36]. The World Health Organization (WHO) produced the World Report on Road Traffic Injury Prevention [37]. It was found that proper wearing of motorcycle helmets reduced the risk of traffic fatalities by 40% and the risk of severe head injuries by 70%. Moftakhar et al. conducted a retrospective multicenter study of all consecutive patients admitted to hospitals for TWEV-related injuries (riders and non-riders) between 1 May 2018 and 30 September 2019, and the results showed that 17.7% had serious head injuries. Therefore, proper helmet wear may reduce the chance of injury in TWEV accidents [38]. Hamzani used a retrospective cross-sectional study design. All patients, including those referred to the emergency departments of tertiary care centers for oral and maxillofacial injuries involving TWEVs or P-scooters, from 2014–2020 were investigated. Results showed that patients who used helmets also had lower rates of fractures (18.2%) and alveolar injuries (23.7%) than those who did not (68.8% and 37.3%, respectively) [39].

However, despite the protective advantages of helmets, helmet use among e-bike riders is relatively low [19]. Some riders have low willingness to wear helmets because of their heaviness and inconvenience [40]. Yuan et al. found that helmet wearing among e-bike riders was influenced by area of residence, age, gender, marital status, and education level [41]. Yang et al. developed an integrated theoretical model combining the Theory of Planned Behavior (TPB) and the Health Belief Model (HBM) to test the influence of factors among college students on the willingness of TWEV riders to wear helmets, and the results showed that law enforcement emerged as the most influential factor, highlighting the key role of enforcing regulations and raising awareness [42]. Ma et al. used logistic regression to study the impact of helmet wearing regulations on TWEV riders, and the results showed that mandatory regulation was more effective than encouraging regulation in reducing head injuries [43].

This paper takes Fuzhou city as an example and analyzes the factors affecting the influence of helmet wearing on TWEV riders through video data. It provides a theoretical basis for the safety management of electric vehicles in Fuzhou city.

3. Materials and Methods

3.1. Site Characteristics

Four intersections and four road sections in Fuzhou, China were selected for investigation. The distribution of observation sites is shown in Figure 1, and their characteristics are shown in Table 1. The choice of sites was based on the data provided by the traffic police and the distribution of TWEV traffic volume. The observation was carried out through monitoring by traffic police. The view is shown in Figure 2.



Figure 1. Distribution of observation sites.

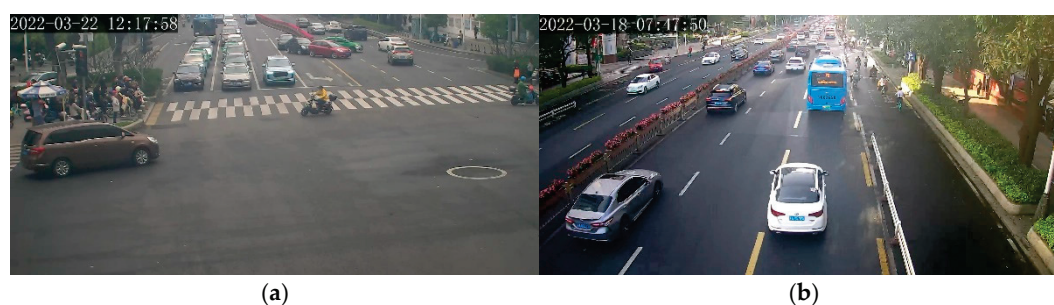


Figure 2. Views observed by surveillance video: (a) intersection, (b) section of road/avenue.

Table 1. Observation site characteristics.

Name	Mark in the Figure	Type	Area	Separate Non-Motor Vehicle Lane
Jiangfu Road-Bincheng Avenue intersection	triangle-1	intersection	suburb	nothing
Jianping Road-Qishan Avenue intersection	triangle-2	intersection	suburb	yes
Wusi Road-Hualin Road intersection	triangle-3	intersection	urban	yes
Cuping road-Hudong Road intersection	triangle-4	intersection	urban	nothing
Bincheng Avenue section	circle-1	road section	suburb	nothing

Table 1. Cont.

Name	Mark in the Figure	Type	Area	Separate Non-Motor Vehicle Lane
Qishan Avenue section	circle-2	road section	suburb	yes
Wusi Road section	circle-3	road section	urban	yes
Guping road section	circle-4	road section	urban	nothing

3.2. Data Collection Procedure

The observational data came from the surveillance video of the observation site from 11 March to 10 April 2022, from which a sunny day and a rainy day were randomly selected as the survey dates. The survey time was from 7:30 am to 8:00 am (morning peak hours), from 12:00 to 12:30 pm (off peak hours) and from 17:30 pm to 17:30 pm (evening peak hours). Each site provided a total of 3 h of surveillance video, and the total amount of video material observed was 24 h.

All observation data were coded by 15 master's degree students majoring in traffic and transportation. The data were collected in a pre-set table. All coders had been trained in different types of TWEV identification, rider characteristics identification and data quality control. The coding helmet use data was reviewed by three experts in the field of traffic engineering and traffic safety to ensure data quality.

4. Data Analysis

4.1. Descriptive Statistics

A total of 16,206 TWEV riders were observed, including 5375 riders not wearing helmets and 10,831 riders wearing helmets. The observed helmet use data at each site is shown in Figure 3. The number on the right side of the bar chart refers to the number of riders observed, and the x-axis as a percentage refers to the proportion of riders without helmets compared to all riders.

