Leveraging Macroscopic Fundamental Diagrams (MFDs) and Control Strategies for Sustainable Transportation Networks

Leveraging Macroscopic Fundamental Diagrams (MFDs) and Control Strategies for Sustainable Transportation Networks

By
MOHAMMAD HALAKOO,
M.Sc., B.Sc.

A Thesis Submitted to the School of Graduate Studies
in the Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
in
Civil Engineering

McMaster University
Hamilton, Ontario

© Copyright by Mohammad Halakoo, December 14, 2023

Doctor of Philosophy (2023)
Civil Engineering
McMaster University
Hamilton, Ontario, Canada

TITLE: Leveraging Macroscopic Fundamental Diagrams (MFDs) and Control Strategies for Sustainable Transportation Networks

AUTHOR:
Mohammad Halakoo

SUPERVISOR:
Dr. Hao Yang
Assistant Professor, Department of Civil Engineering,
McMaster University, ON, Canada

NUMBER OF PAGES: xvii, 166

Lay Abstract

This thesis comprehensively tackles urban transportation challenges, prioritizing resilience and sustainability. It centers on the Macroscopic Fundamental Diagram (MFD) and associated control methods, investigating congestion, bottleneck identification, and environmental impact reduction. By leveraging MFD principles, it studies how network configurations and congestion influence tra□c flow while highlighting the detrimental e□ects of congestion and emissions on e□ciency and the environment. Introducing concepts such as the emission Macroscopic Fundamental Diagram (e-MFD) and an innovative controller illustrates their e□ectiveness in lessening environmental impact. Also, the inclusion of decision tree models strengthens bottleneck management strategies. This thesis o□ers a holistic framework and equips policymakers and transportation authorities with sustainable strategies for enhancing urban transportation performance and environmental well-being. Its insights pave the way for future research and underscore the significance of holistic approaches in shaping sustainable urban transportation networks.

Abstract

Urban transportation networks form the backbone of modern societies, enabling essential mobility and economic activities while presenting intricate challenges associated with tra□c congestion, emissions, and environmental sustainability. This thesis presents a comprehensive and integrated study framework to enhance the robustness and sustainability of urban transportation infrastructure. With a primary focus on the Macroscopic Fundamental Diagram (MFD) and its associated control strategies, the research focuses on the complex dynamics of network performance, emphasizing the critical interplay between congestion management, bottleneck identification, and the reduction of environmental impact within complex urban environments.

The initial chapters provide a comprehensive overview of the challenges inherent within contemporary urban transportation systems, emphasizing the adverse impacts of tra□c congestion and emissions on network e□ciency and the environment. Leveraging the principles of the MFD, the study investigates the diverse dynamics of network performance, demonstrating the pivotal role of network configurations and congestion distributions in shaping tra□c flow patterns and emissions.

Subsequent chapters extend the research framework by introducing novel concepts such as the heterogeneity-aware emission Macroscopic Fundamental Diagram (e-MFD) and the innovative eMFD controller. The study showcases the e□cacy of these novel approaches in mitigating environmental impact and optimizing network performance by leveraging the MFD, eMFD, and the Model Predictive Controller (MPC). Developing comprehensive decision tree models for bottleneck identification and management further enhances the reliability and applicability of tra□c control strategies within complex urban environments.

In conclusion, this thesis serves as a seminal contribution to the field of transportation engineering, o□ering a comprehensive and integrated framework for policymakers and transportation authorities to develop sustainable strategies for enhancing urban transportation infrastructure's performance and environmental sustainability. The findings presented within this research provide a solid foundation for future research endeavours and underscore the imperative of holistic approaches in shaping the future of sustainable urban transportation networks.

Acknowledgements

As I approach the end of this academic journey, I am deeply grateful to those who assisted me in making this endeavour possible. Foremost, I extend my heartfelt appreciation to my family, whose unwavering support and encouragement have been the foundation of my perseverance throughout this journey. Their constant belief in my potential has been a source of strength and inspiration, propelling me through challenges and triumphs.

I am indebted to my supervisor, Dr. Hao Yang, whose profound guidance and mentorship have shaped this thesis into a coherent and impactful contribution to the field of transportation engineering.

Moreover, I would like to express my gratitude to my friends, Amir Hossein, Siyavash, Mohsen, Ehsan and Ali, whose support has been a constant source of encouragement and motivation. Their encouragement and empathy have significantly contributed to the balance and perspective that have sustained me during the rigorous demands of this academic undertaking.

I am also immensely grateful to Toyota Info Tech Lab, Cubic Transportation Systems, and the Natural Sciences and Engineering Research Council (NSERC) for their financial support, which has played a pivotal role in facilitating the execution of this project. Their generous contributions have facilitated the smooth progression of my research and reinforced my belief in the vital role of industry-academia collaboration in the advancement of transportation engineering.

Lastly, my sincere appreciation goes to the esteemed supervisory committee members, whose insightful feedback, astute observations, and invaluable guidance have played a crucial role in shaping my academic and professional growth.

Contents

Lay Abstract iii

Abstract iv

Acknowledgements v

List of Figures x

List of Tables xiii

1 Introduction 1
- 1.1 Background and motivation 1
- 1.2 Tra□c state estimation and data requirements 3
- 1.3 Tra□c control 5
- 1.4 Research objectives 6
- 1.5 Dissertation organization 8
- Bibliography 11

2 Evaluation of Macroscopic Fundamental Diagram Transition in the Era of Connected and Autonomous Vehicles 14
- 2.1 Abstract 15
- 2.2 Introduction 15
- 2.3 Related Work 16
- 2.4 Microscopic Simulation Analysis 18
  - 2.4.1 Simulated Network 19
  - 2.4.2 Features Of HDVs and CAVs 19
  - 2.4.3 Simulation Scenarios 22
- 2.5 Results and Discussion 22
  - 2.5.1 MFDs of Grid Networks 22
  - 2.5.2 Hysteresis Loops and Resilience 23

2.5.3 Accessibility . . . 26
2.6 Conclusions . . . 26
Bibliography . . . 28

3 Quantifying the impact of bottlenecks on network performance and Macroscopic fundamental diagram using clustering and decision tree prediction models 32
3.1 Abstract . . . 33
3.2 Introduction . . . 33
3.3 Literature review . . . 35
3.4 Methods . . . 38
3.4.1 MFD and hysteresis . . . 38
3.4.2 Clustering . . . 39
3.4.3 Classification . . . 42
3.5 Numerical Experiment . . . 44
3.5.1 Grid network . . . 44
3.5.2 Real-world network: Manhattan, New York . . . 45
3.6 Results . . . 47
3.6.1 Grid Network . . . 47
3.6.2 Real-world Manhattan Network . . . 52
3.7 Discussion . . . 56
3.8 Conclusions . . . 59
Bibliography . . . 60

4 Enhancing Macroscopic Fundamental Diagram Accuracy in the Presence of Bottlenecks Through Machine Learning-based Scatter Classification and Correction 65
4.1 Abstract . . . 66
4.2 Introduction . . . 66
4.3 Literature review . . . 68
4.4 Methods . . . 71
4.5 Numerical Experiment . . . 72
4.5.1 9x9 grid network . . . 72
4.6 Simulation results . . . 75
4.7 Results analysis . . . 76
4.7.1 Clustering bottleneck MFDs . . . 76
4.7.2 MFD error modelling for each cluster . . . 81

4.7.3 Predicting MFD type from bottleneck features 83
4.7.4 Proposed bottleneck controller 84
4.8 Discusion 91
4.9 Conclusions 92
Bibliography 93

5 Heterogeneity aware emission macroscopic fundamental diagram (eMFD) 97
5.1 Abstract 98
5.2 Introduction 98
5.3 Review of Methodology 101
5.3.1 Macroscopic fundamental diagram (MFD) 102
5.3.2 Emission estimation 103
5.4 Modelling framework 103
5.5 Numerical experiment 104
5.5.1 Grid network 104
5.5.2 Blacksburg, VA network 111
5.6 Discussion 118
5.7 Conclusion 119
Bibliography 120

6 A Model Predictive Emission Controller Based on the Heterogeneity Aware Emission Macroscopic Fundamental Diagram 125
6.1 Abstract 126
6.2 Introduction 126
6.3 Literature review 128
6.4 Methodology 132
6.4.1 Overview 132
6.4.2 Macroscopic Plant model 132
6.4.3 eMFD Model Predictive Controller 138
6.4.4 Benchmark scenarios 139
6.5 Results 140
6.5.1 Settings of Numerical Experiments 140
6.5.2 Controller performance comparison 141
6.5.3 Sensitivity Analysis 146
6.6 Discussion 151
6.7 Conclusions 153
Bibliography 154

7 Conclusions 158
7.1 Summary 159
7.2 Conclusions and contributions 160
7.2.1 Conclusions and contributions about CAV impact on MFD 160
7.2.2 Conclusions and contributions about bottleneck impacts on MFD . 161
7.2.3 Conclusions and contributions bottleneck management 161
7.2.4 Conclusions and contributions about eMFD 162
7.2.5 Conclusions and contributions about eMFD controller 163
7.2.6 Overall conclusions and contributions 164
7.3 Study limitations and Future work 165

List of Figures

1.1 Emperical MFD from simulation 4
1.2 Thesis structure 10

2.1 Grid Network 20
2.2 Network MFD: A (0% MPR no accident), B (20% MPR no accident), C (50% MPR no accident), D (80% MPR no accident), E (0% MPR with accident), F (20% MPR with accident), G (50% MPR with accident), H (80% MPR with accident) 24
2.3 Network density standard deviation: A (0% MPR no accident), B (20% MPR no accident), C (50% MPR no accident), D (80% MPR no accident), E (0% MPR with accident), F (20% MPR with accident), G (50% MPR with accident), H (80% MPR with accident) 25
2.4 Trip completion: A (OW, no accident), B (TW, no accident), C (TWL, no accident), D (OW, with accident), E (TW, with accident), F (TWL, with accident) 27

3.1 A Clockwise Hysteresis with Residual Density Kr 39
3.2 9x9 Grid network 45
3.3 Manhattan Network 46
3.4 A) Hysteresis area probability distribution, B) Residual density probability distribution, C) Hysteresis area and residual density scatter plot (Grid network) 47
3.5 Clustering validation (grid network) indexes: A) inertia, B) average silhouette, C) Calinski-Harabasz, D) Davies Bouldin for the range of clusters (Grid Network) 49
3.6 Cluster pro ling (grid network) based on: A) hysteresis area, B) residual density, C) MFD capacity 50
3.7 Classi cation model calibration and testing (grid network) A) tree depth hyperparameter tuning, B) Confusion matrix for the training set, C) confusion matrix for testing set 51

3.8 Decision tree model (grid network) 53
3.9 A) Hysteresis area probability distribution, B) Residual density probability distribution, C) Hysteresis area and residual density scatter plot (Manhattan network) 54
3.10 Cluster pro ling (Manhattan) based on: A) hysteresis area, B) residual density, C) MFD capacity 54
3.11 Classi cation model calibration and testing (Manhattan network) A) tree depth hyperparameter tuning, B) Confusion matrix for the training set, C) confusion matrix for the testing set 55
3.12 Decision tree model (Manhattan) 56

4.1 Methods Overview 71
4.2 9x9 Grid network 73
4.3 Network density in the base scenario at A) 60 minutes, B) 120 minutes, C) 180 minutes after the start of simulation 74
4.4 Base case (no bottlenecks) A) MFD, B) Density C) Standard deviation of density 75
4.5 60 minutes bottlenecks e ect on A) MFD, B) density, C) standard deviation of density. 35 minutes bottlenecks e ect on D) MFD, E) density, F) standard deviation of density 15 minutes bottlenecks e ect on G) MFD, H) density, I) standard deviation of density 77
4.6 A) Hysteresis area probability distribution, B) Residual density probability distribution, C) Hysteresis area and residual density scatter plot . . . 78
4.7 Clustering validation indexes: A) inertia, B) average silhouette, C) Calinski-Harabasz, D) Davies Bouldin for the range of clusters 79
4.8 Cluster pro ling A) Hysteresis area, B) Residual density, C) Capacity . . 80
4.9 Error distribution A) high impact MFDs B) low impact MFDs 82
4.10 Decision Tree model A) tree depth hyperparameter tuning, B) Confusion matrix for training set, C) confusion matrix for testing set 83
4.11 Macroscopic plant model 85
4.12 Control loop 88
4.13 Proposed bottleneck controller logic 89
4.14 Decision tree model 90

5.1 Modelling Framework 105
5.2 Grid network 106

5.3 Grid network base scenario results: A) MFD, B) density and CO_2, C) speed and CO_2, D) standard deviation of density and CO_2, E) eMFD . . 107
5.4 Validation results (grid network): A) unseen balanced demand using the Saedi model, B) unseen directional demand using the Saedi model, C) accident using the Saedi model, D) unseen balanced demand using our model, E) unseen directional demand using our model, F) accident using our model 112
5.5 Blacksburg, VA 113
5.6 Simulation results for Blacksburg, VA: A) MFD, B) density and CO_2, C) speed and CO_2, D) standard deviation of density and CO_2, E) eMFD . . 115
5.7 Validation results for Blacksburg Network: A) Saedi model, B) Our model, C) CO_2 rate (Saedi), D) CO_2 rate (Our model), E) Cumulative CO_2 rate (Saedi), F) Cumulative CO_2 rate (Our model) 117

6.1 Blacksburg, VA Network 134
6.2 Modelling overview 135
6.3 Macroscopic plant model 135
6.4 Estimated Model Parameters based on the Blacksburg, A) MFD, B) Speed vs Accumulation, C)Accumulation Standard Deviation vs Accumulation, D) eMFD vs microscopic emission model (Panis et al.; 2006) 137
6.5 In ow rates of di erent controllers 142
6.6 Controllers performance metrics 145
6.7 Cumulative performance comparison with no control case 146
6.8 Sensitivity analysis of controller parameters 148
6.9 Sensitivity analysis of initial conditions 149
6.10 Sensitivity analysis of demand levels 150
6.11 Sensitivity analysis of sampling time 151
6.12 Regional standard deviation of accumulation 153

List of Tables

2.1 Comparison of HDV and CAV features 21

3.1 Model's performance metrics (grid network) 51
3.2 Model's performance metrics (Manhattan) 55

4.1 Error model calibration 82
4.2 Generalized MFD performance improvement with error modelling 83
4.3 Models performance metrics 89

5.1 Literature Review 100
5.2 Notation 102
5.3 Model parameters (grid network) 110
5.4 Training goodness of t indexes (grid network) 110
5.5 Model validation results for the grid network 111
5.6 Calibration results for Blacksburg network 114
5.7 Training goodness of t indexes (Blacksburg network) 114
5.8 Validation results for Blacksburg network 118

6.1 Variables description 133
6.2 Model Parameters 138
6.3 Greedy Controller (GC) logic (Geroliminis et al.; 2012) 140
6.4 Initial Tra c Conditions 147

List of Abbreviations

AI	Arti cial Intelligence
API	Application Programming Interface
AV	Autonomous vehicles
CAV	Connected autonomous vehicles
CBD	Central business district
CACC	Cooperative Adaptive Cruise Control
CMEM	Comprehensive Modal Emission Model
CO	Carbon monoxide
CO_2	Carbon dioxide
COPERT	Computer Programme to calculate Emissions from Road Transport
CS	Critical Score
DUE	Dynamic user equilibrium
DTA	Dynamic Tra c Assignment
EPA	United States Environmental Protection Agency
eMFD	emission Macroscopic Fundamental Diagram
GC	Greedy controller
GPS	Global positioning system
HDV	Human-driven vehicles
IEA	International Energy Agency
LDV	Light-Duty Vehicles
LoS	Level of service
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MFD	Macroscopic fundamental diagram
MPC	Model Predictive Controller
MPR	Market Penetration Rate
MSE	Mean Absolute Error
NEF	Network exit function
NO_x	Nitrogen Oxides
OD	Origin-destination
OLS	Ordinary Least Squares
OW	One way
PI(D)	Proportional integral (derivative) controller

PM	Particulate matter
ROC-AUC	Area Under the Curve of the Receiver Operating Characteristic
RMSE	Root Mean Squared Error
STD	Standard deviation
TW	Two way
TWL	Two-way with prohibited left turns
V2I	Vehicle-to-infrastructure
Veh	Vehicle
VOC	Volatile organic compounds

Declaration of Academic Achievement

I, Mohammad Halakoo, declare that this thesis titled, Leveraging Macroscopic Fundamental Diagrams (MFDs) and Control Strategies for Sustainable Transportation Networks , and works presented in it are my own. The Academic supervisor Dr. Hao Yang provided advice and guidance on the thesis. Amir Hossein Karbasi co-authored and contributed to the simulation model of the paper presented in Chapter 3. Also, Yashar Zeiynali Farid, Seyhan Ucar, and Kentaro Oguchi assisted with the conceptualization and revision of the papers presented in Chapters 3 and 4. Harith Abdulsattar helped conceptualize the paper presented in Chapter 5. Atriya Biswas and Peter Azer assisted me with the conceptualization of the paper presented in Chapter 6. I con rm that this dissertation was prepared following the guidelines for the sandwich thesis format set by the School of Graduate Studies (SGS) at McMaster University. The sandwich thesis is a compilation of journal articles published or prepared for publication. Chapter 5 is already published as a journal article, and Chapter 6 is submitted for publication as a journal article. Chapters 3 and 4 are prepared for submission to a journal. Information presented from outside sources, which has been used for analysis or discussion, has been cited where appropriate; all other materials are the sole work of the author. This thesis consists of the following manuscripts in the following chapters:

- Chapter 2 Halakoo, M., & Yang, H. (2021). Evaluation of Macroscopic Fundamental Diagram Transition in the Era of Connected and Autonomous Vehicles. In 2021 IEEE Intelligent Vehicles Symposium (IV), Nagoya, Japan (pp. 1188-1193). doi: 10.1109/IV48863.2021.9575687.

