Technische Universität München
TUM School of Engineering and Design

Modeling of urban traffic flow: Approximation of the network macroscopic fundamental diagram using semi-analytical methods

Gabriel Tilg

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines *Doktors der Ingenieurwissenschaften (Dr.-Ing.)* genehmigten Dissertation.

Vorsitz: Prof. Dr.-Ing. Klaus Bogenberger

Prüfende der Dissertation:

1. Prof. Dr.-Ing. Fritz Busch
2. Prof. Monica Menendez, Ph.D.
3. Prof. Dr. Constantinos Antoniou

Die Dissertation wurde am 16.11.2021 bei der Technischen Universität München eingereicht und durch die *TUM School of Engineering and Design* am 11.04.2022 angenommen.

Dies ist eine kumulative Dissertation basierend auf Veröffentlichungen in internationalen Fachzeitschriften.

Gabriel Tilg: *Modeling of urban traffic flow: Approximation of the network macroscopic fundamental diagram using semi-analytical methods*

Die Zeit ist ein dampfendes Ross.

Acknowledgements

First and foremost I would like to express my gratitude to my supervisor Prof. Fritz Busch, who made me feel welcome from the very first moment at the Chair of Traffic Engineering and Control. Thank you for introducing me to the German traffic research community, offering me the freedom of choice regarding the dissertation topic, highlighting the importance of the applicability in research, and providing full support throughout the doctoral studies. I am equally grateful for my second supervisor, Prof. Monica Menendez, who tirelessly gave me invaluable feedback and supported me in all matters. Thank you for sharing your knowledge about traffic flow theory and inviting me to the research stay at NYU Abu Dhabi. Additionally, I would like to thank Prof. Constantinos Antoniou for being the third examiner, Prof. Ludovic Leclercq for his scientific contributions to parts of this work, and Prof. Klaus Bogenberger for being the examinations committee's chairman.

I want to thank the German Federal Ministry of Transport and Digital Infrastructure for funding the research projects which enabled me to write this dissertation.

Of course, all current and former colleagues at the Chair of Traffic Engineering and Control deserve a great thank you. Even as an Austrian with a quite noticeable accent, or especially because of that, they warmly welcomed me in the Bavarian capital. Thank you Sabine Krause, Frederik Bachmann, Fabian Fehn, Martin Margreiter, Florian Noack, Eftychis Papapanagiotou, Matthias Spangler, Philipp Stüger, and Antonis Tsakarestos. Special thanks go to Lukas Ambühl, Sasan Amini, and Sergio Batista. The discussions with you on the MFD and related topics greatly contributed to this dissertation. I am looking forward to our future projects!

Without a doubt, completing a dissertation affects the personal life and the corresponding support is invaluable. Therefore, I would like to thank my siblings Laura and Alex, and my friends outside the research community. In particular, I would like to thank Judith for her great patience and tolerance regarding the unusual working hours and for the strong mental support whenever needed. Finally, and most importantly, I deeply thank my parents, Anita and Herbert, for their unconditional and ongoing support since the very first second. I dedicate this dissertation to them.

Gabriel Tilg

Executive summary

The worldwide increasing urbanization and population growth lead to densely populated cities around the globe. All people who live, work, study, or are otherwise active in and around urban areas have the need to be mobile. The resulting demand for travel pushes the supply of urban transport infrastructure to its limits. In other words, the demand exceeds the infrastructure's capacity resulting in severe congestion and consequently in more emissions, noise, and psychological stress for travelers. Adequate traffic management strategies are in need to effectively mitigate such congestion in urban areas. The proper design of appropriate strategies requires parsimonious and efficient tools and models for the correct identification and accurate quantification of urban congestion.

Traffic flow theory provides a key concept for the analysis of congestion: the fundamental diagram of traffic flow. It describes the relationship of traffic flow, traffic density, and the momentary mean speed for a specific location. Originally, it was formulated based on empirical data from freeways. It allows for the efficient modeling of traffic dynamics and the identification of the capacity of a homogeneous road stretch. It is the base for describing the severity of congestion and an important input for the design of effective traffic management strategies.