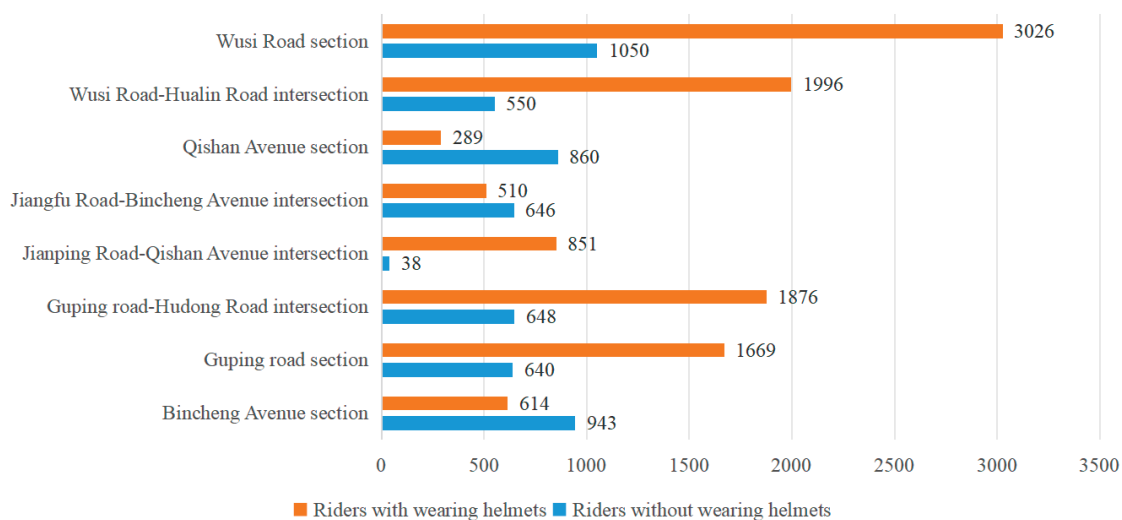


Figure 3. Statistics on the number and proportion of riders riding without helmets.

4.2. Multivariate Logistic Regression Model

The multivariate logit model is applicable in cases where the result variable is a category variable (with or without helmet), and multiple explained variables are classified variables or sequence variables. Therefore, a multivariate logistic regression model was

built to analyze the relationship between helmet use and other factors. The formula for the model is shown in Equation (1) below:

$$\text{logit}[P(y = j|X)] = \log\left(\frac{P(y = j|X)}{1 - P(y = j|X)}\right) = \alpha_j - x'_i\beta_j, \quad (1)$$

where j is the level of the result variable and i is the number of variables to be explained, $i \in [1, I], j \in [1, J]$. y is the result variable. x'_i is the change value of the i -th explained variable. β_j is the slope of the i -th explained variable. α_j is the constant term of the result variable corresponding to the j -th level.

When y is the j -th level, the formula can be written as Equation (2):

$$P_j = P(y = j|X) = \frac{1}{1 + \exp[-(\alpha_j - x'_i\beta_j)]} \quad (2)$$

In a logistic model, a certain group of explained variables is selected as the baseline to estimate the impact of other explained variables on the result variables. The impact is generally reflected in the odds ratio. Assuming that a selected baseline is the first group, then the impact of other explained variables on the result variable relative to the baseline can be written as follows in Equation (3):

$$\frac{P_j}{P_1} = \frac{P(y = j|X)}{P(y = 1|X)} = e^{x'_i\beta_j}, j = 2, \dots, J \quad (3)$$

The effect of the change of l -th explained variables on the odds ratio can be written as $e^{\Delta x_{il}\beta_{jl}}$, where, β_{jl} is the l -th element in the j -th group coefficient vector β_j . The equation shows that, when the other explained variables remain unchanged, for each additional unit of x_{il} , i.e., $\Delta x_{il} = 1$, the odds ratio of selecting the j -th groups relative to baseline is $e^{\beta_{jl}}$. It can be seen that if, and only if, $\beta_{jl} > 0$, the odds ratio of the j -th group relative to the baseline is greater than 1, and the increase in x_{il} will increase the odds ratio of selecting the j -th group relative to the baseline.

4.3. Data Preparation

In this study, helmet use was the result variable. The other factors recorded during observation were explained variables. Finally, 16,206 rows of table data were obtained. Table 2 shows the result variable, explained variables and their codes. STATA 17.0 MP software was used to build the logistic model, with mlogit command.

Table 2. Result variables, explained variables and their codes.

Result Variable	Explained Variable	Data Type	Classification	Code
Helmet use	Area	Category	Urban, Suburban	1, 0
	Time	Category	Morning Peak hours, Off Peak hours, Evening Peak hours	0, 1, 2
	Level of service (LOS)	Sequence	1, 2, 3, 4 Levels	1, 2, 3, 4
	Weather	Category	Rainy, Sunny	1, 0
	Existence of controller (abbreviated as EC)	Category	Yes, No	1, 0
	Average daily temperature (abbreviated as ADT)	Category	≥ 20 °C, < 20 °C	1, 0
	Whether the rider is a food delivery worker (abbreviated as FD)	Category	Yes, No	1, 0

Table 2. Cont.

Result Variable	Explained Variable	Data Type	Classification	Code
Helmet use	Separate non-vehicle lane (abbreviated as SNL)	Category	Yes, No	1, 0
	Gender	Category	Male, Female	1, 0
	Age	Category	≥35 Years old, <35 Years old	1, 0
	Vehicle type	Category	Electric Moped, Electric Bicycle	1, 0

5. Results

5.1. Multicollinearity Diagnostics

Multicollinearity diagnostics are used to examine whether there are correlations among the influencing factors in the electric bicycle rider helmet use model. Severe multicollinearity can increase the standard deviation and variance of model parameters, reduce the significance level of variables, and even cause the model to fail to converge, leading to ineffective predictions. Therefore, this study employs the Variance Inflation Factor (VIF) method and the correlation coefficient matrix method to diagnose multicollinearity among the 11 factor variables included in the model, and Pearson correlation analysis is used to analyze the correlation between variables.