- Chapter 3 Halakoo, M., Yang, H., Karbasi, A., Farid, Y., Ucar, S., & Oguchi, K. (Prepared for submission). Quantifying the Impact of Bottlenecks on Network Performance and Macroscopic Fundamental Diagram Using Clustering and Decision Tree Prediction Models.

- Chapter 4 Halakoo, M., Yang, H., Farid, Y., Ucar, S., & Oguchi, K. (Prepared for submission). Enhancing Macroscopic Fundamental Diagram Accuracy in the Presence of Bottlenecks Through Machine Learning-based Scatter Classi cation and Correction.

- Chapter 5 Halakoo, M., Yang, H., & Abdulsattar, H. (2023). Heterogeneity Aware Emission Macroscopic Fundamental Diagram (e-MFD). Sustainability, 15(2), 1653. doi: 10.3390/su15021653.
- Chapter 6 Halakoo, M., Yang, H., Biswas, A., & Azer, P. (Submitted to the Journal of Cleaner Production). A Model Predictive Emission Controller Based on the Heterogeneity Aware Emission Macroscopic Fundamental Diagram.

Chapter 1

Introduction

1.1 Background and motivation

Transportation infrastructure serves as the lifeblood of a prosperous society, similar to the intricate network of veins within the human body. It enables individuals to engage in various activities, ful ll essential needs such as medical care, sustenance, and education and facilitates the overall growth of communities. Therefore, the transportation system's e ciency plays a pivotal role in fostering societal prosperity.

However, recent urban transportation encounters many challenges, prominently characterized by persistent tra c congestion and the surge in air and noise pollution. These challenges have emerged as a direct consequence of the rapid upsurge in urbanization and the subsequent escalation in vehicle ownership, leading to an overwhelming demand that surpasses the available transportation infrastructure capacity (Jifeng et al.; 2008; Metz; 2018).

The consequences of these challenges are far-reaching, encompassing not only adverse environmental implications, such as the contribution to climate change through the emission of CO_2 from internal combustion engines, but also posing a signi cant threat to human health by precipitating respiratory issues. Moreover, the economic toll is substantial, primarily due to the productivity losses incurred during prolonged tra c congestions, as evident in the case of Greater Toronto and Hamilton Area, where the annual cost of congestion is estimated to be approximately 6 billion dollars, encompassing both wasted productivity and fuel, greenhouse emissions and the accidents (Urban Transportation Task Force; 2012).

The severity of these repercussions made global entities such as the United Nations

initiate discussions on integrating sustainable development principles within transportation systems. This necessitates adopting more sustainable approaches to the design and operation of transportation networks, with a comprehensive focus on economic, environmental, and social sustainability aspects (Mihyeon Jeon and Amekudzi; 2005).

Economic sustainability in the context of transportation entails the prudent allocation of resources to ensure long-term viability and e ciency. Environmental sustainability involves minimizing the negative ecological footprint caused by transportation activities and reducing the impact on climate and ecosystems. Social sustainability emphasizes the equitable distribution of transportation resources to foster inclusivity and enhance the society's overall well-being (Mihyeon Jeon and Amekudzi; 2005).

As urban areas have space limitations, the conventional solution of road expansion to alleviate congestion becomes increasingly impractical. It fails to address the persistent mismatch between the rapidly increasing demand and the limited supply of transportation infrastructure (Metz; 2018). In this context, intelligent tra c control systems emerge as a promising alternative, o ering not only a cost-e ective solution compared to extensive road construction but also mitigating the challenges associated with latent demand, which often nulli es the projected bene ts of road expansion (Ravish and Swamy; 2021; P eiderer and Dietrich; 2017).

Moreover, the recent discourse on autonomous vehicles has catalyzed conversations about their potential contributions to transportation networks. Considering short reaction times (Mahmassani; 2016), autonomous vehicles appear as a promising solution to address the limitations inherent in human-driven vehicles, particularly safety concerns (Rahman and Thill; 2023). Nonetheless, at the initial stages of their adoption, these vehicles are expected to interact with human-driven counterparts. This necessitates a comprehensive examination of the implications at various market penetration rates to inform future planning (Rahman and Thill; 2023; Halakoo and Yang; 2021).

Given the interconnected nature of these discussions, investing in intelligent tra c control systems not only serves the immediate goal of achieving sustainable transportation but also provides valuable insights for the seamless integration of connected autonomous vehicles into the transportation landscape, facilitating a comprehensive understanding of their impacts throughout di erent stages of adoption until fully autonomous transportation becomes a reality.

However, implementing such systems poses challenges, ranging from data requirements to the complexities of state estimation and control optimization (Ravish and

Swamy; 2021), which will be discussed in the following paragraphs.

1.2 Tra c state estimation and data requirements

The process of state estimation serves as a critical prerequisite for executing control actions within transportation networks. It involves determining the current congestion levels, enabling the assessment of crucial factors such as travel time, speed, out ow (trip completion rate), and Level of Service (LoS). While local measurement devices like loop detectors aid in estimating congestion levels on road segments, the complexities escalate signi cantly as the scale of state estimation expands across larger networks. One of the foremost challenges encountered in tra c state estimation within large-scale networks revolves around the issue of data availability, exacerbated by the inherently dynamic nature of tra c demand. This is particularly evident in urban environments, where the availability of a limited number of tra c sensors poses a signi cant issue to accurate state estimation. Furthermore, acquiring Origin-Destination (OD) information for comprehensive state estimation presents challenges, primarily characterized by the high costs associated with conducting OD studies and the subsequent analytical processes. Consequently, the demand for a robust, dependable, and e cient tra c state estimation tool has become indispensable in developing and successfully implementing intelligent tra c control systems, especially within complex urban settings.

The Macroscopic Fundamental Diagram (MFD) emerges as a promising solution to the challenges mentioned earlier. Originating from the work of (Godfrey; 1969) and subsequently rede ned by (Daganzo and Geroliminis; 2008), the MFD represents a concave and crisp curve that delineates the intricate relationship between network density and ow within a uniformly loaded network. Empirical studies conducted by (Geroliminis and Daganzo; 2008) observed this fundamental relationship in diverse urban landscapes, ranging from Yokohama to San Francisco. They observed a low scattered relationship between network ow and density when the highly scattered individual loop detector measurements were aggregated. Other studies such as (Tsubota et al.; 2014) and (Loder et al.; 2017) con rmed the relationship for cities such as Brisbane, Australia; Bern, Switzerland; London, UK; and Madrid, Spain. Due to its inherent characteristics, in- uenced by network properties, control policies, and driver behaviours, the MFD is an invaluable tool for accurate state estimation within complex transportation networks.

(Geroliminis and Daganzo; 2008) also showed a direct relationship between the circulating ow and the out ow. This nding was important from the practical point of

Figure 1.1: Emperical MFD from simulation

view since measuring the out ow was a nontrivial task in practice, while circulating ow could be easily measured by loop detectors. Therefore, MFD could be used for measuring trip completion rates in the network. These unique characteristics also resulted in various applications of MFD for tra c management, such as perimeter control (Keyvan-Ekbatani et al.; 2012), safety evaluation (Alsalhi et al.; 2018), and planning implications (Ortigosa and Menendez; 2014), etc. Notably, studies such as (Ortigosa et al.; 2015) have demonstrated that even with a minimal coverage of as low as 25 percent of the network links, the MFD can e ectively estimate the network state with a reasonable degree of accuracy, further underscoring its signi cance in addressing the challenges of state estimation within large-scale transportation systems.

In the literature, various methods, including an analytical method based on variational theory (Daganzo and Geroliminis; 2008), an empirical method with local tra c measurements or tra c simulation (Buisson and Ladier; 2009) were used to calibrate the MFD. Figure 1.1 shows an empirical MFD.

Also, a macroscopic emission model becomes indispensable to implement an emission controller hinged upon the MFD. This model estimates emission levels and consequently formulates control actions that work to minimize emissions. Using microscopic models

and aggregating their results is an option for macroscopic emission control; this process requires high computational power and detailed vehicle trajectories. On the other hand, common macroscopic models were shown to be more computationally e cient while compromising accuracy (Halakoo et al.; 2023; Saedi et al.; 2020). The Emission Macroscopic Fundamental Diagram (eMFD) extends the MFD to encompass the relationship between speed (or ow), density, and emissions within a network (Barmpounakis et al.; 2021; Saedi et al.; 2020). This framework o ers insights into how varying tra c conditions in uence emissions, facilitating emission reduction strategies and improving air quality. The eMFD can be tailored to model emissions such as CO_2, NO_x, and PM. It quanti es the impact of tra c condition changes on emission levels and sheds light on the emissions' sensitivity to congestion (Barmpounakis et al.; 2021; Halakoo et al.; 2023). As tra c density escalates and speeds decline, the e-MFD reveals heightened emissions from idling and stop-and-go tra c patterns. This understanding is a cornerstone for devising tra c control strategies that alleviate congestion and reduce emissions. Recent research underscores the e-MFD's robustness and accuracy for emission estimation, demonstrating comparable precision to microscopic models while o ering practical advantages for large-scale applications (Halakoo et al.; 2023).

1.3 Tra c control

E ective tra c control is essential to prevent gridlock within the network. While local tra c control algorithms o er solutions for the immediate vicinity of intersections, they deal with a signi cant challenge, particularly in handling the oversaturated congestion and spillback phenomena (Kouvelas et al.; 2017). These algorithms often overlook protecting the upstream of the protected intersection, thereby limiting their e cacy (Sirmatel and Geroliminis; 2021). Moreover, the applicability of these methods diminishes as network complexity grows alongside the network's expansion in size. Developing real-time tra c controllers for urban regions poses additional challenges, including those related to the previously discussed state estimation problem and the inherent complexities embedded within the network structure.

An alternative approach that has recently captured the attention of researchers for large-scale networks is perimeter control. This innovative strategy involves partitioning the network into homogeneously congested regions and regulating the ow transfer between these regions to optimize overall network performance. The existing body of literature demonstrates the e ectiveness of such strategies in single-region (Keyvan-Ekbatani

et al.; 2012) or monocentric congestion pocket networks, as well as in multi-centric congestion pockets (Ramezani et al.; 2015; Geroliminis et al.; 2012). Furthermore, studies have revealed that coupling perimeter control with local controllers can enhance the bene ts of perimeter control (Keyvan-Ekbatani et al.; 2019), o ering a more holistic approach to tra c management within complex networks.

Criticism directed at perimeter control within the literature often revolves around the potential creation of congestion pockets along the boundaries of the protected region due to the gating of vehicles. However, recent studies such as (Keyvan-Ekbatani et al.; 2021) have indicated that issues of this nature can be mitigated by implementing appropriate green time distributions among the border intersections, thereby facilitating a smoother ow of tra c and minimizing the risk of congestion hotspots. Despite its potential advantages, perimeter control relies heavily on the MFD for modelling tra c dynamics and optimizing control actions. While MFD o ers the advantages mentioned previously, including the simplicity of tra c state representation, its application can encounter challenges, such as the emergence of hysteresis loops especially when the network deals with bottlenecks (Buisson and Ladier; 2009; Tsubota et al.; 2014). Hence, a nuanced understanding and careful consideration of the hysteresis phenomena are imperative to ensure the successful deployment of perimeter control strategies within complex urban networks.

Moreover, previous research has underscored the potential for integrating environmental objectives into perimeter control (Ingole et al.; 2020), leading to an enhanced focus on promoting environmental sustainability within transportation systems. By aligning tra c control strategies with eco-friendly objectives, perimeter control can reduce carbon and other emissions, foster energy e ciency, and promote sustainable transportation practices.

1.4 Research objectives

The primary aim of this thesis is to utilize the MFD in conjunction with innovative control strategies to enhance performance and reduce the environmental impacts of urban passenger transportation infrastructure. By enhancing performance, measurable improvements can be achieved, such as reducing delays and emissions, ultimately fostering a more sustainable transportation infrastructure management. In line with this overarching goal, the following research objectives have been outlined for this project:

1. Investigate the transformation of network performance across multiple stages during the introduction of connected autonomous vehicles within various urban transportation network layouts. This investigation aims to uncover the dynamic interplay between the integration of connected autonomous vehicles and the corresponding shifts in network performance, encompassing factors such as tra c ow, congestion, and overall network e ciency.

2. Analyze the hysteresis phenomena and its correlation with bottleneck attributes to understand how di erent bottleneck attributes in uence the occurrence and magnitude of hysteresis. By studying the intricate relationship between bottleneck characteristics and hysteresis, this objective seeks to o er valuable insights into the underlying mechanisms contributing to tra c delays and network ine ciencies.

3. Propose modi cations to the MFD in the presence of bottlenecks to enhance its accuracy and reliability. This objective involves re ning the MFD to accommodate the complexities associated with bottlenecks, thereby ensuring its applicability as a robust modelling tool for accurately representing tra c dynamics and optimizing control actions within urban transportation networks.

4. Formulate a perimeter controller correction algorithm that integrates bottleneck information and incorporates appropriate modi cations to the MFD based on the discerned relationship between bottleneck attributes and the potential associated hysteresis or errors. This algorithm aims to enhance the precision and e cacy of perimeter control strategies, o ering a more re ned approach to mitigating congestion and optimizing tra c ow within complex urban environments.

5. Propose a robust and e cient macroscopic emission estimation model that can e ectively account for variations in network congestion distribution, including directional demand uctuations and the in uence of bottlenecks. By devising an advanced emission estimation model, this objective seeks to provide a comprehensive framework for accurately assessing and mitigating the environmental footprint of transportation systems.

6. Develop a comprehensive perimeter control strategy to minimize the environmental footprint of transportation systems. By integrating environmental considerations into the design and implementation of perimeter control, this objective aims to promote eco-friendly transportation practices, reduce carbon emissions, and contribute to developing sustainable urban transportation networks.

1.5 Dissertation organization

This section provides a concise overview of this dissertation's various chapters and their relevance to the research objectives.

Chapter 1 presents a comprehensive background and motivation for the research, outlining the core objectives of the dissertation. It establishes the foundational framework for the subsequent chapters, providing an organizational structure that delineates the key themes and methodologies employed throughout the research.

Chapter 2 directly addresses the rst research objective by thoroughly analyzing the implications of introducing connected autonomous vehicles at varying penetration rates on the MFD across diverse street layouts. This chapter o ers valuable insights into the potential bene ts of di erent network con gurations in supporting the seamless integration and sustainable adoption of connected autonomous vehicles within urban transportation systems.

Chapter 3, in response to the second research objective, focuses on a comprehensive analysis of the hysteresis phenomena induced by bottlenecks within transportation networks. It clusters the manifestations of hysteresis and introduces a robust classi cation model capable of predicting speci c clusters based on distinct bottleneck characteristics, thus providing a nuanced understanding of the complex dynamics in uencing tra c ow and network performance.

Chapter 4 collectively addresses the second, third, and fourth research objectives by proposing advanced modi cations to the MFD in the presence of bottlenecks, integrating the ndings and decision tree model developed in Chapter 3. These modi cations enhance the MFD controller's robustness, accuracy, and e ciency by e ectively incorporating bottleneck information and optimizing the control strategies to mitigate congestion and streamline tra c ow within complex urban environments.

Chapter 5 addresses the fth research question by introducing the heterogeneity-aware eMFD, a robust model capable of accurately estimating emission levels, even in directional demand variations and bottleneck presence. Notably, this model closely approximates microscopic emission models but with signi cantly reduced computational costs and data requirements. As a result, it becomes a viable option for conducting emission estimations within large-scale transportation networks.

Finally, Chapter 6 focuses on addressing the sixth research objective by leveraging the insights and methodologies outlined in Chapter 5 to implement the eMFD in minimizing emissions and improving overall network performance. By integrating the eMFD into a comprehensive perimeter control strategy, this chapter aims to foster eco-friendly transportation practices and contribute to developing sustainable and environmentally conscious urban transportation networks.

It is noteworthy that Chapter 2 has been published in an esteemed peer-reviewed conference proceeding, underscoring the signi cance and relevance of the research ndings in urban transportation infrastructure. Additionally, the ndings presented in Chapter 5 have been published in a distinguished, high-quality, peer-reviewed journal, contributing valuable insights to the existing knowledge in sustainable transportation practices. The content of Chapters 3, 4 are prepared for submission to high-quality journals, and the content of Chapter 6 has been submitted to a high-quality peer-reviewed journal, re ecting the meticulous research and rigorous methodologies employed throughout the dissertation. Collectively, these chapters form a cohesive research body, providing a holistic understanding of the complexities inherent in urban transportation systems and o ering innovative solutions to address pertinent challenges.