Recently, a similar functional relationship was found for urban areas: the network or macroscopic fundamental diagram (MFD) of urban traffic. It describes the relationship of traffic flow, traffic density, and the momentary mean speed not just for a specific location but for a given network. In the last two decades, many studies have explored the properties of the MFD, found empirical evidence for it, and integrated the concept into the design of traffic control and management measures. The corresponding literature clearly shows the capabilities of the MFD for modeling, monitoring, and control of urban traffic. However, prior to its application, its functional form needs to be defined. In case no empirical or traffic simulation-based data exist, semi-analytical approaches can be utilized for its estimation. Yet, these approaches apply only to the corridor level, i.e. for a sequence of links without substantial turning flows at intersections. Thus, the application of such methods to realistic urban networks is rather limited.

This dissertation contributes to closing this research gap. First, we thoroughly analyze the state-of-the-art semi-analytical methods to estimate the MFD at the corridor level. More specifically, we evaluate the *method of cuts* and the *stochastic approximation* of the MFD. We choose a highly complex urban arterial in Munich, Germany, and compare the semi-analytically approximated MFDs with the corresponding empirical one. Furthermore, we conduct a sensitivity study to identify the most important input parameters. This analysis reveals that the MFDs resulting from the approximation methods studied contain substantial

errors. Nevertheless, the estimate by the method of cuts matches the empirical MFD considerably well.

Based on these results, we choose the method of cuts as the base for developing a methodology to semi-analytically approximate the MFD for realistic urban networks. First, we focus our attention on the variational theory (VT) of traffic as it is the theoretical base of the method of cuts. The VT in its original form also applies merely to the corridor level. Therefore, we extend it to urban networks by accounting for the effects of turning flows at intersections and showcase the applicability of the extended VT in a case study for the Sioux Falls network. We show that our extended version successfully inherits important properties, such as the numerical precision and the ability to model complex intra-link bottlenecks, e.g. random pedestrian crossings, from the original VT.

Finally, we focus on the semi-analytical approximation of the MFD for networks. For this purpose, we develop a flexible framework that utilizes analytical approaches as well as the proposed extended VT. The inclusion of several approaches allows the framework user to trade-off modeling complexity, computational efficiency, and estimation accuracy. In a case study for the Sioux Falls network, we illustrate the improvements provided by our MFD approximation, which is up to five times more accurate than the state-of-the-art methods. Moreover, we highlight the effects of spatial demand patterns on the MFD.

In summary, this dissertation studies existing methods to semi-analytically approximate the corridor MFD and proposes extensions based on such to estimate the network MFD. Semi-analytically approximated MFDs can be beneficial for use cases where no empirical or simulated data exist, e.g. hypothetical scenario analyses, or in case the computational cost of traditional traffic simulations are too high, e.g. in model-based optimization frameworks or real-time applications. The resulting MFDs can be utilized for the efficient design of traffic management measures to reduce congestion in urban areas. Furthermore, the developed extension of VT applies beyond the MFD approximation and thus contributes to computationally efficient traffic modeling for urban networks in general. The model-based design of traffic management measures for urban areas is essential to improve traffic operations and lower the external costs of traffic such as congestion, emissions, and noise. By extending the scope of existing efficient traffic models to the network level, this dissertation can hopefully contribute to the rapid creation of sustainable urban transportation systems.

Kurzfassung

Die weltweit zunehmende Urbanisierung und das Bevölkerungswachstum führen zu dicht besiedelten Städten rund um den Globus. Alle Menschen, die dort leben, arbeiten, studieren, oder anderweitig aktiv sind, haben das Bedürfnis mobil zu sein. Die daraus resultierende Verkehrsnachfrage bringt das Angebot der städtischen Verkehrsinfrastruktur an seine Grenzen. Die Überschreitung der Kapazität der Infrastruktur führt jedoch zu schweren Staus und damit unter anderem zu mehr Emissionen, Lärm und psychischem Stress für die Verkehrsteilnehmer. Um solche Staus in städtischen Gebieten wirksam zu reduzieren, sind geeignete Verkehrsmanagementstrategien erforderlich. Die effektive Entwicklung geeigneter Strategien erfordert einfache und wirksame Instrumente und Modelle zur korrekten Identifizierung und genauen Quantifizierung von Verkehrsstaus in Städten.