Based on the diagnostic results of the VIF and the correlation coefficient matrix method, it can be concluded that there is no severe multicollinearity among the 11 factor variables in the model (Table 3). Specifically, the VIF values are all below 5 and within a reasonable range, indicating that the correlations among the factors are low and the variables are relatively independent. Therefore, the parameter estimates of the model are reliable, and the standard deviations or variances will not be inflated due to multicollinearity.

Table 3. Results of multicollinearity diagnostics for the independent variables.

Variable	VIF	1/VIF
Los	1.67	0.60
ADT	1.67	0.60
SNL	1.12	0.89
Age	1.12	0.89
Vehicle type	1.11	0.90
Weather	1.08	0.92
EC	1.08	0.93
Time	1.06	0.94
Gender	1.06	0.95
FD	1.04	0.96

Mean VIF: 1.20.

In the correlation coefficient matrix, the correlations between variables do not show significant multicollinearity issues, and the highest correlation coefficients between variable combinations are still within an acceptable range, with no near-perfect correlations close to 1 or −1 (Figure 4). This further indicates that the factor variables in the model can function independently, ensuring the stability of the model.

Overall, the multicollinearity diagnostics indicate that the correlations between selected variables do not significantly affect the regression results, and the model's parameter estimation and predictive performance remain unaffected by multicollinearity. Therefore, the model can be used for further analysis and prediction.

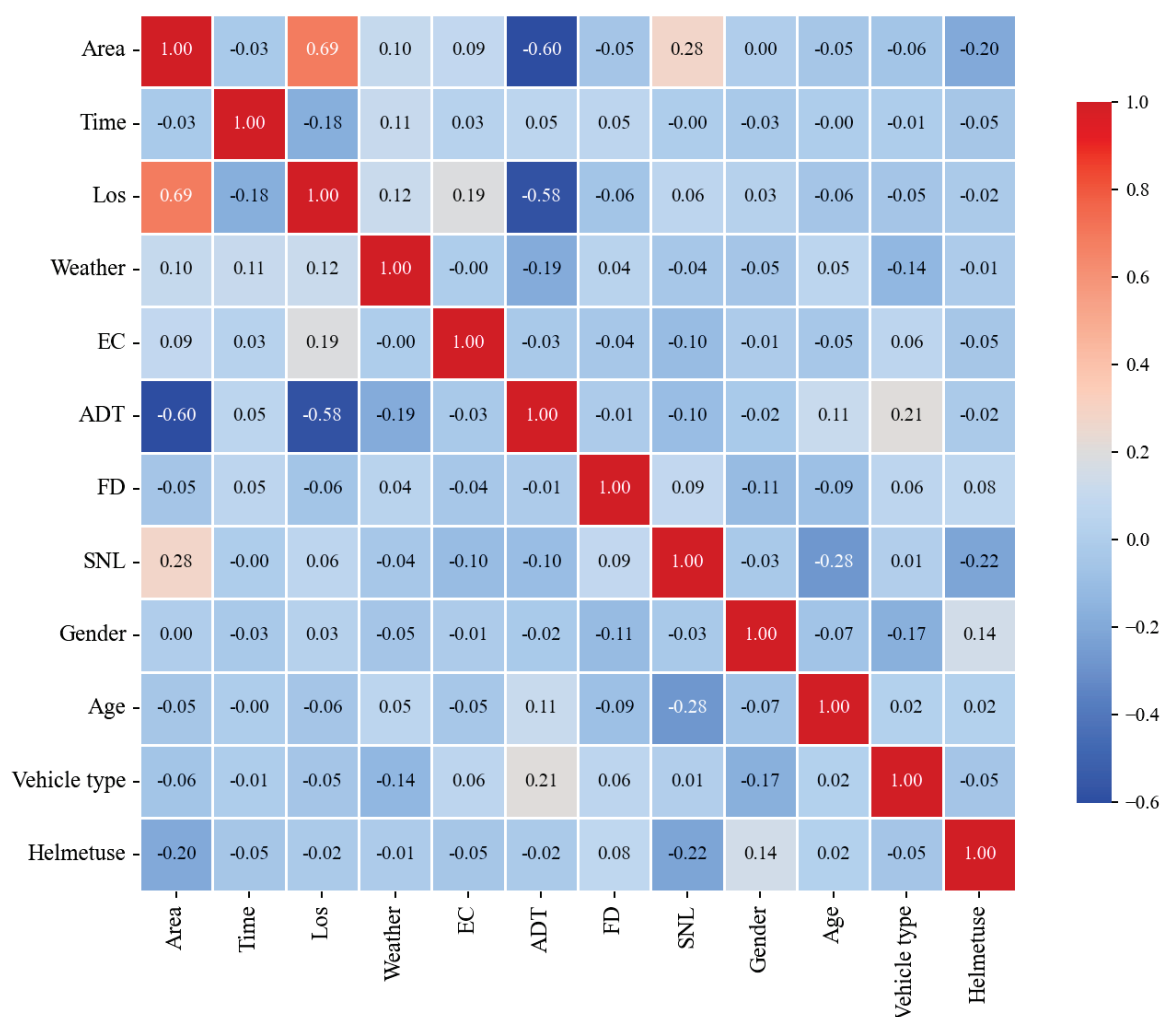


Figure 4. Correlation heatmap.