It is important to note that while these chapters contribute to a comprehensive research narrative, there might be a certain degree of overlap to ensure the completeness and coherence of each manuscript (chapter). Figure 1.2 illustrates the sequential structure of the dissertation and consolidates how each chapter's objectives contribute to the overarching research framework.

Figure 1.2: Thesis structure

Bibliography

Alsalhi, R., Dixit, V. V. and Gayah, V. V. (2018). On the existence of network macroscopic safety diagrams: Theory, simulation and empirical evidence, PloS one 13(8): e0200541.

Barmpounakis, E., Montesinos-Ferrer, M., Gonzales, E. J. and Geroliminis, N. (2021). Empirical investigation of the emission-macroscopic fundamental diagram, Transportation Research Part D: Transport and Environment 101: 103090.

Buisson, C. and Ladier, C. (2009). Exploring the impact of homogeneity of tra c measurements on the existence of macroscopic fundamental diagrams, Transportation Research Record 2124(1): 127 136.

Daganzo, C. F. and Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban tra c, Transportation Research Part B: Methodological 42(9): 771 781.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental ndings, Transportation Research Part B: Methodological 42(9): 759 770.

Geroliminis, N., Haddad, J. and Ramezani, M. (2012). Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach, IEEE Transactions on Intelligent Transportation Systems 14(1): 348 359.

Godfrey, J. (1969). The mechanism of a road network, Tra c Engineering & Control 8(8).

Halakoo, M. and Yang, H. (2021). Evaluation of macroscopic fundamental diagram transition in the era of connected and autonomous vehicles, 2021 IEEE Intelligent Vehicles Symposium (IV), IEEE, pp. 1188 1193.

Halakoo, M., Yang, H. and Abdulsattar, H. (2023). Heterogeneity aware emission macroscopic fundamental diagram (e-mfd), Sustainability 15(2): 1653.

Ingole, D., Mariotte, G. and Leclercq, L. (2020). Minimizing network-wide emissions by optimal routing through inner-city gating, Transportation research part D: Transport and Environment 86: 102411.

Jifeng, W., Huapu, L. and Hu, P. (2008). System dynamics model of urban transportation system and its application, Journal of Transportation Systems engineering and information technology 8(3): 83 89.

Keyvan-Ekbatani, M., Carlson, R. C., Knoop, V. L. and Papageorgiou, M. (2021). Optimizing distribution of metered tra c ow in perimeter control: Queue and delay balancing approaches,Control Engineering Practice 110: 104762.

Keyvan-Ekbatani, M., Gao, X. S., Gayah, V. V. and Knoop, V. L. (2019). Tra c-responsive signals combined with perimeter control: investigating the bene ts,Transportmetrica B: Transport Dynamics .

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I. and Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating,Transportation Research Part B: Methodological 46(10): 1393 1403.

Kouvelas, A., Saeedmanesh, M. and Geroliminis, N. (2017). Enhancing model-based feedback perimeter control with data-driven online adaptive optimization, Transportation Research Part B: Methodological 96: 26 45.

Loder, A., Ambühl, L., Menendez, M. and Axhausen, K. W. (2017). Network features and mfd parameters, 17th Swiss Transport Research Conference (STRC 2017), IVT, ETH Zurich.

Mahmassani, H. S. (2016). 50th anniversary invited article autonomous vehicles and connected vehicle systems: Flow and operations considerations, Transportation Science 50(4): 1140 1162.

Metz, D. (2018). Tackling urban tra c congestion: the experience of london, stockholm and singapore,Case Studies on Transport Policy 6: 494 498.

Mihyeon Jeon, C. and Amekudzi, A. (2005). Addressing sustainability in transportation systems: de nitions, indicators, and metrics, Journal of infrastructure systems 11(1): 31 50.

Ortigosa, J. and Menendez, M. (2014). Tra c performance on quasi-grid urban structures, Cities 36: 18 27.

Ortigosa, J., Menéndez, M. and Tapia, H. J. C. L. (2015). Study on the number and location of measurement points for an mfd perimeter control scheme: a case study of zurich, EURO Journal on Transportation and Logistics 3: 245 266.

P eiderer, R. and Dietrich, M. (2017). New roads generate new tra c, The Earthscan Reader on World Transport Policy and Practice, Routledge, pp. 177 181.

Rahman, M. M. and Thill, J. (2023). Impacts of connected and autonomous vehicles on urban transportation and environment: a comprehensive review, Sustainable Cities and Society 96: 104649.

Ramezani, M., Haddad, J. and Geroliminis, N. (2015). Dynamics of heterogeneity in urban networks: aggregated tra c modeling and hierarchical control, Transportation Research Part B: Methodological 74: 1 19.

Ravish, R. and Swamy, S. R. (2021). Intelligent tra c management: a review of challenges, solutions, and future perspectives, Transport and Telecommunication Journal 22: 163 182.

Saedi, R., Verma, R., Zockaie, A., Ghamami, M. and Gates, T. J. (2020). Comparison of support vector and non-linear regression models for estimating large-scale vehicular emissions, incorporating network-wide fundamental diagram for heterogeneous vehicles, Transportation Research Record 2674 (5): 70 84.

Sirmatel, I. I. and Geroliminis, N. (2021). Stabilization of city-scale road tra c networks via macroscopic fundamental diagram-based model predictive perimeter control, Control Engineering Practice 109: 104750.

Tsubota, T., Bhaskar, A. and Chung, E. (2014). Macroscopic fundamental diagram for brisbane, australia: empirical ndings on network partitioning and incident detection, Transportation Research Record 2421 (1): 12 21.

Urban Transportation Task Force (2012). The high cost of congestion in canadian cities, Council of Ministers Transportation and Highway Safety .

Chapter 2

Evaluation of Macroscopic Fundamental Diagram Transition in the Era of Connected and Autonomous Vehicles

The content of this chapter is the manuscript text for publication under the following citation:

Halakoo, M. and Yang, H. (2021). Evaluation of macroscopic fundamental diagram transition in the era of connected and autonomous vehicles. 2021 IEEE Intelligent Vehicles Symposium (IV).

2.1 Abstract

Introducing connected and autonomous vehicles (CAVs) could bring practical solutions to the existing challenges with transportation infrastructures, such as accidents and congestion. However, the transition to the era of CAVs would be gradual, and it could be expected that both CAVs and human-driven vehicles (HDVs) would exist in the network for some time, which could change the fundamental properties of urban networks. In this paper, the impact of CAVs on the macroscopic fundamental diagram (MFD) is analyzed with microscopic tra c simulations, and the sensitivity analysis of market penetration rates of CAVs and network con gurations is conducted. The analysis shows that one-way grid networks o er the most accessible and resilient environment during various phases of CAV introduction. Moreover, introducing CAVs improves the aggregated network performance and accessibility (measured by trip completion rate) of HDVs.

2.2 Introduction

Connected and autonomous vehicles with capabilities such as shorter reaction time and tighter headways seem like a promising solution for current challenges with transportation infrastructures, such as accidents and tra c congestion. With the recent advances in communication technology, implementing CAVs looks feasible soon. In some countries, such as the Netherlands, under certain conditions, CAVs are allowed on public roads (KPMG; 2019). Considering their continuous sensing and almost instant reaction time advantage (Mahmassani; 2016), CAVs can reduce human errors, which have been the major cause of crashes (Fagnant and Kockelman; 2015). In addition, their shorter reaction times result in closer spacing between vehicles to increase road capacities (Mahmassani; 2016). Moreover, CAV-enabled adaptive routing can lead to the optimal spread of vehicles in the network to reduce the risk of gridlock (Daganzo et al.; 2011).

Ideally, in a fully connected and autonomous environment where 100% of vehicles are CAVs, no tra c signal or road signs would be required to operate transportation systems. However, removing all human-driven vehicles would not happen overnight, i.e., in the early stages of CAV development, CAVs and HDVs have to share roads. Therefore, proper planning for the transition process is important, speci cally in urban areas with limited road expansion space. An existing study proved that even at low market penetration rates (MPRs), CAVs can improve tra c stability and network throughput (Talebpour and Mahmassani; 2016). However, the impacts of network layout on network performance in the mixed environment of CAVs and HVs have not been well investigated.

This paper aims to study the impacts of di erent MPRs of CAVs on the MFD of gird networks with common con gurations (i.e. two-way, two-way with prohibited left turns, and one-way). This will aid policymakers in e ectively planning for e cient use of the limited urban space and achieving the highest capacity throughout the transition to 100% CAVs. To quantitatively assess the impact, MFD relating the average network accumulation (number of vehicles inside the network) to network production (total distance travelled by vehicles per unit of time, which is a measure of tra c ow) is utilized. MFD not only allows for network performance evaluation in an aggregated manner but is also independent of the tra c demand (Geroliminis and Daganzo; 2008), which makes it appropriate for control applications. In this paper, CAVs with di erent MPRs are loaded to the same grid network with the same demand but di erent network con gurations (i.e., one-way, two-way, two-way with banned left turns) to discover the optimal network layout with the highest trip serving capacity and resiliency.

The rest of this paper is organized as follows. Section 2.3 reviews the related works to this study. Section 2.4 introduces the simulated network settings, vehicle types, and tra c scenarios. Section 2.5 discusses the impact of CAVs on MFD resilience and network accessibility with microscopic tra c simulations. Finally, Section 2.6 concludes the ndings of this paper and the recommendations for future research.

2.3 Related Work

The research about macroscopic tra c monitoring can be traced back to the 1960s (Wardrop; 1968), (Godfrey; 1969). However, it was not well investigated until the last decade when (Daganzo and Geroliminis; 2008) suggested if the network is redundant, homogeneous, uniformly loaded, and minimally a ected by turning vehicles, the average network ow (or production) is a function of network density (or occupancy). Later, (Geroliminis and Daganzo; 2008) suggested the existence of the MFD by eld data and showed that when the highly scattered link fundamental diagrams were aggregated, the scatters disappeared, and points showed a clear relationship between network density and ow. Other studies empirically showed the existence of MFD by analyzing real traf- c data for Brisbane, Australia (Tsubota et al.; 2014), Bern, Switzerland, London, UK, and Madrid, Spain (Loder et al.; 2017). (Geroliminis and Daganzo; 2008) also showed a direct relationship between the circulating ow and the out ow. This nding was important from the practical point of view since measuring the out ow was a nontrivial task in practice, while circulating ow could be easily measured by loop detectors. Therefore, MFD could be used for measuring trip completion rate and accessibility in

the network. These unique characteristics also resulted in various applications of MFD for tra c management, such as perimeter control (Keyvan-Ekbatani et al.; 2012), safety evaluation (Alsalhi et al.; 2018), and planning implications (Ortigosa and Menendez; 2014), etc. In the literature, various methods, including an analytical method based on variational theory (Daganzo and Geroliminis; 2008), an empirical method with local tra c measurements (Buisson and Ladier; 2009), and a trajectory-based method, were used to calibrate MFD. (Courbon and Leclercq; 2011) compared di erent methods for MFD calculation and concluded that the analytical MFD is the upper bound of empirical MFD.

Given that MFD is a property of the network, some studies in the literature investigated the impacts of di erent network con gurations on the MFD and trip-serving capacities. (Gayah and Daganzo; 2012) analytically compared the capacity of one-way and two-way networks and concluded that the average trip length was a signi cant factor in determining the trip serving capacity of the two networks. In other words, when the trip lengths were short, the two-way layout resulted in higher trip completion and lower lost time, and one-way networks were o set by more circuity. Banning the left turns also led to higher trip serving in two-way networks, even when trips were long. (DePrator et al.; 2017) investigated the impact of prohibiting left turns on improving urban network e ciency, and it showed that the bene t of left-turn movement prohibition on improving the e ciency of the network (considering network exit function (NEF) and MFD as proxies for that purpose) depended on the density of the network. When the network was loaded lightly or extremely heavily, the prohibition of left turns did not increase the trip completion rate. Hence, they proposed that the left-turn accommodation policy should be based on network accumulation to achieve the most e cient results. (Ortigosa et al.; 2015) studied the impacts of various grid network con gurations, including two-way (TW), two-way with prohibited left turns (TWL), and one-way (OW), on NEF. The study showed that the NEF ow for TW networks was smaller than OW and TWL, and the length of the left-turn pocket was an e ective factor in the performance of the TW networks, meaning TW networks with shorter left-turn pockets (5m) gridlocked faster than the longer (25m) ones. In addition, with a higher Kirchho parameter, TWL networks outperform in average speed, delay and trip length, which was attributed to the lack of redundant paths in TWL networks. Moreover, (Ortigosa and Menendez; 2014) studied the impact of link removal in a TW grid network on network performance indicators such as total travel time and showed that removing central links was more detrimental to network performance than peripheral link removal.

The literature also studied the impact of the technologies of connectivity and autonomy, from adaptive driving and routing to fully autonomous vehicles, on network e ciency indicators, such as trip completion rate and maximum throughput. (Saberi et al.; 2015) showed that providing real-time information about the tra c condition (i.e., adaptive driving) could improve network trip completion rate by more than 10% when 100% adaptive driving was applied. However, the improvement from 30% to 100% adaptivity was very trivial, and more scatters were observed in the MFD. (Talebpour and Mahmassani; 2016) studied the impact of CAVs on ow stability and throughput, and they found that at low MPRs, combining autonomy and connectivity could signi cantly increase ow stability. Moreover, increasing MPRs reduced the scatters of MFD and increased maximum throughput, and autonomous vehicles only contributed to higher throughput compared to connected vehicles at the same MPR. Moreover, (Makridis et al.; 2018) expanded the scope of studying the impact of CAVs from a highway segment to a highway network in Antwerp, Belgium, and showed that at high demand levels and higher MPR, CAVs' shorter gaps improved the network's average harmonic speed. (Gu and Saberi; 2019) investigated the impact of autonomous vehicles (AVs) on the MFD of the idealized two-ring network. They concluded that AVs' turning and merging behaviours had negative impacts on the stability of the network, and it worsened as the probability of turn increased. In conclusion, they suggested that AVs without connectivity, despite increasing network ow, led to increasing network instability due to their merging and turning maneuvers (even less frequent ones).

The aforementioned review indicated that it was very challenging to understand the relationship between network con guration and network trip serving capacity under a mixed CAV and HDV environment. To ll this gap, this paper will use microsimulation tools to model HDV sand CAVs' behaviour.

2.4 Microscopic Simulation Analysis

To understand the impact of CAVs on MFD, especially network trip serving capacity, under di erent network con gurations, a grid network is designed with a microscopic tra c simulator, Aimsun (Aimsun; 2020), that was used in the literature for simulating the mixed ow of HDVs and CAVs (Makridis et al.; 2018; Kouvelas et al.; 2017). Both HDVs and CAVs and their behaviours are modelled in the simulation. The settings of the networks, features of HDVs and CAVs, and simulation scenarios are described in the rest of this section.

2.4.1 Simulated Network

In this subsection, a 5X5 square grid network is designed, and Figure 2.1a shows the geometry of the network. The lengths of all links are set as 200 meters, and 60 centroids are placed in the middle of the links to emulate trip generation/ attraction along the route (Ortigosa et al.; 2015).

The capacity of each lane is set as 900 vph. In the simulation, all links of the OW network are modelled as two-lane sections, while they are set as one-lane links in each direction for both TW and TWL networks. In addition, 20m left-turn pockets are considered for the TW network, and their speed limit is set to 50 km/h.

Moreover, the network is loaded with a uniform demand (30,000 veh), which gradually increases during the rst hour and remains constant in the second hour. No vehicles are added to the network in the third hour to allow the network to recover from congestion. The demand pro le is shown in Figure 2.1b. Aimsun dynamic tra c assignment module with logit model (with scale parameter 1) is applied for trip assignment. The costs of links are updated every 5 minutes for route choices.

In addition, all intersections are controlled by xed time signals with a cycle length of 60 seconds and a yellow time of 2 seconds. In the TW network, the four-legged intersections are controlled with four phases: two through movements and right turns (20 s) and two protected left turns (6 s). The three-legged intersections are controlled with three phases in the TW network: through movement and right turn (20 s), protected left turn (6 s), and right and left turns from the perpendicular link (26 s). In the TWL and OW network, a simple two-phase signal (two through and right-turn movements 28 s) is applied. All three-legged intersections in TWL and OW networks are modelled as yellow boxes, with which vehicles do not enter the intersection unless there is enough gap ahead.

2.4.2 Features Of HDVs and CAVs

This subsection describes the behaviours of both HDVs and CAVs simulated in Aimsun. Three distinct features of HDVs and CAVs are described, including their deterministic behaviour, real-time routing, and shorter reaction time. Table 2.1 summarizes the parameter values for the two types of vehicles.