Für Autobahnen bietet die Verkehrsflusstheorie ein entsprechendes und lang etabliertes Konzept, das auf der Grundlage empirischer Daten entdeckt wurde, das Fundamentaldiagramm des Verkehrsflusses. Es beschreibt den Zusammenhang von Verkehrsfluss, Verkehrsdichte, und momentaner mittlerer Geschwindigkeit. Damit ermöglicht es die effiziente Modellierung der Verkehrsdynamik und die Ermittlung der Kapazität eines Straßenabschnitts. Dies ist die Grundlage für die Beschreibung der Verkehrsüberlastung und ein wichtiger Faktor für die Entwicklung geeigneter Verkehrsmanagementstrategien.

Kürzlich wurde ein ähnlicher funktionaler Zusammenhang für städtische Gebiete entdeckt: das netzweite oder makroskopische Fundamentaldiagramm (MFD) des Stadtverkehrs. Es beschreibt den Zusammenhang zwischen Verkehrsfluss, Verkehrsdichte und der momentanen Durchschnittsgeschwindigkeit nicht nur für einen konkreten Ort, sondern für ein ganzes Netz. In den letzten zwei Jahrzehnten haben viele Studien die Eigenschaften des MFDs untersucht, empirische Belege dafür gefunden und das Konzept in die Gestaltung von Verkehrssteuerungs- und Verkehrsmanagementmaßnahmen integriert. Die entsprechende Literatur zeigt die Fähigkeit des MFDs, den städtischen Verkehr zu modellieren, zu überwachen und zu steuern, deutlich auf. Vor der Anwendung des Konzepts muss jedoch seine funktionale Form definiert werden. Liegen keine empirischen oder verkehrssimulationsbasierten Daten vor, können semi-analytische Ansätze zur Schätzung verwendet werden. Diese Methoden gelten jedoch nur auf Korridorebene, d.h. für eine Abfolge von Straßenabschnitten, auf denen keine wesentlichen Abbiegeströme auftreten. Daher ist die Anwendung solcher Methoden für realistische städtische Netze stark eingeschränkt.

Diese Dissertation trägt dazu bei, diese Forschungslücke zu schließen. Zunächst werden die semi-analytischen Methoden zur Schätzung des MFDs auf Korridorebene entsprechend dem Stand der Technik gründlich analysiert. Im Speziellen sind dies die *Method of Cuts* und die *Stochastic Approximation* des MFDs. Wir wählen einen hochkomplexen, städtischen Ver-

kehrskorridor in München, Deutschland, und vergleichen die semi-analytisch approximierten MFDs mit dem entsprechenden Empirischem. Darüber hinaus führen wir eine Sensitivitätsstudie durch, um die wichtigsten Eingabeparameter zu identifizieren. Diese Analyse zeigt, dass die aus den untersuchten Approximationsmethoden resultierenden MFDs substantielle Fehler enthalten. Dennoch stimmt die Schätzung durch die Method of Cuts mit dem empirischen MFD sehr gut überein.

Auf der Grundlage dieser Ergebnisse wählen wir die Method of Cuts als Basis für die Entwicklung einer Methodik zur semi-analytischen Approximation des MFDs für realistische städtische Netze. Zunächst richten wir unsere Aufmerksamkeit auf die Variational Theory (VT) des Verkehrs, welche die theoretische Grundlage der Method of Cuts ist. Die VT gilt in ihrer ursprünglichen Form ebenfalls nur für die Korridorebene. Daher erweitern wir sie auf städtische Netze, indem wir die Auswirkungen abbiegender Verkehrsströme an Kreuzungen berücksichtigen und demonstrieren die Anwendbarkeit der erweiterten VT in einer Fallstudie für das Sioux Falls-Netz. Wir zeigen, dass unsere erweiterte Methodik wichtige Eigenschaften der ursprünglichen VT erfolgreich übernimmt, wie die numerische Präzision und die Fähigkeit zur Modellierung komplexer Engpässe auf der Straßenabschnitssebene, z.B. zufällige Fußgängerüberquerungen.

Schließlich konzentrieren wir uns auf die semi-analytische Approximation des MFDs für Netze. Zu diesem Zweck entwickeln wir eine flexible Methodik, die sowohl analytische Ansätze als auch die vorgeschlagene erweiterte VT nutzt. Durch die Einbeziehung mehrerer Ansätze kann der Modellnutzer einen Kompromiss zwischen Modellierungskomplexität, Berechnungseffizienz und Schätzgenauigkeit eingehen. In einer Fallstudie für das Sioux Falls-Netz veranschaulichen wir die Verbesserungen durch unsere MFD-Approximation, die bis zu fünfmal genauer ist als die Methoden nach dem Stand der Technik. Darüber hinaus zeigen wir die Auswirkungen der räumlichen Nachfragemuster auf das MFD auf.