5.2. Analysis of Multivariate Logistic Regression Model Results

After 8 iterations, the model converges (log likelihood = -8444.285). The model prob $> \chi^2 = 0.000$, pseudo $R^2 = 0.2147$, which indicates that the model fits better than the model containing only constant terms, but the ability of independent variable x to explain the total variation is low. The results of the multivariate logistic model are shown in Table 4. Among them, the odds ratio refers to the probability of using helmets when all other explained variables are controlled at the same level. Taking area as an example, riders in urban areas are more likely to use helmets than those in suburbs by 99.3%. After the main effect analysis, the interaction analysis was carried out. The following Table 5 lists all statistically significant ($p < 0.05$) interaction factors without singular values.

Table 4. Results of main effect analysis.

Explained Variable	Odds Ratio	Possibility of Helmet Use	Standard Error	z	P > z	95% Confidence Interval	
						Lower Limit	Upper Limit
LOS	1			Baseline			
	2	2.798	2.630	1.090	0.274	0.443	17.658
	3	0.398	0.360	−1.020	0.309	0.067	2.349
	4	0.115	0.171	−1.450	0.146	0.006	2.126
Area	Suburb			Baseline			
	Urban	0.007	0.002	−14.460	0.000	0.003	0.013
Weather	Sunny			Baseline			
	Rainy	0.062	0.026	−6.570	0.000	0.027	0.142
ADT	Below 20 °C			Baseline			
	Above 20 °C	0.000	0.000	−6.630	0.000	0.000	0.001
EC	No			Baseline			
	Yes	4.460	2.898	2.300	0.021	1.248	15.940
FD	No			Baseline			
	Yes	3.466	3.645	1.180	0.237	0.441	27.229
SNL	No			Baseline			
	Yes	0.204	0.141	−2.290	0.022	0.052	0.794
Time	Morning peak hours			Baseline			
	Off peak hours	0.036	0.060	−1.980	0.047	0.001	0.962
	Evening peak hours	0.025	0.030	−3.110	0.002	0.002	0.258
	Female			Baseline			
Gender	Male	1.316	0.326	1.110	0.269	0.809	2.139
Age	≤35 years old			1 baseline			
	>35 years old	2.922	0.751	4.170	0.000	1.766	4.837
Vehicle type	Electric bicycle			Baseline			
	Electric moped	0.326	0.097	−3.760	0.000	0.182	0.585

Table 5. Result of the interaction analysis.

Explained Variable	Odds	Possibility of Using Helmet	Standard Error	z	P > z	95% Confidence Interval	
						Lower Limit	Upper Limit
Area*Weather	14.227	1322.7%	4.419	Baseline	0.000	7.740	26.151
Area*FD	0.032	−96.8%	0.039	Baseline	0.005	0.003	0.360
				−2.790			
Area*Gender	1.483	48.3%	0.269	Baseline	0.030	1.040	2.117
				2.170			
Weather*EC	0.608	−39.2%	0.072	Baseline	0.000	0.482	0.767
				−4.190			
Weather*ADT	0.001	−99.9%	0.001	Baseline	0.000	0.000	0.010
				−5.350			
Weather*FD	3.957	295.7%	1.424	Baseline	0.000	1.954	8.012
				3.820			
Weather*SNL	1.405	40.5%	0.147	Baseline	0.001	1.145	1.725
				3.260			
Weather*Gender	1.356	35.6%	0.121	Baseline	0.001	1.138	1.615
				3.400			
Weather*Vehicle type	1.355	35.5%	0.135	Baseline	0.002	1.115	1.646
				3.060			
ADT*EC	0.002	−99.8%	0.002	Baseline	0.000	0.000	0.010
				−7.490			
ADT*FD	0.179	−82.1%	0.139	Baseline	0.027	0.039	0.819
				−2.220			
ADT*Age	1.746	74.6%	0.342	Baseline	0.004	1.189	2.564
				2.850			
ADT*Vehicle type	2.843	184.3%	0.765	Baseline	0.000	1.677	4.819
				3.880			
EC*FD	3.468	246.8%	1.566	Baseline	0.006	1.431	8.404
				2.750			
EC*SNL	0.399	−60.1%	0.051	Baseline	0.000	0.310	0.513
				−7.160			

Table 5. Cont.

Explained Variable	Odds	Possibility of Using Helmet	Standard Error	z	P > z	95% Confidence Interval	
						Lower Limit	Upper Limit
EC*Gender							
		EC = No * Gender = Female		Baseline			
	1.711	EC = Yes * Gender = Male	0.176	5.230	0.000	1.399	2.093
EC*Age		EC = No * Age = ≥ 35 Years old		Baseline			
	0.738	EC = Yes * Age = < 35 Years old	0.088	−2.550	0.011	0.584	0.932
FD*Age		FD = No * Age = ≥ 35 Years old		Baseline			
	0.374	FD = Yes * Age = < 35 Years old	0.131	−2.810	0.005	0.189	0.742
SNL*Age		SNL = No * Age = ≥ 35 Years old		Baseline			
	0.764	SNL = Yes * Age = < 35 Years old	0.082	−2.520	0.012	0.620	0.942
Gender*Age		Gender = Female * Age = ≥ 35 Years old		Baseline			
	1.242	Gender = Male * Age = < 35 Years old	0.111	2.430	0.015	1.043	1.479

6. Discussion

The main effect analysis shows that the possibility of helmet use for riders in urban areas is higher than that of riders in suburban areas, and riders are more reluctant to wear helmets on rainy days. This finding aligns with previous studies, which indicate that stricter law enforcement and better infrastructure in urban areas contribute to higher helmet use rates [44–46]. However, our study further reveals that regulatory differences regarding TWEV modifications play a crucial role. An interaction is found, increasing the possibility of helmet use to 1322.7% relative to the baseline group of Area = Suburb * Weather = Sunny. Apart from the stricter enforcement of helmet use laws in urban areas, this enormous increase may be due to the different regulations regarding TWEV modification in suburbs and urban areas [47]. In urban areas, installing rain canopies on TWEVs is prohibited, making riders more likely to wear helmets on rainy days. However, in suburban areas, regulations on such modifications are more relaxed, leading riders to forgo helmets in favor of installing rain shelters for protection. Compared to similar studies that focus solely on law enforcement, our findings offer a novel perspective by highlighting the indirect effects of vehicle modifications on helmet use.