In the simulation, both HDVs and CAVs are modelled based on the Gipps car-following model, the default car-following model in Aimsun. The reaction time of CAVs is set as 0.2 s and 0.8 s for HDVs. This idea is based on (Kouvelas et al.; 2017) approach

(a) Simulated network and accident

(b) Simulated demand pro le

Figure 2.1: Grid Network

for modelling AVs which assumes that they drive su ciently similar to HDVs only with some enhanced competencies such as shorter reaction time. Also, the minimum clearance at stop is reduced from 1 m for HDVs to 0.5 m for CAVs (similar to (Atkins; 2016; Lu et al.; 2020)). The settings allow CAVs to quickly respond to tra c changes on roads and maintain a closer gap with their leaders. In addition, the stochasticity in parameters, such as minimum/maximum acceleration/ deceleration and speed limit acceptance, is removed from CAVs to allow them to accurately control dynamics (Kouvelas et al.; 2017). Moreover, CAVs deploy the cooperative adaptive cruise control (CACC), and the maximum platoon size is ten vehicles.

Both HDVs and CAVs implement dynamic user equilibrium (DUE) for routing behaviours, and the results are saved as a path assignment for all designed experiments. In addition, the CAVs are marked as en-route vehicles, which can change their route after assignment to a route, while HDVs follow the path being assigned at their entry based on the recently updated link travel time. Moreover, the distribution of CAVs in each origin-destination (OD) pair is assumed to be uniform, i.e., the numbers of HDVs and CAVs are determined based on the de ned MPRs.

Table 2.1: Comparison of HDV and CAV features

Feature (unit)	HDV	CAV
Reaction time (s)	0.8	0.2
Reaction time at stop (s)	1.2	0.2
Reaction time for the vehicle ahead of the queue (s)	1.6	0.2
Platoon size (veh)	-	10
Speed limit acceptance	1.1	1
Maximum desired speed (km/h)	110	110
Maximum acceleration (m/s^2)	3	3
Normal deceleration (m/s^2)	4	4
Maximum deceleration (m/s^2)	6	6
Clearance (m)	1	0.5
Sensitivity factor	1	1
Vehicle length (m)	4	4
Speed gain	-	0.0125
Distance gain (s^1)	-	0.45
Time gap leader (s)	-	1.5
Time gap follower (s)	-	0.6
Lower gap threshold (s)	-	1.5
Higher gap threshold (s)	-	2
Assignment method	DUE (Logit)	DUE (Logit)
Update cycle for routing information (min)	5	5
En-route percentage (%)	0	100

2.4.3 Simulation Scenarios

In this paper, the impact of MPRs of CAVs on MFD will be analyzed to understand the changes in network properties at di erent stages of CAV development. In the simulation, three di erent MPRs of CAVs (20%, 50%, 80%) are modelled with the same demand under three grid network con gurations (OW, TW, TWL). In the analysis, tra c information, including link ow and density and completed trip rates, is collected every minute.

For the estimation of MFD, the trip production ((vehicle km/min), which is the sum of all distance travelled in the network, and the number of vehicles inside the network (veh) are calculated at every minute (Geroliminis et al.; 2007). Also, for comparing the resiliency of di erent network con gurations in handling disruptions of road events, an incident closing one lane in the center of the network is modelled. The length of the incident is 6 meters, and it is active during [60-105] minutes (45 minutes duration). Also, no accident broadcasting or roadside units are simulated to inform the users. Nonetheless, based on updated path travel time, the CAVs can adapt their route to avoid the a ected links. Since this paper focuses on understanding MFD at mixed CAV and HDV environments, the scenario with 100% MPR will not be discussed.

2.5 Results and Discussion

This section discusses the results in the aforementioned tra c scenarios to understand MFD at di erent MPRs of CAVs and network con gurations. The resilience of the network and vehicle accessibility are also discussed.

2.5.1 MFDs of Grid Networks

Figure 2.2 A-D show the MFD from four di erent MPRs of CAVs (0%, 20%, 50%, 80%) under three network con gurations without deploying the incident, while Figure 2.2 E-H illustrate the other four scenarios with the incident. The gures demonstrate that increasing CAVs' MPRs results in higher network capacities, which can be explained by shorter reaction times and a smaller minimum clearance of CAVs. Moreover, the OW network o ers the highest capacity, followed by the TWL network, and TW is the worst. In the simulation, the MFD from the OW network only shows the free- ow state (even in the presence of accidents), while the other two networks gridlock when the accident happens despite having the same demand levels.

Moreover, when there is no incident, the capacity di erence between the OW and TWL networks decreases at higher MPRs, which indicates the support of real-time routing with CAVs can increase network mobility. However, when the MPR reaches 80%, the standard deviation of density in the TWL network reaches its highest value. Around 400 vehicles are still in the network while network production is zero which suggests some parts of the network are gridlocked (see Figure 2.2 D and Figure 2.3 D). Also, the congested branch of MFD in the TW network looks oscillatory at MPR=80%, which indicates the instability of the network. Capacity improvement in the OW might seem to decrease at higher MPRs. However, it can be seen that with the increase in MPRs, the critical accumulation decreases while production either remains the same or increases, which indicates that network capacity is increased. In this case, since the demand was not su ciently high, the network recovered before reaching its new higher capacity. The capacity of the intersections is signi cantly higher in OW and TWL networks due to more green time (28 s vs 20 s) and less lost time (4 s vs 8 s) compared to the TW network. Therefore, OW and TWL networks demonstrate higher production values than the TW network.

2.5.2 Hysteresis Loops and Resilience

In this subsection, the hysteresis of MFD and the network's resilience are evaluated. In transportation networks, hysteresis loops are caused during the loading and recovery of the network when the congestion is unevenly distributed (Daganzo et al.; 2011). The area inside the hysteresis loop shows network congestion dissipation and gridlock recovery, i.e., the smaller the area, the faster recovery from gridlock (Fakhrmoosavi et al.; 2020).

Figure 2.2 shows that the hysteresis loop size in the OW network is considerably smaller than the other two networks, i.e., the OW network can better recover from congestion. The higher path redundancy in the OW network can explain this. With that, the OW network can recover from congestion faster and generate a smaller loop area, while the other two networks are gridlocked. The result is similar to the study in (Ortigosa et al.; 2015). In addition, according to (Ortigosa and Menendez; 2014), in symmetric grid networks with uniform demand, most of the congestion is concentrated in the center. In that sense, if a central link is temporarily closed by an incident during a busy time, which is also the most disruptive scenario in a grid network, the network with high path redundancy, i.e., the OW network, can minimize the impact of the incident with the alternative shortest paths. Higher MPRs of CAVs can better take advantage of the path redundancy, which explains the smaller hysteresis loop size. Moreover, when the MPR increases from 50% to 80%, the hysteresis loop of MFD in the OW network

Figure 2.2: Network MFD: A (0% MPR no accident), B (20% MPR no accident), C (50% MPR no accident), D (80% MPR no accident), E (0% MPR with accident), F (20% MPR with accident), G (50% MPR with accident), H (80% MPR with accident)

Figure 2.3: Network density standard deviation: A (0% MPR no accident), B (20% MPR no accident), C (50% MPR no accident), D (80% MPR no accident), E (0% MPR with accident), F (20% MPR with accident), G (50% MPR with accident), H (80% MPR with accident)

slightly expands, which implies the homogeneity of congestion is slightly decreased. The other two networks also show a similar pattern. This observation can be attributed to the extra turning maneuvers caused by adaptive routing. This issue is anticipated to be resolved if more advanced signal control strategies or nely tuned routing parameters are applied.

The multi-valuedness in the right tail of MFD with an accident in TWL networks (see Figure 2.2 F-H) indicates this network is not completely gridlocked and some vehicles still have mobility. This results from the uneven distribution of congestion during the recovery and loading process. Increasing MPRs of CAVs can push this multi-valuedness point to the right, which means the network can sustain higher density, i.e., the network has higher resilience.

2.5.3 Accessibility

In this subsection, the network accessibility is evaluated with the simulations. The accessibility can be measured with the trip completion rate (Geroliminis and Daganzo; 2008). It has been demonstrated that the OW network with higher shortest path redundancy has higher trip completion rates (Ortigosa and Menendez; 2014), and the production in MFD is highly correlated to the rate (Geroliminis et al.; 2007). Hence, the system accessibility improvement is larger in the OW network with a higher MPR. Nevertheless, the accessibility bene t to HDVs is not well investigated.

Figure 2.4 shows the trip completion rates (the value is normalized by dividing the MPRs of HDVs) under di erent MPRs and network con gurations. The gure veri es that the accessibility for HDVs can be increased at higher CAV MPRs. However, the bene t is considerable for the OW and TW networks with and without incident. While the TWL network gains insigni cant accessibility improvement at higher MPRs, the occurrence of an accident highlights the bene ts, especially at high MPRs. The result demonstrates the accessibility bene ts of connectivity and autonomy on both CAVs and HDVs under di erent network con gurations. In conclusion, the accessibility improvement for HDVs with higher MPRs is more pronounced in the TW network.

2.6 Conclusions

This paper investigated the e ect of network con gurations under di erent MPRs of CAVs on urban network performance. The results showed that the capacity and resilience of the network to disruption increased at higher MPRs of CAVs. Also, the OW

Figure 2.4: Trip completion: A (OW, no accident), B (TW, no accident), C (TWL, no accident), D (OW, with accident), E (TW, with accident), F (TWL, with accident)

network brought up more bene ts (i.e., higher accessibility and capacity and a better response to disruptions). In addition, the accessibility of HDVs throughout increasing MPRs was discussed. The results showed that the increasing MPR bene ted CAVs and improved accessibility for HDVs. Especially when disruption happens, the bene ts are more pronounced. The results suggested the OW network worked the best throughout the transition from the existing situation to the fully connected and autonomous environment.

Based on the ndings in this paper, there are some recommendations for future work. First, for simplicity, this paper only investigated a xed time control plan for a grid network. The transition process to connected and autonomous environments with more advanced tra c control systems and vehicle-to-infrastructure (V2I) communication will be studied to understand the bene ts of CAVs more accurately. Moreover, a sensitivity analysis of demand levels and patterns (e.g., directional demand) will be conducted to understand the impact of CAVs on MFD at di erent congestion levels. Also, the e ects of the locations and types of incidents will be investigated in the future. Aimsun microscopic simulator was used in this study, but the replication and comparison of results with other simulators, such as VISSIM and SUMO, will be investigated.

Last but not least, due to the limited implementation of CAVs at the time of writing this paper, the experiments are all done in the simulated environment. The replication of the experiment with real CAVs would be an interesting topic for future works.

Bibliography

Aimsun (2020). Aimsun next 20 users manual.

Alsalhi, R., Dixit, V. V. and Gayah, V. V. (2018). On the existence of network macroscopic safety diagrams: Theory, simulation and empirical evidence, PloS one 13(8): e0200541.

Atkins, W. (2016). Research on the impacts of connected and autonomous vehicles (cavs) on tra c ow, Stage 2: Tra c Modelling and Analysis Technical Report .

Buisson, C. and Ladier, C. (2009). Exploring the impact of homogeneity of tra c measurements on the existence of macroscopic fundamental diagrams, Transportation Research Record 2124(1): 127 136.

Courbon, T. and Leclercq, L. (2011). Cross-comparison of macroscopic fundamental diagram estimation methods, Procedia-Social and Behavioral Sciences 20: 417 426.

Daganzo, C. F., Gayah, V. V. and Gonzales, E. J. (2011). Macroscopic relations of urban tra c variables: Bifurcations, multivaluedness and instability, Transportation Research Part B: Methodological45(1): 278 288.

Daganzo, C. F. and Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban tra c, Transportation Research Part B: Methodological 42(9): 771 781.

DePrator, A. J., Hitchcock, O. and Gayah, V. V. (2017). Improving urban street network e ciency by prohibiting con icting left turns at signalized intersections, Transportation Research Record2622(1): 58 69.

Fagnant, D. J. and Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations,Transportation Research Part A: Policy and Practice 77: 167 181.

Fakhrmoosavi, F., Saedi, R., Zockaie, A. and Talebpour, A. (2020). Impacts of connected and autonomous vehicles on tra c ow with heterogeneous drivers spatially distributed over large-scale networks,Transportation research record 2674(10): 817 830.

Gayah, V. V. and Daganzo, C. F. (2012). Analytical capacity comparison of one-way and two-way signalized street networks,Transportation Research Record2301(1): 76 85.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental ndings, Transportation Research Part B: Methodological42(9): 759 770.

Geroliminis, N., Daganzo, C. F. et al. (2007). Macroscopic modeling of tra c in cities, Transportation Research Board 86th Annual Meeting number 07-0413, No. 07-0413.

Godfrey, J. (1969). The mechanism of a road network,Tra c Engineering & Control 8(8).

Gu, Z. and Saberi, M. (2019). E ects of turning and merging on network tra c instability: A simulation-based analysis of human-driven and autonomous vehicles,CoRR .

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I. and Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating,Transportation Research Part B: Methodological46(10): 1393 1403.

Kouvelas, A., Perrin, J.-P., Fokri, S. and Geroliminis, N. (2017). Exploring the impact of autonomous vehicles in urban networks and potential new capabilities for perimeter control, 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) , IEEE, pp. 19 24.

KPMG, I. (2019). 2019 autonomous vehicles readiness index: assessing countries' preparedness for autonomous vehicles.

Loder, A., Ambühl, L., Menendez, M. and Axhausen, K. W. (2017). Network features and mfd parameters, 17th Swiss Transport Research Conference (STRC 2017), IVT, ETH Zurich.

Lu, Q., Tettamanti, T., Hörcher, D. and Varga, I. (2020). The impact of autonomous vehicles on urban tra c network capacity: an experimental analysis by microscopic tra c simulation, Transportation Letters 12(8): 540 549.

Mahmassani, H. S. (2016). 50th anniversary invited article autonomous vehicles and connected vehicle systems: Flow and operations considerations, Transportation Science 50(4): 1140 1162.

Makridis, M., Mattas, K., Ciu o, B., Raposo, M. A., Toledo, T. and Thiel, C. (2018). Connected and automated vehicles on a freeway scenario. e ect on tra c congestion and network capacity, 7th Transport Research Arena 13.

Ortigosa, J. and Menendez, M. (2014). Tra c performance on quasi-grid urban structures, Cities 36: 18 27.

Ortigosa, J., Menendez, M. and Gayah, V. V. (2015). Analysis of network exit functions for various urban grid network con gurations, Transportation Research Record 2491(1): 12 21.

Saberi, M., Mahmassani, H. S. and Zockaie, A. (2015). Network capacity, tra c instability, and adaptive driving: ndings from simulated urban network experiments, EURO Journal on Transportation and Logistics 3(3): 289 308.

Talebpour, A. and Mahmassani, H. S. (2016). In uence of connected and autonomous vehicles on tra c ow stability and throughput, Transportation research part C: emerging technologies 71: 143 163.

Tsubota, T., Bhaskar, A. and Chung, E. (2014). Macroscopic fundamental diagram for brisbane, australia: empirical ndings on network partitioning and incident detection, Transportation Research Record 2421(1): 12 21.

Wardrop, J. (1968). Journey speed and ow in central urban areas,Tra c Engineering & Control 8(8).

Chapter 3

Quantifying the impact of bottlenecks on network performance and Macroscopic fundamental diagram using clustering and decision tree prediction models

3.1 Abstract

Understanding how bottlenecks a ect tra c networks is crucial but challenging because it typically requires detailed origin-destination data. The Macroscopic Fundamental Diagram (MFD) o ers a promising solution because it does not rely on speci c demand data. Previous research showed that MFDs can re ect the presence of bottlenecks, but there has not been a clear classi cation of the types of MFDs a ected by bottlenecks. Such a classi cation could enhance our understanding of how bottlenecks impact network performance.

Our study used microsimulators to model two networks and applied unsupervised machine-learning techniques to analyze the resulting MFDs. Our ndings contribute to the existing literature by identifying two major types of MFDs that emerge when bottlenecks are present: "high-impact" and "low-impact" MFDs. We named them based on their characteristics, such as hysteresis area, residual density, and capacity.

Additionally, we developed a decision tree model to uncover the relationship between bottleneck characteristics and the resulting MFD type. Interestingly, our analysis revealed that while the structures of the models in the two networks were similar, they were not identical. Both networks exhibited low-impact MFDs when bottlenecks occurred during the unloading phase. Furthermore, the MFD type during the loading phase depended on factors such as the duration of the bottleneck and the number of lanes. Surprisingly, bottleneck location did not seem to make a signi cant di erence.

Our study goes beyond previous work by shedding light on the intricate interactions between bottleneck characteristics that were often simpli ed in earlier research. Ultimately, we provide practical policy recommendations for network managers based on our model's insights.

3.2 Introduction

Bottlenecks in transportation networks invariably lead to delays, which can manifest as static or mobile impediments. A static bottleneck can arise from various factors, such as road construction or bottlenecks caused by certain operations. Conversely, a mobile bottleneck is typically caused by slower-moving vehicles, such as trucks or tractors. These bottlenecks restrict the capacity of the a ected road section, and if the incoming tra c surpasses the bottleneck's capacity, queues start forming. Extensive queues can extend

backward into upstream road sections in extreme cases, resulting in a network-wide gridlock.