Zusammenfassend analysiert diese Dissertation bestehende Methoden zur semi-analytischen Approximation des Korridor-MFDs und empfiehlt darauf aufbauende Erweiterungen zur Annäherung des netzweiten MFDs. Solche Approximationen können in Anwendungsfällen von Vorteil sein, in denen keine empirischen oder simulierten Daten vorliegen, z.B. bei der Analyse hypothetischer Szenarien, oder wenn der Rechenaufwand traditioneller Verkehrssimulationen zu hoch ist, z.B. bei modellbasierten Optimierungsverfahren oder Echtzeitanwendungen. Die resultierenden MFDs können für die effiziente Gestaltung von Verkehrsmanagementmaßnahmen zur Reduzierung von Staus in städtischen Gebieten eingesetzt werden. Darüber hinaus ist die entwickelte Erweiterung der VT über die MFD-Approximation hinaus anwendbar und trägt somit zu einer rechnerisch effizienten Verkehrsmodellierung für urbane Netze im Allgemeinen bei. Der modellgestützte Entwurf von Verkehrsmanagementmaßnahmen für städtische Gebiete ist von entscheidender Bedeutung, um die Verkehrsabläufe zu verbessern und die externen Kosten des Verkehrs wie Staus, Emissionen und Lärm zu senken. Indem diese Dissertation den Anwendungsbereich bestehender, effizienter Verkehrsmodelle auf die Netzebene erweitert, kann sie hoffentlich zur raschen Schaffung nachhaltiger, städtischer Verkehrssysteme beitragen.

Contents

Preface

Paper-based dissertation

This dissertation is written based on the following manuscripts that were submitted or published in peer-reviewed scientific journals:

- G. TILG, S. AMINI & F. BUSCH [2020a]: Evaluation of analytical approximation methods for the macroscopic fundamental diagram. In: *Transportation Research Part C: Emerging Technologies* 114, pp. 1–19. ISSN: 0968-090x. DOI: doi.org/10.1016/j.trc.2020.02.003
- G. TILG, L. AMBÜHL, S. F. BATISTA, M. MENENDEZ & F. BUSCH [2021b]: On the application of variational theory to urban networks. In: *Transportation Research Part B: Methodological* 150, pp. 435–456. DOI: 10.1016/j.trb.2021.06.019
- G. TILG, L. AMBÜHL, S. F. BATISTA, M. MENENDEZ, L. LECLERCQ & F. BUSCH [n.d.]: From corridor to network macroscopic fundamental diagrams: A semi-analytical estimation approach. Submitted for publication.

G. Tilg conducted the majority of the works independently. The detailed author's contributions along with short summaries are provided in Appendix C.1. The respective published and submitted versions of the manuscripts are included in Appendix C.3. Other scientific works which were authored and co-authored during the doctorate and published as journal or conference papers, held as presentations, or presented as posters can be found in Appendix B.

The chapters 3 to 5 were written based on the aforementioned papers. However, terms and symbols were altered for the sake of consistency throughout the dissertation. Moreover, the introductory and concluding sections of these chapters were modified to reduce redundancies. The remaining chapters of this dissertation are original and are not included in any other manuscript.

Chapter 1

Introduction

1.1 Background and problem setting

1.1.1 Congestion

Mobility is nothing less than a basic human need that every one of us desires to satisfy. However, while being mobile we consume space, and such is especially scarce in urban areas. Megatrends such as increasing urbanization, a rising world population, and growing prosperity lead to an expansion and densification of modern cities around the globe. At the same time, unsustainable travel modes such as the private car seem to be on a never-ending rise [e.g. EUROSTAT, 2021], not to speak of the recent emergence of so-called SUVs that consume even more space. Furthermore, the spread of delivery services and corresponding logistics trips is constantly increasing. Such developments lead to what most of us, living in cities or near them, experience every day. Congestion.