The rise of ADT also reduces the probability of riders wearing helmets. This contradicts previous research suggesting that higher traffic density leads to increased safety awareness among riders [48]. Interaction analysis proves that this interacts with age. When the ADT rises, older riders are more willing to wear helmets. The reason may be that older riders are more safety conscious, which is also verified by the main effect analysis of age variables. This highlights that helmet use behavior is not only dependent on external traffic conditions but also significantly influenced by demographic factors, an aspect that previous models may have overlooked [44,49].

The existence of traffic controllers (EC) greatly increases the probability of helmet use. On the contrary, separated non-motor vehicle lanes (SNLs) do not improve the use of helmets but increase the probability of riders not wearing helmets. This contrasts with studies, which suggest that dedicated lanes improve overall traffic safety [50,51]. In interaction analysis, even if traffic controllers exist, the separated non-motor vehicle lane still reduces the possibility of helmet use. The source of this problem may be that riders overestimate the protective effect of lane separation, thus reducing helmet use, seen as implementation of an “over-protection” measure [52]. Unlike previous research that primarily focuses on the benefits of lane separation, our study reveals a potential unintended consequence—riders’ false sense of security leading to lower helmet use.

Time of day affects helmet use. Compared with the morning peak hours, the probability of riders wearing helmets during off peak hours and evening peak hours is reduced. The reasons for this situation may be very complicated, involving traffic flow, the characteristics of riders, the existence of controllers, etc. Our findings support previous research indicating that law enforcement intensity varies throughout the day [53], but we further highlight that this variation directly influences riders’ compliance behavior. Most likely, stricter law enforcement during morning peak hours may lead to this situation.

It is noteworthy that, in the main effect analysis, gender does not have a significant impact on helmet use ($p > 0.05$). However, regarding interaction, that between gender and age shows a significant impact. Over 35-years-old male riders are more likely to wear helmets, which may reflect the different attitudes of male and female riders of different ages towards helmet use. This nuanced finding extends previous studies that focus solely on gender differences by demonstrating the necessity of considering interaction effects between demographic variables [49,54]. At the same time, although being a food delivery worker has no influence on helmet use, in interactive analysis, the probability of the over 35-years-old food delivery worker wearing the helmet is less than that of under-35 years-

old riders. This may indicate that work-related urgency or risk perception differences contribute to helmet use behavior, an area that warrants further study.

Types of TWEV have a significant influence on helmet use. Compared with electric bicycle riders, electric moped riders are more reluctant to wear helmets. This aligns with previous research suggesting that riders of larger, more powerful vehicles often exhibit greater risk-taking behavior [55,56]. However, in interactive analysis, when the control temperature is different, the possibility of electric moped riders wearing helmets is higher than that of electric bicycle riders. Further experiments are needed to check and verify this conclusion. This temperature-related factor has been rarely explored in prior studies, suggesting that external environmental conditions could play a more critical role than previously assumed. Further experiments are needed to check and verify this conclusion.

From Tables 3 and 4, it seems that different LOS has no influence on helmet use. All the p values for this are greater than 0.05 in the main effect analysis. This result is inconsistent with previous studies on traffic density and risky behaviors [57], which often report a correlation between congestion levels and safety compliance. The reason for this phenomenon may be that helmet use is not closely related to aggressive riding behavior, which is obstructed by other vehicles in high-density traffic flow. Our findings suggest that helmet use decisions may be less about risk-taking tendencies and more about external regulatory and environmental factors, distinguishing our study from previous behavioral-focused research.

Despite the strong explanatory power of the multinomial Logit model, this study has several limitations. First, while the model provides valuable insights into the influence of various factors on helmet use, it assumes that the relationships between variables are linear and independent. This may oversimplify real-world behavior, as helmet use decisions are influenced by complex and dynamic factors that may require more flexible models, such as mixed-effects models or machine learning approaches. Second, the study relies on observational data collected at specific locations in Fuzhou, which may limit the generalizability of the findings to other cities or regions with different regulatory and cultural contexts. Third, some unobserved variables, such as individual risk perception, attitudes toward law enforcement, and socioeconomic factors, were not included in the model, potentially affecting the robustness of the conclusions. Future research should consider incorporating more comprehensive behavioral data and exploring alternative modeling techniques to enhance predictive accuracy and generalizability.

7. Conclusions

This study conducted an observational analysis to collect helmet use data from 16,206 two-wheeled electric vehicle (TWEV) riders in Fuzhou, China. A multivariate logistic regression model was developed to investigate the influence of various factors, including level of service (LOS), area type (urban/suburban), average daily traffic (ADT), weather, etc., on helmet use. The model demonstrated high goodness of fit, effectively explaining how these factors impact helmet-wearing behavior.

The findings revealed that several factors, such as area type, weather conditions, average daily temperature, the presence of traffic controllers, separated non-motor vehicle lanes, time of day, rider age, and vehicle type, significantly influenced helmet use. Moreover, the interactions among these factors were notable, with weather and average daily temperature showing particularly strong interaction effects. In contrast, LOS, gender, and whether the rider was a food delivery worker or not did not have significant direct effects on helmet use but exhibited significant interaction effects with other factors.