Quantifying the impact of bottlenecks on the network is imperative for mitigating their adverse e ects. For instance, during the planning phase of a construction operation, it is essential to consider its potential impact on the network. Similarly, network managers must prioritize their responses to multiple bottlenecks, allocating more resources to those with a greater network-wide impact. However, obtaining detailed origin-destination (OD) matrices is essential to determine both the in ow to the bottleneck and its repercussions on network performance. Furthermore, OD matrices can uctuate with changes in demand, adding complexity to the problem.

MFD, which has garnered increasing attention from researchers, emerges as a promising solution to address these challenges due to its independence from demand (Daganzo and Geroliminis; 2008). Its reliance on network geometry (Daganzo and Geroliminis; 2008; Halakoo and Yang; 2021), tra c control (Daganzo; 2007), and routing (Saberi et al.; 2015) makes the MFD a robust tool for monitoring network performance. Additionally, existing literature has demonstrated its e ectiveness in detecting macroscopic-level bottlenecks(Tsubota et al.; 2014). Moreover, once the e ects of bottlenecks on the MFD are known, it becomes possible to devise tra c control systems such as the one introduced in (Ramezani et al.; 2015) that can adjust signal settings to manage in ow at bottleneck locations.

However, it is noteworthy that, to the best of the authors' knowledge, there has been no systematic categorization of MFDs in the presence of bottlenecks in existing literature. Such categorization would simplify practitioners' understanding of the scale of bottleneck impact on network performance. Additionally, developing a predictive model that can categorize resulting MFDs based on bottleneck characteristics remains an unexplored avenue in the literature despite the advantages outlined in the preceding paragraphs. In this study, we are trying to answer the following questions:

- Is it possible to nd meaningful categories for MFDs a ected by bottlenecks?
- How are bottleneck characteristics related to the resulting MFD category?
- What similarities and di erences can be expected regarding the e ects of bottlenecks on the MFD of di erent networks?
- What policies can be recommended based on a model that relates bottleneck characteristics to MFD?

to address our research inquiries, we employed a combination of unsupervised and supervised machine learning algorithms. Initially, we conducted simulations on two distinct transportation networks: a 9x9 grid network (area = 3.2 km^2) and a designated portion of Manhattan, NY, with prede ned bottlenecks. Utilizing metrics derived from the MFD, such as the hysteresis area (the area inside the MFD) and residual density (density at the end of simulation to show the extent of gridlock), we formulated an unsupervised machine learning model (speci cally, k-means) to categorize the various impacted MFDs. Subsequently, after assigning labels to these MFD categories, we developed a supervised machine learning model (in this case, a decision tree) to predict the MFD categories based on bottleneck characteristics. Ultimately, we derived policy recommendations based on the insights gleaned from our models.

The outcomes presented in this paper possess practical implications in several domains, including construction planning, tra c control system development, and resource allocation for responding to bottleneck incidents. These applications aim to mitigate the adverse e ects of bottlenecks on network performance.

The subsequent sections of this paper follow a structured organization. Section 3.3 comprehensively reviews pertinent literature. Section 3.4 o ers a concise introduction to the MFD, the concept of the hysteresis loop, clustering, and classi cation methods. Section 3.5 outlines the speci cs of our simulated network and various tra c scenarios. Moving forward, Section 3.6 summarizes the simulation results and examines how bottlenecks impact MFDs. Section 3.7 is dedicated to discussing the implications of our ndings. Lastly, we present our conclusions and o er recommendations for future research in Section 3.8.

3.3 Literature review

MFD, initially introduced in the 1960s by Godfrey (Godfrey; 1969) and later reformulated by Daganzo in recent years (Daganzo and Geroliminis; 2008), maps network density to network ow. Daganzo's work (Daganzo and Geroliminis; 2008) suggested that the average network ow could be formulated as a function of network density (or occupancy) if a network is homogeneously congested. After validating the existence of the MFD by real data in Yokohama and San Francisco (Geroliminis and Daganzo; 2008), MFD attracted researchers' attention, leading to various proposed applications, such as tra c control (Keyvan-Ekbatani et al.; 2012), routing (Yildirimoglu et al.; 2015), taxi dispatch (Ramezani et al.; 2015), emission estimation (Halakoo et al.; 2023), and safety evaluation

(Alsalhi et al.; 2018). However, the homogeneous congestion distribution might not hold in reality. Buisson et al. (Buisson and Ladier; 2009) detected a hysteresis loop in the MFD of Toulouse, France, due to truck drivers' strike and intentional speed reduction, an example of a moving bottleneck. Simulation results in (Mazloumian et al.; 2010) showed the network ow should be a function of another factor in addition to the density that accounted for the homogeneity of congestion. They suggested the standard deviation of density as a congestion homogeneity indicator and showed its negative relationship with network ow. Saberi et al. (Saberi et al.; 2015) introduced the H index, which is the ratio of the di erence between the length of the congested and uncongested sections to the network's total length. Later, the generalized macroscopic fundamental diagram was introduced, which related network ow to network density and the standard deviation of density as a proxy for congestion homogeneity (Knoop and Hoogendoorn; 2013).

Gayah and Daganzo (Gayah and Daganzo; 2011) showed that networks are more prone to becoming heterogeneously congested in the unloading phase (when no additional load was added and the network recovered from the capacity state) than in the loading phase (when the network was initially empty and vehicles were added to the network before reaching the capacity). This behaviour resulted in a hysteresis loop in MFD, expressed as two ow branches for the same density range. Geroliminis and Sun (Geroliminis and Sun; 2011) suggested the synchronized occurrence of transient periods and capacity drop as another cause for hysteresis phenomena in freeways MFD. Hysteresis loops were further categorized as clockwise, counter-clockwise, and gure-8. Clockwise hysteresis happens when the ow in the unloading phase is lower than in the loading phase.

On the contrary, counter-clockwise hysteresis occurs when the ow during the loading phase is lower than the unloading capacity. This pattern is often observed upstream of a temporary bottleneck before and after removal. In contrast, gure-8 hysteresis combines clockwise and counter-clockwise hysteresis elements, where lower ow is observed before the onset of congestion, followed by higher ow after congestion dissipates (He et al.; 2015).

In a study by Shim et al. (Shim et al.; 2019), lower ow values were observed in the afternoon compared to the morning in Daegu, South Korea, for the same density range. The reduction in the ow during the afternoon peak was attributed to detouring drivers deviating from their usual routes, which did not lead to improved travel times for themselves or other travellers. In contrast, drivers exhibited more consistent route choices during the morning peak, resulting in higher ow rates in the MFD.

Saberi et al. (Saberi et al.; 2015) demonstrated that the hysteresis area diminishes as the percentage of adaptive drivers increases, indicating faster network recovery from congestion. Additionally, Halakoo et al. (Halakoo and Yang; 2021) illustrated that network con guration, such as one-way, two-way, and two-way with prohibited left turn links, can also in uence the reduction of hysteresis through adaptive routing.

Bottlenecks associated with capacity drops are considered non-recurrent disruptions in the network. Consequently, the presence of a bottleneck typically leads to the formation of a hysteresis loop in the MFD. Tsubota et al. (Tsubota et al.; 2014) reported anomalous scattering in the MFD for Brisbane, Australia, during a major bottleneck event. By aligning the scattered timestamps with bottleneck descriptions, such as detection and clearance times, they suggested that the regional MFD can capture the impact of a local bottleneck. Ji et al. (Ji et al.; 2015) observed di erent hysteresis patterns in the MFDs upstream and downstream of bottleneck locations on the M3 Paci c motorway in Brisbane, Australia. The upstream MFD (of motorway sections upstream of the bottleneck location) exhibited a counter-clockwise hysteresis, while a clockwise pattern was observed in the downstream MFD. The size of the hysteresis loop in the MFD was positively associated with the bottleneck's duration and the demand level.

In comparing the impact of inclement weather and accidents on the MFD of Seoul, South Korea, Schorr et al. (Schorr et al.; 2016) found that accidents had a more severe impact on MFD capacity reduction than inclement weather.

In another study, Amini et al. (Amini and Tilg; 2018) introduced the Critical Score (CS) of links, which quanti ed the loss of path redundancy after removing a link, as an explanation for the size of the hysteresis loop. They demonstrated that links with larger CS index values resulted in larger hysteresis areas during bottleneck occurrences. However, it is worth noting that in this study, the e ects of bottlenecks were modelled as permanent blockages throughout the entire simulation time (210 minutes), which might be overly conservative and could lead to a new user equilibrium. Additionally, the CS parameter was treated as a xed value, whereas CS can dynamically change depending on the applied routing strategy, particularly in scenarios involving dynamic tra c assignment.

As far as the authors know, only a few studies have focused on the e ects of isolated bottleneck characteristics, such as location or duration, primarily within freeway networks. However, in practice, multiple bottlenecks with diverse attributes can occur simultaneously. Consequently, this paper aims to develop a model that captures the

in uence of bottleneck characteristics, encompassing factors such as duration, location, and initiation time, on the MFD of urban transportation networks. Urban networks present added complexity compared to freeway networks due to multiple routing options.

3.4 Methods

3.4.1 MFD and hysteresis

MFD depicts the relationship between network-wide ow and density. Various network characteristics, including geometry, capacity, speed limits, and management strategies such as signal control, determine its form. Two main approaches have been proposed for deriving the MFD: the analytical method, which relies on network features and control policies without requiring eld measurements (Daganzo and Geroliminis; 2008), and the empirical method, which draws from eld or simulation data (Geroliminis and Daganzo; 2008; Tsubota et al.; 2014).

Previous research has demonstrated that the analytical method provides an upper bound for the empirical MFD (Loder et al.; 2019). As a result, the analytical approach is often applied in network design contexts, while the empirical method nds utility in policy analysis. In this study, we adopt the empirical method due to its potential to investigate the impact of bottlenecks on the MFD.

To derive the empirical MFD, gathering data on ow and density for each link at prede ned intervals is necessary. Subsequently, these ow and density values for each interval are aggregated to compute network-wide density and ow at that speci c time interval. Equations 3.1 and 3.2 show the estimation of network-wide ow and density (Saberi and Mahmassani; 2013; Tsubota et al.; 2014). These equations represent weighted averages of ow (q_i^t) and density (k_i^t) for each link, with weights being determined by the length of each link (l_i) and the number of lanes (n_i).

$$Q(t) = \frac{\sum_{i=1}^{N} q_i(t) \quad l_i}{\sum_{i}^{N} l_i} \tag{3.1}$$

$$K(t) = \frac{\sum_{i=1}^{N} k_i(t) \quad l_i}{\sum_{i}^{N} l_i} \tag{3.2}$$

Two key metrics, namely the hysteresis area and residual density, were employed to assess and compare the e ect of bottlenecks across di erent scenarios. The hysteresis

Figure 3.1: A Clockwise Hysteresis with Residual Density Kr

area, depicted as the dotted region in Figure 3.1, serves as an e ective indicator of system stability; a smaller hysteresis area signi es more e cient system recovery during the unloading phase, as highlighted in prior research (Fakhrmoosavi et al.; 2020). The area calculation relies on Equation 3.3, where $Q(k)$ represents the MFD, a function determined by the network's average density, and L represents the hysteresis loop. Additionally, a smaller maximum density (or hysteresis width) in the MFD corresponds to improved recovery from congestion, as established in previous work (Fakhrmoosavi et al.; 2020). Furthermore, the metric of residual density (K_r) provides insight into the extent of gridlock; a higher K_r value indicates a more extensive gridlock scenario.

$$A_{Hys} = \int_{L}^{l} Q(k)dk \tag{3.3}$$

3.4.2 Clustering

Clustering is a method in data analysis that involves partitioning a dataset into clusters based on the inherent patterns and similarities within the data points. It aims to identify distinct groups within a dataset, allowing for the discovery of underlying structures and relationships. Data points with similar characteristics are grouped through clustering, while dissimilar ones are separated into clusters. This process is crucial for various applications, including pattern recognition, image segmentation, market research, and anomaly detection (Bijuraj; 2013).

K-means clustering introduced by (MacQueen et al.; 1967) is one of the most widely used and e cient clustering methods. It operates by partitioning the dataset into K clusters, where K represents the predetermined number of clusters. The algorithm iteratively assigns data points to the nearest cluster centroid and recalculates the centroids based on the mean of the data points in each cluster. This process continues until the centroids stabilize, resulting in well-de ned clusters. K-means is advantageous due to its simplicity, scalability to large datasets, and computational e ciency, making it an attractive choice for various data mining and pattern recognition tasks. We used the k-means++ initialization method (Arthur and Vassilvitskii; 2007), which improves the algorithm's performance by enhancing the selection of initial cluster centroids.

3.4.2.1 Clustering performance metrics

To evaluate the performance of our clustering model, we employed the following metrics along with their corresponding equations and references:

1. Inertia , also known as the within-cluster sum of squares, quanti es the sum of squares of each data point's distance from its cluster center. Mathematically, it is expressed as:

$$\text{Inertia} = \sum_{i=1}^{N} \min_{\text{center}_j \in \text{centers}} \|x_i - \text{center}_j\|^2 \tag{3.4}$$

   where N is the number of data points, x_i represents a data point, and center_j denotes the cluster center. Lower inertia values indicate that data points are closely associated with their respective cluster centers. Increasing the number of clusters invariably reduces inertia, making it necessary to strike a balance and select the smallest number of clusters that minimizes inertia (Chavent; 1998).

2. The average silhouette index measures how well each data point is clustered. It ranges between -1 and +1, with values closer to +1 indicating that data points exhibit strong coherence within their clusters and distinct separation from neighbouring clusters. Conversely, values nearing -1 suggest that data points have been incorrectly assigned to clusters (Aranganayagi and Thangavel; 2007). The formula for the average silhouette index is as follows:

$$\text{Average Silhouette Index} = \frac{1}{N} \sum_{i=1}^{N} \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{3.5}$$

   N is the number of data points, a_i is the average distance from x_i to other points within the same cluster, and b_i is the smallest average distance from x_i to data

points in a di erent cluster.

3. The Calinski-Harabasz index , also known as the variance ratio criterion, measures the similarity of data points within the same cluster and the separation of a data point from neighbouring clusters. It is computed as follows:

$$\text{Calinski-Harabasz Index} = \frac{B_k}{W_k} \quad \frac{N \quad k}{k \quad 1} \tag{3.6}$$

Where B_k is the between-cluster dispersion, which represents the sum of the distances between the cluster centroids and the overall dataset mean, and W_k is the within-cluster dispersion, which represents the sum of the distances of individual data points from their respective cluster centroids. N is the total number of data points, and k is the number of clusters.

A higher Calinski-Harabasz index indicates better cluster cohesion and clearer separation from neighbouring clusters, representing a more favourable clustering result (Cali«ski and Harabasz; 1974).

4. The Davies-Bouldin index quanti es the similarity of each data point to its nearest neighbouring cluster. It ranges between 0 and 1, with lower values indicating better cluster cohesion and greater separation from neighbouring clusters. The formula for the Davies-Bouldin index is expressed as:

$$\text{Davies-Bouldin Index} = \frac{1}{N} \sum_{i=1}^{N} \max_{j \neq i} \left(\frac{S_i + S_j}{d(c_i; c_j)} \right) \tag{3.7}$$

where N is the number of data points, S_i and S_j are the average distances from x_i and x_j to the data points within their respective clusters, and $d(c_i; c_j)$ represents the distance between cluster centers c_i and c_j. Lower Davies-Bouldin index values indicate improved cluster cohesion and better separation from neighbouring clusters (Davies and Bouldin; 1979).

We utilized various clustering evaluation metrics to comprehensively assess cluster validity (Arbelaitz et al.; 2013). Our selection of the cluster validity index encompasses two types, namely optimization-like criteria, which are those where higher (maximization) or lower (minimization) values inherently indicate the best partitions (e.g., Calinski Harabasz index, Davies-Bouldin index, and average silhouette), and stopping rule criteria (like the elbow method), which can be employed to halt the hierarchical clustering algorithm (Vendramin et al.; 2010).

3.4.3 Classi cation

Classi cation, a key task in supervised machine learning, involves categorizing data points into prede ned classes or labels based on their features. The goal is to develop a model to accurately predict the class of new, unseen data points. Supervised classi - cation algorithms learn from labelled training data, where each data point is associated with a known class, enabling the model to generalize and make predictions on unseen data. This process is crucial for various applications, including spam detection, sentiment analysis, and medical diagnosis (Hastie et al.; 2009). The decision tree algorithm is popular in classi cation due to its interpretability, versatility, and capability to handle numerical and categorical data (Hastie et al.; 2009). Decision trees are a owchart-like structure, facilitating easy comprehension of the decision-making process (Breiman et al.; 1984). They can e ectively handle non-linear relationships between features and the target variable, making them suitable for complex datasets. Calibrating a decision tree model involves ne-tuning the parameters to improve its performance metric and ensure optimal generalization to unseen data. Several techniques can be employed for model calibration, including hyperparameter tuning, cross-validation, and ensemble methods. In this study, we used hyperparameter tuning for decision tree calibration. Adjusting the hyperparameters, such as the maximum depth, minimum samples for splitting, and leaf nodes, can control model complexity, preventing over tting and improving its predictive accuracy (Hastie et al.; 2009). One major limitation of hyperparameter tuning is the computational cost associated with exhaustively searching for the optimal combination of hyperparameters, especially in large datasets with numerous features. Grid search and random search, commonly used techniques for hyperparameter tuning, can be computationally expensive and time-consuming. We used grid search in this study to speed up the hyperparameter tuning process.