While congestion is only one out of many externalities related to traffic, it is the cause of significant costs to society. Nevertheless, other aspects of external costs such as greenhouse gas emission, accidents, and noise, are equally important to address. Still, congestion may be put in focus as it intensifies the impact of the other externalities. For example, vehicles have a higher number of stops during congestion, they accelerate more frequently, and drivers experience higher inconvenience. These aspects might lead to even more greenhouse gas emissions, accidents, and noise. As an illustration, studies based on data from navigation systems estimated that a driver in Munich, which was 'awarded' the most congested city of Germany many times in recent years, loses 65 h in congestion per year on average [INRIX, 2020]. Furthermore, the cost of congestion were calculated as 585 dollars per driver and year. These numbers point out the effects of congestion from the user's perspective. Extrapolating them to entire cities further demonstrates the large impact of congestion from an economic point of view. Thus, the minimization of such is of high importance. Next to direct positive effects of congestion mitigation, research has shown that other aspects such as the reliability of transport systems as well as public health benefit from congestion reduction, too [SYSTEMATICS, 2005; LEVY, BUONOCORE & VON STACKELBERG, 2010]. Lastly, it is important to note that ensuring the high performance of transportation systems is of advantage for the entire society since these systems contribute substantially to economic development and prosperity [BUITELAAR, VAN DER HEIJDEN & ARGIOLU, 2007].

Strategies to mitigate congestion depend on correct identification and accurate quantification of such. At first sight, it is, however, a traffic phenomenon for which multiple definitions and consequently numerous measures exist. A recent review paper by AFRIN & YODO [2020] explored and evaluated these. For example, a single road user might interpret congestion as the occurrence of delays, i.e. excess travel times [SYSTEMATICS, 2005]. Other researchers defined congestion as an incremental increase of a road user's generalized costs due to the disruption of the usual traffic flow [LITMAN, 2011]. By focusing on the system's perspective, the highway capacity manual [TRB, 2016] defines six levels of service related to congestion severity. The level of service corresponds to the ratio of the average speed in traffic and the free-flow speed, that is the speed under optimal conditions. Very similar to that, the German pendant, the 'Handbuch für die Bemessung von Straßenverkehrsanlagen', also employs levels of service to quantify congestion based on average speed values. The level-of-service approach in well-known guidelines, as well as the results from AFRIN & YODO [2020], indicate the suitability of using indicators related to the average speed and its reduction to identify and quantify congestion.

1.1.2 Demand, supply, and capacity

But why do average speeds drop? This can be explained based on the concept of demand, supply, and capacity. More specifically, the average speeds in traffic drop when the travel demand exceeds the transportation system's capacity which is governed by its supply.

The demand for travel results from the basic need for mobility. People travel to work, to study, to meet friends and family, to buy groceries, for the activity of moving itself, and much more. Thereby, they choose a certain time to start their trip. Additionally, some people drive their car, others walk, and again others ride a bus. Last, they will choose a specific route to reach their destination, given their experience and perhaps some up-to-date information on expected travel times. Consequently, in transportation science, demand is modeled by three different choices: (i) the departure time choice, (ii) the mode choice, and (iii) the route choice. These choices represent the flexibility with which individuals travel. On the other hand, there is the supply of a transport system. It includes the sheer road network, but also traffic control facilities, public transport systems, and other transport-related infrastructure. Therefore, it dictates possible modes and routes for travelers.

A crucial indicator of the system's supply is its capacity. But what is *capacity*? For its explanation, let us first elaborate on the term *traffic state*. The state of traffic is the result of the interplay of demand for travel and the supply provided by the transportation infrastructure. In analogy to structural engineering, capacity is the state where additional load leads to a collapse. In terms of transportation systems, this equals a certain flow measured in vehicles per hour that can be served given the existing supply conditions. Once the demand exceeds this capacity, congestion occurs. Therefore, both demand and supply determine when and where congestion occurs, and how it propagates throughout the network. Congestion is reflected by queues at intersections lasting for multiple traffic signal cycles, vehicles driving at speeds lower than the limits, and increased noise and emissions. Note that, in reality, capacity is rather a range of flows than a single fixed value. The highway capacity manual

defines it as “The maximum sustainable flow rate at which vehicles or persons reasonably can be expected to traverse a point or uniform segment of a lane or roadway during a specified time period under given roadway, geometric, traffic, environmental, and control conditions” [TRB, 2016]. Similar to that, the German Road and Transportation Research Association defines it as maximum traffic volume that can be achieved by a traffic flow under the given constructional and traffic conditions[1] [HANKE, BRAAM, BREITENSTEIN, HABERMEHL, HERBER, STEPHAN & WETTERLING, 2012]. While the exact wording is slightly different, the general idea is clear. The capacity is a property of the supply. Therefore, congestion can be minimized by decreasing the demand or extending the supply and consequently increasing the capacity.