A key observation was that separated non-motor vehicle lanes might create a false sense of security, leading to lower helmet use among riders. Additionally, the analysis

highlighted discrepancies between the conclusions drawn from interaction effect analysis and those from main effect analysis. These findings align with previous research but provide additional insights into the behavioral dynamics of helmet use, emphasizing the role of environmental and regulatory factors.

Future research should focus on several key areas to build upon these findings. First, alternative modeling approaches, such as mixed-effects models or machine learning methods, should be explored to better capture the complexity of helmet use behaviors. Second, additional behavioral variables, including riders' individual risk perceptions, attitudes toward law enforcement, and socioeconomic factors, should be integrated to enhance the robustness of the analysis. Third, expanding the study to different cities or regions with varying regulatory environments will help assess the generalizability of the conclusions. Finally, controlled experiments could be conducted to further validate the interaction effects identified in this study, ensuring a more precise understanding of the factors influencing helmet use.

Author Contributions: Conceptualization, W.L. and C.L.; formal analysis, W.L. and W.Z.; methodology, W.Z. and Y.Y.; resources, Y.Y.; data curation, L.W.; writing—original draft, W.L. and C.L.; writing—review and editing, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Evaluation Standard for the Effectiveness of Traffic Safety Facilities on National and Provincial Trunk Roads (2023 Guangxi Transportation Standardization Project), and the Central Government's Special Fund for Local Scientific and Technological Development (Project No. 2023L3033).

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Author Congying Li was employed by Guangxi Communications Design Group Co., Ltd. Author Weibin Zheng was employed by Fuzhou Fuda Automation Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. eBicycles. E-Bike Facts & Statistics for 2021. 2021. Available online: <https://www.ebicycles.com/ebike-facts-statistics/> (accessed on 9 February 2025).
2. Gibson, K. Big Jump in e-Scooter, e-Bike Injuries the Last Four Years, CPSC Finds. 2021. Available online: <https://www.cbsnews.com/news/electric-scooter-electric-bike-accidents-rise-last-four-years/> (accessed on 9 February 2025).
3. Bambach, M.R.; Mitchell, R.; Grzebieta, R.H.; Olivier, J. The effectiveness of helmets in bicycle collisions with motor vehicles: A case-control study. *Accid. Anal. Prev.* **2013**, *53*, 78–88. [CrossRef] [PubMed]
4. Kwiatkowski, M.A.; Grzelak-Kostulska, E.; Biegańska, J. Could it be a bike for everyone? The electric bicycle in Poland. *Energies* **2021**, *14*, 4878. [CrossRef]
5. Mohan, D. Evidence-based interventions for road traffic injuries in South Asia. *J. Coll. Physicians Surg.-Pak. JCPSP* **2004**, *14*, 746–747.
6. Tang, T.; Guo, Y.; Wang, H.; Li, X.; Agrawal, S. Determinants of helmet use intention among E-bikers in China: An application of the theory of planned behavior, the health belief model, and the locus of control. *Transp. Res. Rec.* **2024**, *2678*, 753–769. [CrossRef]
7. Branas, C.C.; Knudson, M.M. Helmet laws and motorcycle rider death rates. *Accid. Anal. Prev.* **2001**, *33*, 641–648. [CrossRef] [PubMed]
8. Dodds, N.; Johnson, R.; Walton, B.; Bouamra, O.; Yates, D.; Lecky, F.E.; Thompson, J. Evaluating the impact of cycle helmet use on severe traumatic brain injury and death in a national cohort of over 11000 pedal cyclists: A retrospective study from the NHS England Trauma Audit and Research Network dataset. *BMJ Open* **2019**, *9*, e027845. [CrossRef]
9. Karkhaneh, M.; Kalenga, J.-C.; Hagel, B.E.; Rowe, B.H. Effectiveness of bicycle helmet legislation to increase helmet use: A systematic review. *Inj. Prev.* **2006**, *12*, 76–82. [CrossRef] [PubMed]
10. Siebert, F.W.; Albers, D.; Naing, U.A.; Perego, P.; Santikarn, C. Patterns of motorcycle helmet use—A naturalistic observation study in Myanmar. *Accid. Anal. Prev.* **2019**, *124*, 146–150. [CrossRef]