3.4.3.1 Classi cation performance metrics

We used precision, recall, F1 score, accuracy, and area under the receiver operating characteristic curve (ROC AUC) to evaluate the performance of our classi cation prediction model. These metrics are calculated as follows:

1. Precision

   Precision measures the fraction of true positive predictions to the sum of true positive and false positive predictions. Precision values range from 0 to 1, with higher values being better. An optimal precision score is close to 1, indicating a high proportion of true positive predictions relative to false positive predictions.

Precision is calculated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3.8}$$

2. Recall

   Recall shows the fraction of true positives to the sum of true positives and false negatives. Recall values also range from 0 to 1, with higher values being better. An optimal recall score is close to 1, signifying a greater proportion of true positives than false negatives. Recall is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3.9}$$

3. F1 Score

   The F1 score is the harmonic mean of precision and recall, providing a balanced measure. It ranges from 0 to 1, with higher values being better. An optimal F1 score is close to 1, re ecting a balanced trade-o between precision and recall, indicating high true positive rates and low false positive and false negative rates. F1 Score is calculated as:

$$\text{F1} = \frac{2 \quad \text{Precision} \quad \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.10}$$

4. Accuracy represents the ratio of true positives to the total number of predictions. Accuracy values range from 0 to 1, with higher values being better. An optimal accuracy score is close to 1, suggesting a high proportion of correct predictions among all predictions. Accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{True Positives}}{\text{Total Predictions}} \tag{3.11}$$

5. Area Under the Receiver Operating Characteristic Curve (ROC AUC)

   ROC AUC quanti es how well the model distinguishes between positives and negatives. ROC AUC values range from 0 to 1, with higher values being better. An optimal ROC AUC score is close to 1, indicating excellent discrimination between positive and negative cases. A value of 0.5 suggests random guessing, while values above 0.5 indicate better-than-random performance.

3.5 Numerical Experiment

3.5.1 Grid network

In this section, we conducted simulations on a 9x9 grid network covering an area of 3.24 km², with 200 m block sizes, using the Aimsun simulation software. This network size is similar to large cities' central business district areas worldwide, including the Downtown Vancouver, BC Central Business District (CBD) area (3.7 km^2) and Sydney, Australia (2.8 km^2). In this network, all intersections were represented as yellow boxes. The road sections had a speed limit of 50 km/h, with a capacity of 1800 vehicles per hour (or 900 vehicles per hour per lane). All road segments were designed as one-way, two-lane sections.

As illustrated in Figure 3.2b, the demand pro le involved incremental increases every 10 minutes during the rst hour. Throughout the second hour, demand increments remained constant. Starting from the third hour, no additional load was introduced to the network. Our tra c signals consisted of xed signals with a 60-second cycle length.

Dynamic Tra c Assignment (DTA) was utilized for vehicle routing, with route costs being updated every ve minutes to align with the network's average travel time in the base scenario. A logit model with a scale parameter of one was employed to distribute vehicles based on the most recent travel cost of available paths for each origin-destination (OD) pair. Prede ned OD routes, en-route options, and initial path assignments were intentionally excluded from the simulations to maintain consistency.

To introduce bottlenecks and record network states, we utilized the Aimsun API module. Bottleneck locations were randomly selected from links within and outside the shaded area in Figure 3.2a. The shaded area encompassed the network's busiest links, resulting from the shortest paths traversing central links in a grid network under uniform demand (Ortigosa and Menendez; 2014). We considered two bottleneck start times: during the loading phase (60 minutes after the simulation began) and during the unloading phase (180 minutes after simulation commencement). Three bottleneck durations 15 minutes, 35 minutes, and 60 minutes were examined alongside one-lane and two-lane blockage bottleneck con gurations. We generated 24 scenarios by combining 2 locations, 2 start times, 3 durations, and 2 lane con gurations. Each scenario was replicated 15 times with di erent random seeds, resulting in 360 simulations, each recording ow and density data to construct MFDs and calculate hysteresis area and residual density.

(a) Network geometry

(b) Demand pro le

Figure 3.2: 9x9 Grid network

This comprehensive experimental setup allowed us to investigate the impact of various bottleneck characteristics on network performance and derive valuable insights from the resulting MFDs.

3.5.2 Real-world network: Manhattan, New York

In this paper, we also conducted simulations in a section of Manhattan, New York, USA, to validate our models. Figure 3.3a provides an overview of the network, with the shaded area highlighting the most congested region, akin to the central section of the grid network depicted in Figure 3.2a. This network covers a total area of 2.3 km^2 and features a maximum speed limit of 50 km/h. The number of lanes varies from one to ve.

Figure 3.3b portrays the demand pro le, which follows a similar pattern to that observed in a grid network. During the initial hour, demand gradually increased, with consistent demand increments in the subsequent hour. Starting from the third hour, no additional load was introduced to the network. The demand was generated using Randomtrip.py, a tool within the SUMO simulator, resulting in a more exible and homogeneous network compared to one with xed routes (Karbasi and O'Hern; 2022). This network comprised 170 tra c signals and was generated using OSMWebWizard,

(a) Manhattan, NY, USA

(b) Demand pro le

Figure 3.3: Manhattan Network

a tool that extracts various network details, including geometry and tra c light data, and generates a SUMO network le. As per the extracted data, all tra c lights in this network operate as actuated tra c lights. The driving behaviour in this context is modelled using the Krauss car-following model (Krauss; 1998), a component of the SUMO simulator. The parameters associated with this car-following model, including time headway, minimum distance headway, acceleration, and deceleration, have been con gured to their default values as de ned in SUMO.

Also, similar to the grid network, we randomly selected bottleneck locations from congested and non-congested areas, as illustrated in Figure 3.3a. We explored two distinct starting times for these bottlenecks: one during the loading phase, occurring 60 minutes after the simulation started, and another during the unloading phase, taking place 180 minutes after the simulation began. Additionally, we examined three di erent bottleneck durations: 15 minutes, 30 minutes, and 60 minutes. Furthermore, we introduced one-lane and two-lane blockages to account for varying bottleneck sizes. We investigated 24 distinct scenarios, and for each scenario, we conducted 10 replications, each with di erent random seeds, resulting in the generation of 240 MFDs.

Figure 3.4: A) Hysteresis area probability distribution, B) Residual density probability distribution, C) Hysteresis area and residual density scatter plot (Grid network)

3.6 Results

3.6.1 Grid Network

3.6.1.1 Clustering analysis

We applied the k-means clustering algorithm to reveal underlying patterns within the MFD data. Our approach to constructing and interpreting the clustering model involved several steps. Initially, we visually inspected the distributions of our clustering variables, namely hysteresis area and residual density. Subsequently, we employed hyperparameter tuning techniques and multiple cluster validity indices to determine the optimal number of clusters. Lastly, we developed the clustering model and pro led the clusters based on their distinctive attributes.

Figure 3.4 shows the hysteresis area and residual density plots. Figure 3.4 A shows the distribution of the hysteresis area, which exhibits two peaks. Similarly, Figure 3.4 B shows that the residual density distribution also manifests two peaks. In Figure 3.4 C, the scatter plot illustrates hysteresis area values plotted against the corresponding residual density, revealing two dense clusters with two connecting points. We selected two clusters as our initial guess, considering the presence of two peaks in the distributions and two dense spots in the scatter plots.

We explored a range of cluster counts, varying from 2 to 7, and evaluated several clustering validity indices, including inertia, average silhouette, Calinski-Harabasz index, and Davies Bouldin index. Our primary aim was to determine the optimal number of clusters for the MFD dataset. Figure 3.5 presents the clustering indices for cluster

counts ranging from 1 to 7. Notably, the average silhouette, Calinski-Harabasz index, and Davies-Bouldin index require a minimum of two clusters, rendering the 1-cluster case irrelevant to their analysis. Figure 3.5 A showcases the inertia values for various cluster counts. This graph, often called the elbow plot, aids in identifying the optimal number of clusters. The point at which the graph exhibits the rst signi cant decrease in inertia is considered the optimal cluster count. In this case, the elbow plot indicates a substantial drop in inertia at two clusters, while the decrease in inertia for three or more clusters is minimal.

Figure 3.5 B underscores that the highest average silhouette index is attained with two clusters, indicating the suitability of the two-cluster model. Figure 3.5 C suggests that the Calinski-Harabasz index peaks at seven clusters, followed by two clusters. However, since the other indices do not support the seven-cluster model, we favoured the two-cluster model based on the Calinski-Harabasz index. Figure 3.5 D reveals that the lowest Davies-Bouldin index is associated with two clusters, further supporting the choice of this cluster count. After carefully considering all index results, we selected the two-cluster model as our nal choice.

Figure 3.6 illustrates cluster pro ling. While our k-means model was developed using hysteresis area and residual density, we incorporated capacity and the other two variables for cluster pro ling to comprehensively describe our clusters.

Figure 3.6 A reveals that the two clusters exhibit distinct hysteresis area values, with cluster 0 having larger values than cluster 1. In Figure 3.6 B, cluster 1's bottlenecks exhibit lower residual density than those in cluster 0. Figure 3.5 C shows that the bottlenecks in cluster 1 possess larger capacities than those in cluster 0.

It is worth noting that the inclusion of capacity in the cluster pro ling was a deliberate choice to enhance the description of our clusters. However, as the k-means model was primarily trained on hysteresis area and residual density, we observed that seven points in Figure 3.5 C for cluster 0 exhibited capacities similar to those of cluster 1. This occurred because the model was not explicitly trained to distinguish data points based on capacity. We experimented with di erent k-means models, incorporating hysteresis area, residual density, and capacity and various combinations of these variables. The results consistently indicated that the most e ective model, based on the clustering indices discussed earlier, was the one that included only hysteresis area and residual density.

Figure 3.5: Clustering validation (grid network) indexes: A) inertia, B) average silhouette, C) Calinski-Harabasz, D) Davies Bouldin for the range of clusters (Grid Network)

Figure 3.6: Cluster pro ling (grid network) based on: A) hysteresis area, B) residual density, C) MFD capacity

Considering the collective results of all indices, we concluded that cluster 1 encompasses bottlenecks with lower residual densities and hysteresis areas while possessing higher capacities than the bottlenecks in cluster 0. Consequently, we chose to pro le cluster 1 as low-impact MFDs and cluster 0 as high-impact MFDs. Throughout the remainder of this paper, we refer to cluster pro les as MFD types. Notably, approximately 80% of the data points belonged to the low-impact clusters, while the remaining 20% resided in the high-impact cluster.

3.6.1.2 Classi cation prediction analysis

We developed a decision tree classi cation model based on the outcomes of our clustering model. Our model utilizes bottleneck characteristics, including duration, start time (phase), location, and the number of a ected lanes, to predict the corresponding MFD cluster (low-impact or high-impact). We divided the data into training and testing sets, employing a 70/30 split. The training set was used for hyperparameter tuning with ve-fold cross-validation. Hyperparameter tuning helped us nd the optimal minimum number of samples in each node, ranging from one to 30. This parameter de nes the minimum number of data points in each leaf node and controls tree depth. A larger minimum sample size in leaf nodes results in reduced tree depth and more generalized rules, while a smaller minimum sample size allows for taller trees. We used the F1 score for selecting the best hyperparameters in our grid search space. We chose the F1 score due to its e ectiveness in imbalanced classi cation problems where a large portion of data represents a speci c class (Chawla et al.; 2002). We also explored other metrics, such as ROC-AUC, precision, and recall, but found that they yielded similar results to the F1 score.

Figure 3.7: Classi cation model calibration and testing (grid network) A) tree depth hyperparameter tuning, B) Confusion matrix for the training set, C) confusion matrix for testing set

Model set	Precision	F1	Recall	Accuracy	ROC-AUC
Train	0.990	0.988	0.986	0.980	0.969
Test	0.989	0.983	0.978	0.972	0.961

Table 3.1: Model's performance metrics (grid network)

Figure 3.7 A demonstrates that the model exhibited minimal sensitivity to changes in the minimum sample size in leaf nodes, ranging from 1 to 15. We selected 15 as the minimum sample size in leaf nodes because it provides more examples to each leaf node while yielding the highest F1 score. Table 3.1 showcases the outstanding performance of our decision tree model on both the training and testing sets. Notably, the model's performance on the test set closely mirrored its performance on the training set, indicating that the model did not over t or under t the data (Goodfellow et al.; 2016). As a result, our model is likely to perform well on unseen data and exhibit good generalization.

Figure 3.7 B and C display the confusion matrices of the decision tree model for the training and test sets. These matrices illustrate that the model achieved high true positive and true negative values for both sets. In contrast, false negative and false positive values constituted a small fraction of the data, demonstrating the model's excellent performance.

The decision tree visualization in Figure 3.8 provides insights into the model's decision-making process. The branches on the right side of each upstream node represent outcomes when the condition in the upstream node is false, while the left branches represent

outcomes when the condition is true. In the decision tree, gini impurity measures how often a randomly chosen element from the set would be incorrectly labelled if it was randomly labelled according to the distribution of labels in the subset. A node's gini impurity is the weighted sum of the gini impurities of each class. The "value" in the decision tree leaves represents the class distribution for the samples that reach that leaf. It shows the count of samples from each class that belong to the leaf node.

The number of lanes is positioned at the top of the decision tree. Two downstream nodes are connected to the number of lanes. The left branch indicates that if the number of a ected lanes is less than two (equivalent to one, given that we had either one or two-lane closures), the resulting MFD is classi ed as low impact. The right branch of the rst node pertains to cases where two lanes were a ected. Subsequently, another node explains the outcome, indicating that the resulting MFD is classi ed as low impact if the bottleneck a ects two lanes during the unloading phase. The duration node determines the outcome if the bottleneck a ects two lanes during the loading phase. In the duration node, if the bottleneck a ects two lanes during the loading phase and the duration is 60 minutes, the resulting MFD is classi ed as high impact. If the duration is less than 60 minutes, a further evaluation is conducted: if the duration is 35 minutes, regardless of the location, the resulting MFD is high impact; otherwise (for a 15-minute duration), irrespective of the location, the resulting MFD is classi ed as low impact.

3.6.2 Real-world Manhattan Network

3.6.2.1 Clustering analysis

Following a methodology similar to the grid network analysis, we utilized distribution and scatter plots of hysteresis area and residual density to estimate the optimal number of clusters for the bottleneck MFDs within the Manhattan network. The results of this analysis are depicted in Figure 3.9. Remarkably, 3.9 A and B exhibited a bimodal distribution pattern in both hysteresis area and residual density, mirroring the observations made in the grid network. This clear bimodal distribution provided compelling evidence to select two clusters as a suitable choice for our analysis.

Figure 3.10 showcases the cluster pro les speci c to the Manhattan network. These pro les rea rm the existence of two distinct clusters. Cluster 0 is distinguished by its larger hysteresis area and higher residual density, while a smaller hysteresis area and lower residual density characterize Cluster 1. As a result, we assigned the label "low-impact" to cluster 0 and "high-impacts" to cluster 1 within the Manhattan network, aligning with our previous ndings in the grid network. Given our objective to assess the

Figure 3.8: Decision tree model (grid network)

Figure 3.9: A) Hysteresis area probability distribution, B) Residual density probability distribution, C) Hysteresis area and residual density scatter plot (Manhattan network)

Figure 3.10: Cluster pro ling (Manhattan) based on: A) hysteresis area, B) residual density, C) MFD capacity

transferability of results from the grid network to the Manhattan network, we con dently settled on a nal cluster count of 2 for this analysis.

3.6.2.2 Classi cation prediction analysis

We followed a parallel approach to the grid network analysis to establish a decision tree model for the Manhattan network and capture the intricate relationship between bottleneck characteristics and MFD types. As in the grid network analysis, we divided the data into training and test sets, maintaining a split ratio 70/30.

Figures 3.11 A and B illustrate the outcomes of our decision tree parameter hyper-tuning and the confusion matrix for the Manhattan network, respectively. Notably, the decision tree parameter hyper-tuning plot reveals that the model's performance remains unaltered when increasing the minimum sample size in the leaf nodes from 17 to 30, with

Figure 3.11: Classi cation model calibration and testing (Manhattan network) A) tree depth hyperparameter tuning, B) Confusion matrix for the training set, C) confusion matrix for the testing set

Model set	Precision	F1	Recall	Accuracy	ROC-AUC
Train	0.879	0.816	0.761	0.863	0.846
Test	0.920	0.793	0.697	0.833	0.823

Table 3.2: Model's performance metrics (Manhattan)

the F1 score remaining constant within this range. Consequently, we set the minimum samples in the leaf nodes at 17, although the nal model used 20 samples for the leaf nodes.

Table 3.2 provides a comprehensive overview of the model's performance metrics in the Manhattan network. The remarkably similar values observed between the training and test sets signify a reasonable t of the model, underscoring its capacity to generalize e ectively beyond the training data.