For such purposes, appropriate measures need to be designed. Extending the built environment, i.e. the simple addition of more lanes, is rarely an option, as urban space is scarce and such capacity increases often induce more travel demand and hence traffic. Alternatives are traffic management measures that actively modify the operation of the transportation system, e.g. by changing the traffic signal control logic or employing a congestion pricing scheme. Furthermore, they can be pro-active, reactive, and change dynamically with time. On the other hand, they can target longer time horizons with the goal of obtaining a change in travel behavior. Corresponding measures include the optimization of a public transport timetable, e.g. to induce a mode shift, infrastructural changes in the network topology, or other planning-related measures. Independent of whether the measures aim to modify supply and consequently capacity, or if they focus on demand aspects, all types of measures require performance indicators for evaluation, typically including congestion-related ones. Hence, the capacity of a transportation system is of general interest for the mitigation of congestion.

1.1.3 The fundamental diagram of traffic

As suggested above, the capacity is a traffic state that is key for understanding the interaction of demand and supply. A general traffic state is a macroscopic characteristic from a more theoretical perspective. It is defined by the triplet of traffic flow q, traffic density κ, and the average instantaneous speed of vehicles $\overline{v}$. The flow is defined as the number of vehicles that pass a given point per time interval. The density is the number of vehicles per space interval, and the average speed is the mean of the vehicle’s instantaneous speeds. To fully define a traffic state, a subset of two of those variables is sufficient, as they are interrelated with the fundamental equation of traffic flow [e.g. HALL, 1996]:

$$q = \kappa \cdot \overline{v} \tag{1.1}$$

A traffic state is stationary, or in steady-state, if these characteristics do not change for the given road segment and time period, i.e. in a specific time-space region. DAGANZO [1997] stated it to be reasonable to expect some relationship between these variables for stationary

[1] German original: ‘Größte Verkehrsstärke, die ein Verkehrsstrom unter den gegebenen baulichen und verkehrlichen Bedingungen erreichen kann.’

states which is a property of the road, the environment, and the population of travelers as one can expect drivers to behave equally on average given the same average conditions.

As a fact, this relationship was examined and observed in the very early days of traffic flow research and was appropriately named the fundamental diagram (FD) of traffic. The first works concerning the FD were conducted by Greenshields in the 1930s [GREENSHIELDS, BIBBINS, CHANNING & MILLER, 1935]. They proposed a linear relationship between the flow and the average speed. Afterward, many researchers explored this relationship of traffic states and proposed numerous different functional forms. Amongst them are single-regime functions, i.e. where the congested and uncongested traffic states are described by merely one function, such as the exponential one [DRAKE, SCHOFER & MAY, 1966]. But also multi-regime approaches were developed, where different functions describe congested and uncongested conditions. Examples are the triangular FD [NEWELL, 1993], the truncated triangular one [DAGANZO, 1997], the inverse lambda shaped one [KOSHI, 1983], and extensions of such [e.g. WU, 2002]. Furthermore, KERNER & LIEU [2005] proposed to describe the congested regime of the FD by an area rather than a curve.

The FD fostered the understanding of traffic phenomena such as the capacity drop and it has been utilized as a performance indicator to monitor and control traffic. Also, it is essential to traffic flow theory as it is the base for many models to describe traffic dynamics along road stretches, e.g. for the well-known kinematic wave theory (KWT) (see Chapter 2). Despite its origin in the 1930s, related research is still being published showing the ongoing interest in the concept [e.g. KNOOP & DAAMEN, 2017]. However, the FD was originally observed, further developed, and mostly applied for freeways since steady traffic states are more likely observed when no interruptions such as traffic lights, parking vehicles, and pedestrians exist. Thus, its validity as a model and its suitability as a performance indicator are restricted to homogeneous links where supply conditions do not change drastically.