11. Manan, M.M.A.; Várhelyi, A. Motorcyclists' road safety related behavior at access points on primary roads in Malaysia—A case study. *Saf. Sci.* **2015**, *77*, 80–94. [CrossRef]
12. Rusli, R.; Oviedo-Trespalacios, O.; Abd Salam, S.A. Risky riding behaviours among motorcyclists in Malaysia: A roadside survey. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *74*, 446–457. [CrossRef]
13. Li, Y.; Chen, Q.; Ma, Q.; Yu, H.; Huang, Y.; Zhu, L.; Zhang, H.; Li, C.; Lu, G. Injuries and risk factors associated with bicycle and electric bike use in China: A systematic review and meta-analysis. *Saf. Sci.* **2022**, *152*, 105769. [CrossRef]
14. Peltzer, K.; Pengpid, S. Helmet use and associated factors among motorcyclists in the Association of Southeast Asian Nations: Prevalence and effect of interventions. *Afr. Saf. Promot. A J. Inj. Violence Prev.* **2014**, *12*, 72–86.
15. Wang, X.; Chen, J.; Quddus, M.; Zhou, W.; Shen, M. Influence of familiarity with traffic regulations on delivery riders' e-bike crashes and helmet use: Two mediator ordered logit models. *Accid. Anal. Prev.* **2021**, *159*, 106277. [CrossRef]
16. Buhler, T.; Comby, E.; Vaudor, L.; von Pape, T. Beyond 'good' and 'bad' cyclists. On compensation effects between risk taking, safety equipment and secondary tasks. *J. Transp. Health* **2021**, *22*, 101131. [CrossRef]
17. Eccarius, T.; Leung, A.; Shen, C.-W.; Burke, M.; Lu, C.-C. Prospects for shared electric velomobility: Profiling potential adopters at a multi-campus university. *J. Transp. Geogr.* **2021**, *96*, 103190. [CrossRef]
18. Huemer, A.K.; Banach, E.; Bolten, N.; Helweg, S.; Koch, A.; Martin, T. Secondary task engagement, risk-taking, and safety-related equipment use in German bicycle and e-scooter riders—An observation. *Accid. Anal. Prev.* **2022**, *172*, 106685. [CrossRef]
19. Zhou, J.; Zheng, T.; Dong, S.; Mao, X.; Ma, C. Impact of helmet-wearing policy on e-bike safety riding behavior: A bivariate ordered probit analysis in Ningbo, China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2830. [CrossRef]
20. Du, W.; Yang, J.; Powis, B.; Zheng, X.; Ozanne-Smith, J.; Bilston, L.; Wu, M. Understanding on-road practices of electric bike riders: An observational study in a developed city of China. *Accid. Anal. Prev.* **2013**, *59*, 319–326. [CrossRef]
21. Debnath, A.K.; Haworth, N.; Heesch, K.C. Women cycling in Queensland: Results from an observational study. *Accid. Anal. Prev.* **2021**, *151*, 105980. [CrossRef]
22. Pai, C.-W.; Hsu, J.-J.; Chang, J.-L.; Kuo, M.-S. Motorcyclists violating hook-turn area at intersections in Taiwan: An observational study. *Accid. Anal. Prev.* **2013**, *59*, 1–8. [CrossRef]
23. Feng, M.; Wang, X.; Quddus, M. Developing multivariate time series models to examine the interrelations between police enforcement, traffic violations, and traffic crashes. *Anal. Methods Accid. Res.* **2020**, *28*, 100139. [CrossRef]
24. 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 Traffic Psychol. Behav.* **2018**, *56*, 344–353. [CrossRef]
25. Chai, H.; Zhang, Z.; Xue, J.; Hu, H. A quantitative traffic performance comparison study of bicycles and E-bikes at the non-signalized intersections: Evidence from survey data. *Accid. Anal. Prev.* **2022**, *178*, 106853. [CrossRef]
26. Qian, Q.; Qi, Y.; Shi, J. Description and analysis of aberrant riding behaviors of pedal cyclists, e-bike riders and motorcyclists: Based on a self-report questionnaire. *Transp. Res. Part F Traffic Psychol. Behav.* **2024**, *107*, 969–984. [CrossRef]
27. Gioldasis, C.; Christoforou, Z.; Seidowsky, R. Risk-taking behaviors of e-scooter users: A survey in Paris. *Accid. Anal. Prev.* **2021**, *163*, 106427. [CrossRef] [PubMed]
28. Shah, N.R.; Aryal, S.; Wen, Y.; Cherry, C.R. Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology. *J. Saf. Res.* **2021**, *77*, 217–228. [CrossRef]
29. Sadeghi, M.; Aghabayk, K.; Quddus, M. A hybrid Machine learning and statistical modeling approach for analyzing the crash severity of mobility scooter users considering temporal instability. *Accid. Anal. Prev.* **2024**, *206*, 107696. [CrossRef] [PubMed]
30. Gao, D.; Zhang, X. Injury severity analysis of single-vehicle and two-vehicle crashes with electric scooters: A random parameters approach with heterogeneity in means and variances. *Accid. Anal. Prev.* **2024**, *195*, 107408. [CrossRef]
31. Ye, F.; Wang, C.; Cheng, W.; Liu, H. Exploring Factors Associated with Cyclist Injury Severity in Vehicle-Electric Bicycle Crashes Based on a Random Parameter Logit Model. *J. Adv. Transp.* **2021**, *2021*, 5563704. [CrossRef]
32. Azimian, A.; Jiao, J. Modeling factors contributing to dockless e-scooter injury accidents in Austin, Texas. *Traffic Inj. Prev.* **2022**, *23*, 107–111. [CrossRef]
33. Yadukul, S. A Study of Preventive Risk Factors in Fatal Road Traffic Injuries, with Special Emphasis to Helmets and Seatbelts. Ph.D. Dissertation, Rajiv Gandhi University of Health Sciences (India), Bengaluru, India, 2013.
34. Bonnet, E.; Lechat, L.; Ridde, V. What interventions are required to reduce road traffic injuries in Africa? A scoping review of the literature. *PLoS ONE* **2018**, *13*, e0208195. [CrossRef] [PubMed]
35. Cripton, P.A.; Dressler, D.M.; Stuart, C.A.; Dennison, C.R.; Richards, D. Bicycle helmets are highly effective at preventing head injury during head impact: Head-form accelerations and injury criteria for helmeted and unhelmeted impacts. *Accid. Anal. Prev.* **2014**, *70*, 1–7. [CrossRef] [PubMed]
36. Thompson, D.C.; Rivara, F.P.; Thompson, R.S. Effectiveness of bicycle safety helmets in preventing head injuries: A case-control study. *Jama* **1996**, *276*, 1968–1973. [CrossRef] [PubMed]
37. Peden, M.M. *World Report on Road Traffic Injury Prevention*; World Health Organization: Geneva, Switzerland, 2004.