Figure 3.12 presents the decision tree model for the Manhattan network. As the model illustrates, bottlenecks during the unloading phase are associated with low-impact MFDs. This nding aligns with our observations in the grid network, where all two-lane bottlenecks during the unloading phase led to low-impact MFDs.

Like the grid network model, the number of lanes is positioned at the second level of the Manhattan network decision tree. In this context, two-lane bottlenecks, regardless of their speci c locations, result in high-impact MFDs during the loading phase. Conversely, for one-lane bottlenecks during the loading phase, similar to the grid network,

Figure 3.12: Decision tree model (Manhattan)

the duration of the bottleneck dictates the resulting MFD type. Speci cally, 60 and 30-minute bottlenecks during the loading phase are associated with high-impact MFDs, while 15-minute bottlenecks lead to low-impact MFDs.

Thus, although the Manhattan Network decision tree may not exactly replicate the grid network decision tree precisely, the structural consistency underscores the coherence of our observations between the two networks.

3.7 Discussion

Understanding the factors that in uence the shape of the MFD in urban transportation networks is essential for e ective tra c management and congestion mitigation. This study sought to investigate how various factors, including start time, duration, location, and the number of lanes a ected by bottlenecks, contribute to the resulting MFD types

in two distinct urban environments: a simpli ed grid network and the complex streets of Manhattan.

The clustering analysis based on the hysteresis area and residual in the grid and Manhattan network showed two types of MFDs that could be observed due to network bottlenecks. These two types represent two distinctive characteristics; one is associated with a larger hysteresis area and higher residual density, while the other encompasses smaller hysteresis areas and lower residual density. These ndings, for the rst time represented in the literature and apart from the research novelty, entail some important practical applications. First, they provide a simple what-if analysis regarding the impact of bottlenecks on network performance.

Moreover, this simple what-if analysis can be extended to controller design. For example, the previous research (Daganzo; 2007) has shown that congestion management strategies are essential for improving tra c ow. Our study extends this by identifying distinct MFD types associated with high and low-impact scenarios, providing insights into the e ectiveness of di erent congestion management approaches. As a result, one can calibrate more adaptive controllers, considering the e ects of the bottlenecks on the MFD. In such a controller, high-impact MFDs require more aggressive control strategies, while low-impact MFDs allow for more exible strategies.

The study in (Witte et al.; 2012) suggested that the conception of bottlenecks should be extended by incorporating other dimensions to tackle the cumulating and culminating e ects of bottlenecks. They also highlighted the importance of an integrative perspective on analyzing bottlenecks, which can add important insights to the present body of knowledge. In our study, we took the cumulating e ects of bottlenecks on network e ciency by exploring their impacts on the MFD, which has already been shown to be an advantageous tool for network performance measurement (Daganzo and Geroliminis; 2008; Geroliminis and Daganzo; 2008).

Also, the clustering analysis extends the ndings in the literature by demonstrating that high-impact MFDs can negatively a ect network e ciency, emphasizing the importance of proactive network design and management. Furthermore, regarding the environmental impacts of the bottlenecks, the literature shows a consensus on the direct relationship between congestion and emission levels (Halakoo et al.; 2023). Therefore, it entails high-impact MFDs associated with higher congestion levels and lower capacity and is most likely to entail higher emission amounts. Also, from an economic perspective, higher congestion levels past the critical state mean higher delays and economic

losses. Hence, network managers should take high-impact bottlenecks more seriously if they want to reduce the congestion cost.

In the grid network, the results clearly show how di erent factors impact MFD types. Notably, one-lane bottlenecks consistently led to low-impact MFDs, irrespective of their start time, duration, or location. This nding suggests that a single-lane bottleneck is insu cient to induce high congestion levels in a network with limited routing options and a simpli ed structure. For two-lane bottlenecks, the start time and duration emerged as crucial determinants of MFD types. Speci cally, bottlenecks during the unloading phase consistently resulted in low-impact MFDs. However, during the loading phase, longer durations (35 and 60 minutes) led to high-impact MFDs, while shorter durations (15 minutes) still resulted in low-impact MFDs. These ndings have practical implications for grid-like networks with constrained routing options. Transportation authorities in such environments may focus on strategies to manage tra c during the loading phase when longer-duration bottlenecks can lead to high congestion levels. Additionally, recognizing that one-lane bottlenecks typically result in low-impact MFDs can guide resource allocation and tra c management e orts.

With its intricate urban layout and abundant routing options, the Manhattan network presented a more complex set of ndings. First and foremost, all unloading phase bottlenecks consistently led to low-impact MFDs, emphasizing the importance of timing in this congested urban environment. For loading phase bottlenecks in Manhattan, duration and the number of lanes blocked played signi cant roles in determining MFD types. Notably, all two-lane bottlenecks during the loading phase resulted in high-impact MFDs, underlining the sensitivity of Manhattan's tra c to such congestion events. In contrast, for one-lane bottlenecks, the duration of the bottleneck became the key factor, with longer durations (60 and 30 minutes) causing high-impact MFDs. In comparison, shorter durations (15 minutes) led to low-impact MFDs. These ndings have profound implications for tra c management in complex urban environments like Manhattan. The consistent impact of unloading phase bottlenecks on low-impact MFDs highlights the importance of timely accident response and congestion mitigation measures. Furthermore, the sensitivity of Manhattan's tra c to two-lane bottlenecks during the loading phase underscores the need for proactive management and e ective communication during such bottlenecks. Additionally, recognizing the duration-based di erentiation in one-lane bottleneck impacts can guide emergency response strategies and tra c control measures in this bustling urban setting.

One notable aspect of this study is its contribution to the existing literature on the

impact of bottlenecks on MFD types. Previous studies often isolated these factors, discarding the interplay considered here. For instance, earlier research typically did not incorporate a duration factor (Amini et al.; 2023), with lanes blocked for the entire simulation time, or set the number of lanes statically to one or two. In another study, (Amini and Tilg; 2018) only considered the e ect of accident duration and concluded a longer duration always causes a more severe capacity drop in the MFD. In contrast, this study considered multiple levels for the bottleneck characteristics and created models to capture the interplay between these characteristics and their levels. This approach provides a more comprehensive understanding of how these factors change MFDs, o ering valuable insights for practical tra c management strategies in diverse urban environments.

In conclusion, this study has shed light on the nuanced factors in uencing MFD types in di erent urban environments. While the grid network's simplicity led to straightforward results, the complex network structure of Manhattan introduced variations and sensitivities that call for tailored tra c management approaches. These ndings provide valuable insights for urban planners, transportation authorities, and policymakers aiming to optimize tra c ow and mitigate congestion in diverse urban settings. Future research may further explore the interplay of these factors and their implications for sustainable urban mobility.

3.8 Conclusions

In this study, we quanti ed the impact of bottlenecks on network performance using the type of change in MFD they caused. Our MFD clustering analysis showed two major types of MFDs can be expected in the presence of bottlenecks. We named these two MFD types high and low impact based on their hysteresis area, residual density, and capacity. High-impact MFDs were associated with larger hysteresis areas, larger residual density, and lower capacity, which all relate to severely degrading network performance and slowing recovery. On the other hand, low-impact MFDs were associated with smaller hysteresis areas, residual density, and higher capacity, which indicate their lower impact on network performance. This is the rst study in the literature providing such a categorization based on an unsupervised machine learning algorithm for MFDs in the presence of bottlenecks. This categorization not only adds to the novelty of this paper but also can be used for quantifying the extent to which network performance is a ected by bottlenecks. Moreover, a decision tree model with these two MFD types was developed for two networks (a grid network and the Manhattan network) to nd

the relationship between bottleneck characteristics and the resulting MFD types. Our models highlight the importance of bottleneck prevention and clearing (in the case of accidents) during the loading phase when the network demand is high.

This paper simplifies assumptions, which shall be further investigated in the future. For example, currently, a spatially uniform demand for the sake of simplicity was considered. However, testing other demand patterns, such as directional demand, could be more interesting in understanding the real-world impact of bottlenecks. We also plan to replicate the experiments with different network sizes. It would be insightful to evaluate the impact of bottlenecks on regional MFDs to explore the network-level bottleneck impacts.

Bibliography

Alsalhi, R., Dixit, V. V. and Gayah, V. V. (2018). On the existence of network macroscopic safety diagrams: Theory, simulation and empirical evidence, PloS one 13(8): e0200541.

Amini, S. and Tilg, G. (2018). Estimating the effects of temporary bottlenecks on the capacity of urban arterial: An mfd-approach, 7th Symposium of the European Association for Research in Transportation (hEART 2018).

Amini, S., Tilg, G. and Busch, F. (2023). Macroscopic traffic dynamics in urban networks during incidents, Authorea Preprints .

Aranganayagi, S. and Thangavel, K. (2007). Clustering categorical data using silhouette coefficient as a relocating measure, International conference on computational intelligence and multimedia applications (ICCIMA 2007), Vol. 2, IEEE, pp. 13 17.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M. and Perona, I. (2013). An extensive comparative study of cluster validity indices, Pattern Recognition 46: 243 256.

Arthur, D. and Vassilvitskii, S. (2007). K-means++ the advantages of careful seeding, Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp. 1027 1035.

Bijuraj, L. (2013). Clustering and its applications, Proceedings of National Conference on New Horizons in IT-NCNHIT , Vol. 1, pp. 169 172.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classi cation and regression trees. wadsworth int, Group 37(15): 237 251.

Buisson, C. and Ladier, C. (2009). Exploring the impact of homogeneity of tra c measurements on the existence of macroscopic fundamental diagrams, Transportation Research Record 2124(1): 127 136.

Cali«ski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis, Communications in Statistics-theory and Methods 3(1): 1 27.

Chavent, M. (1998). A monothetic clustering method, Pattern Recognition Letters 19(11): 989 996.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique, Journal of arti cial intelligence research 16: 321 357.

Daganzo, C. F. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches, Transportation Research Part B: Methodological 41(1): 49 62.

Daganzo, C. F. and Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban tra c, Transportation Research Part B: Methodological 42(9): 771 781.

Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure, IEEE transactions on pattern analysis and machine intelligence PAMI-1 (2): 224 227.

Fakhrmoosavi, F., Saedi, R., Zockaie, A. and Talebpour, A. (2020). Impacts of connected and autonomous vehicles on tra c ow with heterogeneous drivers spatially distributed over large-scale networks, Transportation research record 2674(10): 817 830.

Gayah, V. V. and Daganzo, C. F. (2011). Clockwise hysteresis loops in the macroscopic fundamental diagram: an e ect of network instability, Transportation Research Part B: Methodological 45(4): 643 655.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental ndings, Transportation Research Part B: Methodological 42(9): 759 770.

Geroliminis, N. and Sun, J. (2011). Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks, Procedia-Social and Behavioral Sciences 17: 213 228.

Godfrey, J. (1969). The mechanism of a road network, Tra c Engineering & Control 8(8).

Goodfellow, I., Bengio, Y. and Courville, A. (2016). Deep learning, MIT press.

Halakoo, M. and Yang, H. (2021). Evaluation of macroscopic fundamental diagram transition in the era of connected and autonomous vehicles, 2021 IEEE Intelligent Vehicles Symposium (IV), IEEE, pp. 1188 1193.

Halakoo, M., Yang, H. and Abdulsattar, H. (2023). Heterogeneity aware emission macroscopic fundamental diagram (e-mfd), Sustainability 15(2): 1653.

Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction, Vol. 2, Springer.

He, Z., He, S. and Guan, W. (2015). A gure-eight hysteresis pattern in macroscopic fundamental diagrams and its microscopic causes, Transportation Letters 7(3): 133 142.

Ji, Y., Jiang, R., Chung, E. and Zhang, X. (2015). The impact of incidents on macroscopic fundamental diagrams, Proceedings of the Institution of Civil Engineers-Transport, Vol. 168, Thomas Telford Ltd, pp. 396 405.

Karbasi, A. and O'Hern, S. (2022). Investigating the impact of connected and automated vehicles on signalized and unsignalized intersections safety in mixed tra c, Future transportation 2(1): 24 40.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I. and Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating, Transportation Research Part B: Methodological 46(10): 1393 1403.

Knoop, V. L. and Hoogendoorn, S. P. (2013). Empirics of a generalized macroscopic fundamental diagram for urban freeways, Transportation Research Record 2391(1): 133 141.

Krauss, S. (1998). Microscopic modeling of tra c ow: Investigation of collision free vehicle dynamics, Technical report, DLR Deutsches Zentrum fuer Luft- und Raumfahrt e.V., Koeln (Germany).

Loder, A., Ambühl, L., Menendez, M. and Axhausen, K. W. (2019). Understanding tra c capacity of urban networks, Scienti c reports 9(1): 16283.

MacQueen, J. et al. (1967). Some methods for classi cation and analysis of multivariate observations, Proceedings of the fth Berkeley symposium on mathematical statistics and probability, Vol. 1, Oakland, CA, USA, pp. 281 297.

Mazloumian, A., Geroliminis, N. and Helbing, D. (2010). The spatial variability of vehicle densities as determinant of urban network capacity, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368 (1928): 4627 4647.

Ortigosa, J. and Menendez, M. (2014). Tra c performance on quasi-grid urban structures, Cities 36: 18 27.

Ramezani, M., Haddad, J. and Geroliminis, N. (2015). Dynamics of heterogeneity in urban networks: aggregated tra c modeling and hierarchical control, Transportation Research Part B: Methodological 74: 1 19.

Saberi, M. and Mahmassani, H. S. (2013). Empirical characterization and interpretation of hysteresis and capacity drop phenomena in freeway networks, Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, DC.

Saberi, M., Mahmassani, H. S. and Zockaie, A. (2015). Network capacity, tra c instability, and adaptive driving: ndings from simulated urban network experiments, EURO Journal on Transportation and Logistics 3(3): 289 308.

Schorr, J. P., Hamdar, S. H., Kang, S., Jang, K. and Kachouch, H. (2016). From structural equation modeling to macroscopic fundamental diagrams: investigating the impact of road segments safety on network level e ciency, 17th International Conference Road Safety On Five Continents (RS5C 2016), Rio de Janeiro, Brazil, 17-19 May 2016, Statens väg-och transportforskningsinstitut, pp. 1 17.

Shim, J., Yeo, J., Lee, S., Hamdar, S. H. and Jang, K. (2019). Empirical evaluation of in uential factors on bifurcation in macroscopic fundamental diagrams, Transportation Research Part C: Emerging Technologies 102: 509 520.

Tsubota, T., Bhaskar, A. and Chung, E. (2014). Macroscopic fundamental diagram for brisbane, australia: empirical ndings on network partitioning and incident detection, Transportation Research Record 2421 (1): 12 21.

Vendramin, L., Campello, R. J. G. B. and Hruschka, E. R. (2010). Relative clustering validity criteria: a comparative overview, Statistical Analysis and Data Mining: The ASA Data Science Journal 3: 209 235.

Witte, P. A., Wiegmans, B. W., van Oort, F. G. and Spit, T. J. (2012). Chokepoints in corridors: Perspectives on bottlenecks in the european transport network, Research in Transportation Business & Management 5: 57 66.

Yildirimoglu, M., Ramezani, M. and Geroliminis, N. (2015). Equilibrium analysis and route guidance in large-scale networks with mfd dynamics, Transportation Research Procedia 9: 185 204.

Chapter 4

Enhancing Macroscopic Fundamental Diagram Accuracy in the Presence of Bottlenecks Through Machine Learning-based Scatter Classi cation and Correction

4.1 Abstract

A macroscopic fundamental diagram (MFD) is an e ective network state estimation and tra c control tool. However, previous studies have shown the presence of bottlenecks could a ect the accuracy of MFD by inhomogeneous tra c distribution, which is one of the main assumptions of the MFD. In such cases, highly scattered MFDs with hysteresis loops can be expected. Previous studies tried to improve MFD applicability by including the congestion distribution term, which led to the introduction of the generalized MFD. Using machine learning, this study clustered the types of MFD scatters that can be observed in the presence of bottlenecks. Based on the MFD scatter types, correction terms for each type were developed to improve the accuracy of the MFD calibrated without bottlenecks. Including the correction term showed to reduce MFD errors caused by bottlenecks by up to %66. Then, a classi cation machine learning model was developed to predict the type of MFD scattered from bottleneck attributes. Using the classi cation, clustering, and error correction terms, we proposed a bottleneck management control policy that takes the bottleneck attributes and adjusts the controller accordingly. The results of this paper are highly impactful in developing accident management scenarios for network control centres.

4.2 Introduction

Macroscopic Fundamental Diagram (MFD), which is a concave curve showing the relationship between network ow and density, is an excellent solution for network state estimation (Daganzo and Geroliminis; 2008) and tra c control(Geroliminis et al.; 2012; Ramezani et al.; 2015). Although it is a great tool for tra c state estimation, previous studies have shown the presence of bottlenecks that cause inhomogeneous congestion distribution makes MFD more scattered and hysteric (Tsubota et al.; 2014; Buisson and Ladier; 2009).