1.1.4 The macroscopic fundamental diagram of urban traffic

Nevertheless, a relationship with the FD's parsimony and efficiency that can be applied for modeling, monitoring, and control of traffic would be beneficial and desirable at the network level as well. Researchers reasonably conjectured that also in a network, the average flow might increase in a predictable manner until capacity is reached. From then on, it is clear that additional demand or queues at the boundaries of the network lead to increased congestion until network-wide traffic reaches a gridlock state – where nothing moves. This leads us to the formal relationship of network-wide average flow, density, and speed, which has been labeled as the *network* or *macroscopic* fundamental diagram in literature. Hereafter, we will refer to the concept as macroscopic fundamental diagram (MFD).

In the following, we introduce the concept in detail including its history of origins. Then, we present research works concerning its properties and empirical evidence of its existence. Subsequently, the wide range of applications is shown to demonstrate its value for traffic engineering and science. Lastly, we will focus on the most important MFD-related aspect for this dissertation - the currently available approximation methods.

Definition and history

In the 1960s, Thomson and Smeed were among the first to investigate network traffic [THOMSON, 1967; SMEED, 1967; SMEED, 1968]. They collected empirical data from cities and described the relationship of network topology, travel demand, modal split, and congestion. Similar to that, ZAHAVI [1972a], ZAHAVI [1972b], WARDROP [1968] and GODFREY [1969] proposed a generic relation between average speed and flow based on network topology parameters. However, these models only apply to light traffic as congestion is not comprehensively included in the proposed functions [GEROLIMINIS & DAGANZO, 2008]. In contrast, the two-fluid theory developed by HERMAN & PRIGOGINE [1979] and empirically verified by HERMAN & ARDEKANI [1984] is able to account for congestion. It relates the average speed to the fraction of stopped vehicles. This theory was the base for MAHMASSANI, WILLIAMS & HERMAN [1987] to describe the overall network performance. They applied microscopic simulation to verify their results, tested the sensitivity of their model against supply parameters, and concluded that the functional relationship of network traffic variables is similar to those for link segments.

Nevertheless, it was DAGANZO [2007] who formulated the explicit mathematical relation of the MFD, i.e. between the trip completion rate of a single neighborhood-sized reservoir and the accumulation of vehicles inside. He assumed slow-varying demand, a constant average trip length, evenly spread congestion, and stationary traffic states. Given these assumptions, he postulated that trip completion rates are proportional to average flows inside the reservoir and that the relationship is a property of the network, and, hence, insensitive to small changes in demand. Therefore, aggregated traffic dynamics can be explained by the accumulation in the reservoir. GEROLIMINIS & DAGANZO [2007] revised this theory by expressing two main postulates:

1. Homogeneously congested reservoirs exhibit an MFD relating *travel production* and *accumulation*. The former is the product of average flow and network length and is measured in vehicle-kilometers traveled. The latter is the product of average density and the network length and is measured in number of vehicles.
2. The trip completion rate is proportional to the travel production and therefore to the average flow.

In an empirical study for Yokohama, Japan, GEROLIMINIS & DAGANZO [2008] investigated this relationship and were able to experimentally validate the existence of the accumulation-based MFD, i.e. a smooth and low-scatter relationship between the network-wide average flow and density. This MFD is illustrated in Figure 1.1. The y-axis shows the network-wide average flow Q, and the x-axis the average density K. The different symbols represent different measurement periods, see GEROLIMINIS & DAGANZO [2008] for more details. The figure reveals a smooth relationship between both variables and demonstrates that the network capacity was reached and congested traffic states occurred. The authors found the MFD to exist for homogeneous regions, i.e. with low spatial variability of traffic density. Also, they conjectured that it is not significantly impacted by demand changes. Furthermore, they showed with empirical data that the average trip length is constant. Thus, the travel production indeed stays proportional to the average flow. Since the accumulation is the state