38. Moftakhar, T.; Wanzel, M.; Vojcsik, A.; Kralinger, F.; Mousavi, M.; Hajdu, S.; Aldrian, S.; Starlinger, J. Incidence and severity of electric scooter related injuries after introduction of an urban rental programme in Vienna: A retrospective multicentre study. *Arch. Orthop. Trauma Surg.* **2021**, *141*, 1207–1213. [CrossRef]
39. Hamzani, Y.; Bar Hai, D.; Cohen, N.; Drescher, M.J.; Chaushu, G.; Yahya, B.H. The impact of helmet use on oral and maxillofacial injuries associated with electric-powered bikes or powered scooter: A retrospective cross-sectional study. *Head Face Med.* **2021**, *17*, 1–8. [CrossRef]
40. Zamani-Alavijeh, F.; Bazargan, M.; Shafiei, A.; Bazargan-Hejazi, S. The frequency and predictors of helmet use among Iranian motorcyclists: A quantitative and qualitative study. *Accid. Anal. Prev.* **2011**, *43*, 1562–1569. [CrossRef]
41. Yuan, Q.; Shi, H.; Zhao, J.; Li, R. Influencing factors analysis of helmet wearing for electric bicycle riders based on ordinal multinomial logistic model. *Transp. Saf. Environ.* **2022**, *4*, tdac001. [CrossRef]
42. Yang, Y.; Li, C.; Cheng, K.; Hu, S. Factors affecting the intention to wear helmets for e-bike riders: The case of Chinese college students. *Int. J. Inj. Control Saf. Promot.* **2024**, *31*, 487–498. [CrossRef]
43. Ma, W.; Liu, Z.; Li, R. Effect of helmet-wearing regulation on electric-bike riders: A case study of two cities in China. *Transp. Saf. Environ.* **2024**, *6*, tdad038. [CrossRef]
44. Jomnonkwo, S.; Watthanaklang, D.; Sangphong, O.; Champahom, T.; Laddawan, N.; Uttra, S.; Ratanavaraha, V. A comparison of motorcycle helmet wearing intention and behavior between urban and rural areas. *Sustainability* **2020**, *12*, 8395. [CrossRef]
45. Guzman, L.A.; Ortiz, A.I.V.; Mesa, V.G.; Camargo, J.P.; Allen, K.A.; Hyder, A.A. The relationships between correct helmet use, enforcement presence, and mortality in a Latin-America city: The case study of Bogotá, Colombia. *Traffic Inj. Prev.* **2020**, *21*, 500–505. [CrossRef] [PubMed]
46. Satiennam, T.; Kumphong, J.; Satiennam, W.; Klungboonkrong, P.; Jaensirisak, S.; Ratanavaraha, V. Change in helmet use behavior enforced by CCTV cameras with automatic helmet use detection system on an urban arterial road. *Traffic Inj. Prev.* **2020**, *21*, 494–499. [CrossRef]
47. Ruan, Y.; Hang, C.-C.; Wang, Y.; Ma, R. The role of government in an emerging disruptive innovation: The case of E-bikes in China. In *Disruptive Innovation in Chinese and Indian Businesses*; Routledge: London, UK, 2013; pp. 50–67.
48. Thompson, J.H.; Wijnands, J.S.; Mavoa, S.; Scully, K.; Stevenson, M.R. Evidence for the ‘safety in density’ effect for cyclists: Validation of agent-based modelling results. *Inj. Prev.* **2019**, *25*, 379–385. [CrossRef] [PubMed]
49. Son, S.H.; Oh, S.H.; Kang, S.H.; Kim, D.-K.; Seo, K.M.; Lee, S.-U.; Lee, S.Y. Independent factors associated with bicycle helmet use in a Korean population: A cross-sectional study. *Traffic Inj. Prev.* **2018**, *19*, 399–403. [CrossRef] [PubMed]
50. Rothenberg, H.; Goodman, D.; Sundstrom, C. Separated Bike Lane Crash Analysis. 2016. Available online: <https://trid.trb.org/View/1393609> (accessed on 9 February 2025).
51. Cicchino, J.B.; McCarthy, M.L.; Newgard, C.D.; Wall, S.P.; DiMaggio, C.J.; Kulie, P.E.; Arnold, B.N.; Zuby, D.S. Not all protected bike lanes are the same: Infrastructure and risk of cyclist collisions and falls leading to emergency department visits in three US cities. *Accid. Anal. Prev.* **2020**, *141*, 105490. [CrossRef]
52. Ziakopoulos, A.; Nikolaou, D.; Yannis, G. Correlations of multiple rider behaviors with self-reported attitudes, perspectives on traffic rule strictness and social desirability. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *80*, 313–327. [CrossRef]
53. Bhatti, J.A.; Ejaz, K.; Razzak, J.A.; Tunio, I.A.; Sodhar, I. Influence of an enforcement campaign on seat-belt and helmet wearing, Karachi-Hala highway, Pakistan. *Ann. Adv. Automot. Med.* **2011**, *55*, 65–70.
54. Mensah, E.; Ketemepi, G.; Bosompem-Acheampong, K.; Alhassan, M. Factors associated with helmet use among motorcyclists presenting with maxillofacial injuries to a tertiary hospital in Ghana. *Ghana Dent. J.* **2023**, *20*, 14–16.
55. Qian, Y.; Sun, Q.; Fei, G.; Li, X.; Stallones, L.; Xiang, H.; Zhang, X. Riding behavior and electric bike traffic crashes: A Chinese case-control study. *Traffic Inj. Prev.* **2020**, *21*, 24–28. [CrossRef]
56. Teoh, E.R.; Campbell, M. Role of motorcycle type in fatal motorcycle crashes. *J. Saf. Res.* **2010**, *41*, 507–512. [CrossRef]
57. 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]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Systems Editorial Office
E-mail: systems@mdpi.com
www.mdpi.com/journal/systems



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-8112-3