Bottlenecks in transportation networks cause delays. Bottlenecks could be static or moving. A static bottleneck can be a result of a bottleneck or construction operation. A slow-moving vehicle, like a truck or a tractor, can cause a moving bottleneck. Bottlenecks reduce the road section capacity and consequently, if the in ow to the a ected section is larger than the bottleneck capacity, a queue is formed. Large queues can spill back to the upstream road sections and result in network gridlock. Also, the presence of bottlenecks could change tra c distribution as drivers try to dodge the areas a ected by the bottlenecks, which could cause imbalanced tra c distribution.

Previous studies tried to improve MFD accuracy by including a congestion distribution term in the MFD, which led to generalized MFD (Knoop and Hoogendoorn; 2013). The generalized MFD, even though it is an excellent tool for network state estimation, is still challenged when bottlenecks are present. Therefore, to make MFD more exible in the presence of bottlenecks, it is necessary to quantify the impact of bottlenecks on the MFD and then use bottleneck information to modify the MFD to account for the capacity lost caused by the bottlenecks.

Quantifying the impact of bottlenecks on the network is also necessary for reducing their negative impacts. For example, when planning a construction operation, they can plan the construction according to its impact on the network or when network managers need to prioritize their response to several bottlenecks and allocate more resources to bottlenecks with a higher impact on the network. Lastly, controllers can be developed to take the bottleneck information and adjust the in ow to the a ected area according to the changes bottlenecks bring to network performance.

Nonetheless, to the best of the authors' knowledge, there have not been any studies in the literature studying the clusters of MFDs in the presence of bottlenecks or any adjustment terms for the MFD in the presence of bottlenecks. Also, predicting the bottleneck impacts on the MFD from bottleneck attributes was another unaddressed topic in the literature.

In this study, we are trying to answer the following questions:

- Is it possible to nd meaningful categories for MFDs a ected by bottlenecks?
- How are bottleneck attributes related to the resulting MFD category?
- How can the MFD be adjusted based on bottleneck attributes?
- Is there any interaction between bottleneck attributes regarding their impact on the MFD?
- What policies can be recommended based on a model that relates bottleneck attributes to MFD?

To provide answers to our research questions, we used unsupervised and supervised machine learning algorithms. First, we simulated a large 9 9 grid network (3.2 km^2 area), which is similar to the size of many CBD areas around the world, including Downtown Vancouver (3.7 km^2) and Downtown Sydney (2.8 km^2) CBD areas, in Aimsun. Initially, we simulated the network without bottlenecks to calibrate the MFD.

Then, we created bottlenecks with various attributes, such as di erent start times and durations, to nd the impact of bottlenecks on the MFD. Using the area inside the MFD and the residual density (the number of vehicles in the network at the end of simulation time), we developed an unsupervised machine learning model (k-means) to nd categories of the changes in MFD. We found two clusters here and named them high-impact and low-impact. After labelling MFD categories, we studied the error caused by the di erence between the ow MFD calibrated without bottlenecks predicted after plugging the values of each cluster and observed ow for them. Then, we developed an error correction term for each cluster to reduce the errors caused by each cluster. We developed a supervised machine learning model (decision tree) to predict MFD clusters from bottleneck attributes. Eventually, this decision tree model can provide the proper error correction term for the MFD based on bottleneck information as the input. Finally, we made policy recommendations and proposed a bottleneck-aware tra c controller based on our models. The result of this paper can be used for construction planning and resource allocation for bottleneck scene responders and developing tra c controllers to respond to bottleneck changes.

The rest of the paper is organized as follows. Section 4.3 comprehensively reviews related work. Section 4.4 introduces MFD and the hysteresis loop. Section 4.5 describes the simulated network and tra c scenarios. Section 4.6 summarizes the simulation results and Section 4.7 analyses the impact of bottlenecks on MFDs. The results are discussed in Section 4.8. Finally, conclusions and future recommendations are provided in Section 4.9.

4.3 Literature review

MFD, rst introduced in the 1960s by Godfrey and reformulated by Daganzo in recent years, relates network ow to network density (Godfrey; 1969; Daganzo and Geroliminis; 2008). Geroliminis validated its existence using empirical data from Yokohama and San Francisco, leading to increased interest in its applications for tra c control, routing, taxi dispatch, emission estimation, and safety evaluation (Geroliminis and Daganzo; 2008; Keyvan-Ekbatani et al.; 2012; Yildirimoglu et al.; 2015; Ramezani and Nourinejad; 2017; Halakoo et al.; 2023; Alsalhi et al.; 2018). However, studies have shown that the assumption of homogeneous congestion distribution might not hold in practical situations.

During a truck drivers ' strike and intentional speed reduction, Buisson observed a

hysteresis loop in Toulouse, France's MFD (Buisson and Ladier; 2009). At the same time, simulation results by Mazloumian indicated that network ow might be in uenced by factors beyond density, explicitly accounting for congestion homogeneity using the standard deviation of density (Mazloumian et al.; 2010). Saberi introduced the H index, which represents the ratio of the di erence between congested and uncongested sections' lengths to the total length of the network, providing another measure for congestion heterogeneity (Saberi et al.; 2015). Subsequently, the generalized macroscopic fundamental diagram incorporated the standard deviation of density as a proxy for congestion homogeneity, o ering a more comprehensive framework (Knoop and Hoogendoorn; 2013).

Gayah et al. (Gayah and Daganzo; 2011) demonstrated that networks experience uneven congestion during the unloading phase, wherein the network returns to its capacity state without any additional load. In contrast, during the loading phase, when the network was initially empty, and vehicles were gradually added, such heterogeneity was less prevalent. This phenomenon was illustrated as a hysteresis loop within the MFD, showcasing two ow branches for the same density range. Geroliminis and Sun (Geroliminis and Sun; 2011) proposed that this hysteresis occurrence was in uenced by synchronized transient periods and a drop in capacity, contributing to the variability observed in freeway MFDs. These hysteresis loops were further classi ed as clockwise, counter-clockwise, and gure-8 patterns, with the clockwise hysteresis indicating lower ow during the unloading phase than the loading phase.

On the contrary, counter-clockwise hysteresis was characterized by a lower ow during the loading process than the unloading capacity. This phenomenon typically manifested upstream of a temporary bottleneck before and after removal. The gure-8 hysteresis mechanism, as investigated by (He et al.; 2015), integrated aspects of both clockwise and counter-clockwise hysteresis. This pattern exhibited a lower ow before the onset of congestion and a higher ow after congestion had dissipated.

(Shim et al.; 2019) noted reduced ow values during the afternoon compared to the morning in Daegu, South Korea. They attributed this decline to drivers diverting from their usual routes, which failed to improve their travel time or that of other commuters. Additionally, the morning peak showcased more consistent route preferences among drivers, leading to higher observed ow rates within the MFD.

(Saberi et al.; 2015) demonstrated that the hysteresis area diminishes with an escalation in the proportion of adaptive drivers, indicating a faster recovery from congestion within the network. (Halakoo and Yang; 2021) highlighted that network con guration,

including one-way, two-way, and two-way with prohibited left turn links, can also in u-ence the reduction of hysteresis through adaptive routing.

Bottlenecks, causing capacity reductions and considered non-recurrent within the network, are anticipated to induce a hysteresis loop in the MFD. (Tsubota et al.; 2014) observed the presence of a signi cantly scattered MFD in Brisbane, Australia, during a major bottleneck event. By aligning the scatter timestamps with bottleneck descriptions, such as detection and clearance times, they proposed that the regional MFD could e ectively capture the impact of localized bottlenecks. In a study by (Ji et al.; 2015), diverse hysteresis patterns were observed in the upstream and downstream MFDs of the studied highway, with a counter-clockwise hysteresis noted in the upstream and a clockwise pattern observed downstream. The duration of the bottleneck and the demand level were positively correlated with the size of the hysteresis loop in the MFD.

Furthermore, (Schorr et al.; 2016) conducted a comparative analysis of adverse weather conditions and accidents on the MFD of Seoul, South Korea, revealing that accidents had a more pronounced e ect on MFD capacity reduction than inclement weather.

In a related study, (Amini et al.; 2023) introduced the Critical Score (CS) of links as an explanatory measure for the size of the hysteresis loop, demonstrating that links with higher CS index values resulted in larger hysteresis areas during bottleneck occurrences. However, the study's approach modelled the e ects of bottlenecks as a permanent blockage throughout the entire simulation period (210 minutes), potentially leading to an over-conservative estimation and creating a new user equilibrium. Additionally, the xed consideration of the CS parameter might not re ect its dynamic nature, especially when employing dynamic tra c assignment strategies.

The intricate nature of bottlenecks, often characterized by multiple attributes, suggests the possibility of interplays between these attributes, an aspect not extensively explored in prior literature. While previous research has primarily focused on understanding the impacts of isolated bottleneck attributes, this paper seeks to comprehensively model the in uence of various bottleneck attributes, including duration, location, and start time, on the MFD of urban transportation networks. The primary objective is to illuminate the integration of bottleneck information into tra c controllers, enabling more e ective tra c management strategies.

Figure 4.1: Methods Overview

4.4 Methods

Figure 4.1 shows the overview of the methods used in this paper. We started with a microsimulation experiment in Aimsun. The network was initially simulated without bottlenecks to nd the base case MFD, which we used later to nd the errors caused by bottlenecks. Then, bottlenecks with pre-speci ed attributes such as duration, start time, location and number of lanes were created. The simulation outputs (ow and density) were processed to calculate the hysteresis area and residual density (the density at the end of simulation time). Then, these processed variables were fed into the clustering algorithm to nd the corresponding MFD types in the presence of bottlenecks. We compared the ow values predicted by plugging each cluster's density and standard deviation of density into the base case MFD with the observed ow to nd the MFD errors associated with each cluster. We developed error terms for each cluster accordingly. Eventually, using the MFD types as the response variable and bottleneck attributes as the independent variables, we developed the decision tree showing what bottleneck attributes lead to what MFD types. Using such a decision tree, we proposed a bottleneck-aware tra c controller which adjusts the MFD based on the bottleneck information.

4.5 Numerical Experiment

4.5.1 9x9 grid network

A 9x9 grid network with a 200 m block size (total area of 3.24 km2) was simulated in Aimsun. The speed limit on the road sections is 50km/h, and the capacity of each link is 1800 veh/h (or 900 veh/h/lane). Figure 4.2b shows the demand pro le. As shown in Figure 4.2b, the demand was incrementally increased every 10 minutes during the rst hour. During the second hour, the demand increments were kept constant. Starting at the third hour, no new load was added to the network. We used xed signals with a 60-second cycle length and 2 seconds of amber time for each phase (which starts at the end of green time). At four-legged intersections, the through and right-turn movement received 22 s green time per cycle, while the protected left-turn movement received 4 s green time. For three-legged intersections located on the edges of the network, three phases were used: (1) 24 s green time for through movement and right turn for the links on the edges of the network (2) 4 s green time for the protected left turn (3) 26 s green time for a protected right and left turn from the perpendicular direction.

Dynamic Tra c Assignment (DTA) was used for routing the vehicles. Route costs were updated every ve minutes, similar to the network's average travel time in the base scenario. A logit model with scale parameter one was used to distribute the vehicles based on the most recent travel cost of the available paths for each OD pair. We opted not to use en-route as it would decrease the area inside the hysteresis loop (Halakoo and Yang; 2021; Saberi et al.; 2015) and create a new user equilibrium, which adds to the complexity of the analysis.

The Aimsun API module was used to create bottlenecks and record network state. Bottleneck locations were randomly chosen from the links inside the shaded area in Figure 4.2a and outside the shaded area. The shaded area embraces the busiest links of the network. This congestion concentration in the middle of the network is due to the large number of shortest paths passing through the central links of a grid network when the demand is uniform (Ortigosa and Menendez; 2014). Figure 4.3 shows that the area marked with a box embraces the busiest links at three timestamps (60, 120, and 180 minutes after the start of the simulation). We considered two bottleneck start times: the loading phase (60 minutes after the simulation) and the unloading phase (180 minutes after the simulation starts). We picked these two start times based on the base case scenario simulation results in which we found equal network density for two timestamps (as shown in Figure 4.4 B for the tested demand, orange dots). Also, since

(a) Network geometry

(b) Demand pro le

Figure 4.2: 9x9 Grid network

the demand level is shown in Figure 4.2b, pushes the network to the capacity state, we tested the network behaviour with %100 more demand to observe network behaviour in the congested branch of the MFD (we call this demand level %200 base since it does not have a bottleneck, but it is %200 more than the pro le shown in Figure 4.2b). Figure 4.4 A shows the MFD without bottleneck. The demand level in Figure 4.2b (shown in blue Figure 4.4 A) does not produce a signi cant hysteresis or scatters in the MFD. We developed a 3rd-degree polynomial generalized MFD (Simoni et al.; 2014; Knoop and Hoogendoorn; 2013) based on %200 level MFD shown in Equation 4.1 K and K_{std} stand for the density and standard deviation of the density, respectively. a,b,c,d are the model parameters and calibrated as: a = 90.7033; b = -0.7272; c = 0.0015; d = -29.6966 (r^2 = 0.985).

$$G(K; K_{std}) = aK + bK^2 + cK^3 + dK_{std} \tag{4.1}$$

Figure 4.4 B shows the base case scenario density evolution. Figure 4.4 C shows the standard deviation of the density evolution in the base scenario. As Figure 4.4 shows, even though the density and standard deviation of it for 60 and 180 minutes have similar values, more scatters are observed for the time interval between 60 and 180, suggesting an unstable tra c state. Matching the density values to the MFD, we can observe these highly scattered points correspond to the capacity state. Three bottleneck durations, 15

Figure 4.3: Network density in the base scenario at A) 60 minutes, B) 120 minutes, C) 180 minutes after the start of simulation

Figure 4.4: Base case (no bottlenecks) A) MFD, B) Density C) Standard deviation of density

minutes, 35 minutes, and 60 minutes were considered. Also, to consider di erent extents of bottlenecks, we created one-lane blockage and two-lane blockage bottlenecks. In total, 2 (locations) Ö 2 (start time) Ö 3 (duration) Ö 2 (number of lanes) = 24 scenarios were created. Each scenario was replicated 15 times with di erent random seeds. Each replication result (ow and density) was recorded to create MFD and calculate hysteresis area and residual density. Therefore, 360 MFDs were simulated (15 replications Ö 24 scenarios).

4.6 Simulation results

Figure 4.5 shows the results of bottleneck scenarios. Each row represents all scenarios with the same bottleneck duration. As Figure 4.5 A, D, and G show, depending on the bottleneck attributes, the range of outcomes on the MFD could be from complete gridlock to an increase in hysteresis size or even subtle changes in the MFD compared to the base scenario. However, it should be noted that all these MFDs still have similar free- ow speeds and the same critical density while some capacity reduction was observed. Also, comparing the bottleneck MFDs with the base %200 MFD, we realized the %200 MFD is crisper and is the upper limit of the cases with the bottlenecks. Therefore, one can modify the base %200 MFD by including an error term to account for the changes caused by the bottlenecks. In previous work such as (Geroliminis et al.; 2012), the calibration noise in MFD was considered as a normally distributed term added to the MFD. We will explore the distribution of the MFD errors caused by bottlenecks in section 4.7.1.2.

The e ect of bottleneck attributes on network congestion is shown in Figure 4.5 B, E, and H, which ranges from an acute increase in density and the gridlock to a more subtle increase in density level. While Figure 4.5 B shows only two trends (density values either close to zero or between 150 to 175 for time intervals after 200 minutes), multiple trends could be observed for density values after 200 minutes in Figure 4.5 E and H. The e ect of bottleneck attributes on the standard deviation of density is shown in Figure 4.5 C, F, and I, which shows a di erent range of results from similar results to the base scenario (maximum increase close to 40 veh/km) to twice higher values (80 veh/km). Also, it can be deduced since the density and standard deviation of the density for some cases with 60-minute bottlenecks were close to zero after 200 minutes; these cases had no residual density, while other scenarios with the same bottleneck caused a gridlock. Therefore, it can be concluded that multiple outcomes for the same bottleneck duration could be observed, which challenges the previous ndings in the literature where only an increase in the duration of the bottleneck was associated with an increase in hysteresis size. Also, our results highlight the interplay between bottleneck attributes. In the next section, we try to nd meaningful clusters in the MFDs and simplify categorizing the range of the MFDs we observed in Figure 4.5.

4.7 Results analysis

4.7.1 Clustering bottleneck MFDs

We used the k-means clustering algorithm to nd the patterns in the MFDs. First, we plotted the distribution of the MFD variables, i.e., hysteresis area and residual density, to nd the possible dense clusters in the data. Figure 4.6 shows the hysteresis area and residual density plots. Figure 4.6 A shows the distribution of the hysteresis area, which clearly shows two peaks. Similarly, Figure 4.6 B shows that residual density distribution represents two peaks. Figure 4.6 C scatter plot shows hysteresis area values plotted against the corresponding residual density, depicting two dense clusters with two connecting points. Considering the two peaks in distributions and two dense spots in the scatter plots, we picked two clusters as our initial guess.

4.7.1.1 Identifying the number of clusters

We tried a range of clusters (2-7). We calculated clustering validity indexes such as inertia (sum of squared sample distance to their centroid), average silhouette, Calinski-Harabasz index, and Davies Bouldin index. We used these indexes to nd the optimal number of clusters for MFDs.