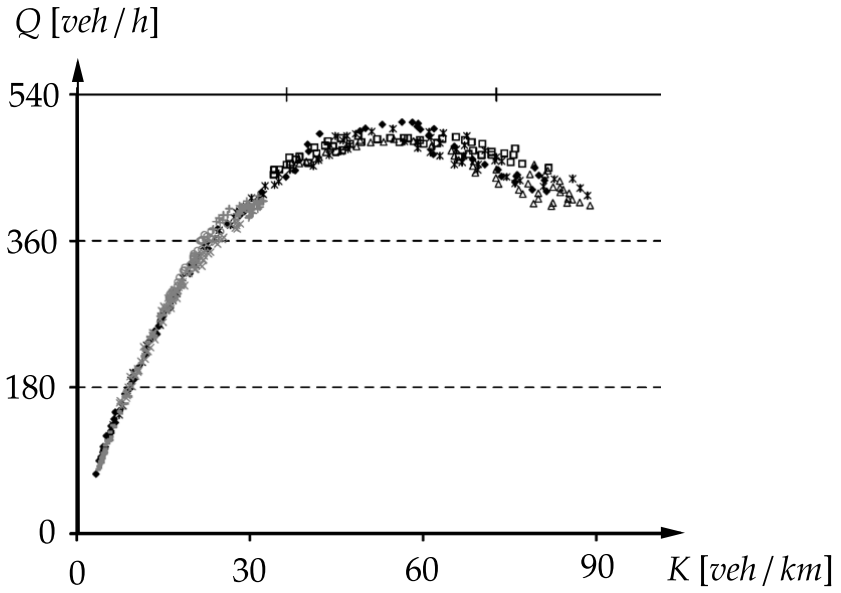


Figure 1.1: MFD for Yokohama (adapted from GEROLIMINIS & DAGANZO [2008]).

variable in this theory, the corresponding MFD is referred to as 'accumulation-based' in the literature.

A different notion of the MFD was introduced by ARNOTT [2013] and VICKREY [2020] that considers the reservoir-specific distribution of trip lengths instead of the average, and was correspondingly labeled as 'trip-based'. Furthermore, a 'delay-based' MFD was introduced by HADDAD & ZHENG [2020] where delays are considered in the outflow calculation. Note that both alternative notions need a speed-density function such as the accumulation-based MFD as input. Thus, in this dissertation, we focus on the accumulation-based MFD introduced by DAGANZO [2007]. For a more detailed discussion of the development history of the MFD and comprehensive classification of related literature, we refer to the recent review by JOHARI, KEYVAN-EKBATANI, LECLERCQ, NGODUY & MAHMASSANI [2021].

Properties and empirical evidence

The idea of a simple relation between network-wide average flows and densities is interesting to many applications. Multiple empirical studies were conducted for cities around the globe to further verify the existence of the MFD and explore its properties [AMBÜHL, LODER, MENENDEZ & AXHAUSEN, 2017; AMBÜHL, LODER, BLIEMER, MENENDEZ & AXHAUSEN, 2018; LODER, AMBÜHL, MENENDEZ & AXHAUSEN, 2019b]. Unlike the results from Yokohama [GEROLIMINIS & DAGANZO, 2008], a substantial number of empirical MFDs revealed considerable scatter. It is important to emphasize that, on the one hand, there are *observed* macroscopic traffic states, and on the other hand there is the *theoretical* upper bound of traffic performance as defined by DAGANZO & GEROLIMINIS [2008]. Throughout this dissertation, we use the term *realized* to refer to the observed MFD and *idealized* to

refer to the upper bound.

DAGANZO & GEROLIMINIS [2008] defined the idealized MFD based on a number of regularity conditions. These contain the demand being slow-varying and well-distributed, as well as the network being redundant, thereby offering many routes, and consisting of similar links. Also, traffic on these links is required to follow an FD that is not affected by turning flows. The idealized MFD was defined as a property of the network, similar to the link FD that is a property of the road. It describes the maximum flows for a given density during stationary traffic conditions. The combination of the demand- and supply-related regularity conditions lead to maximum performance in the network. The idealized MFD was considered to be independent of demand, as such is not an independent characteristic for its approximation. For example, if the supply is less homogeneous, and the network is less redundant, the demand would require a re-distribution to lead to maximum performance. In other words, the idealized MFD is obtained, when the demand perfectly utilizes given supply conditions. Therefore, the demand is not an independent characteristic for the idealized MFD.

Thus, it is clear that the realized MFD is close to the idealized one in case of slow-varying and homogeneously distributed demand in regular networks. This seems to be the case in GEROLIMINIS & DAGANZO [2008] as both MFD definitions appear to be well-matching. They substituted the overall trip completion rate, which theoretically also includes departures inside the region, by the average outflow. Moreover, the authors showed that the average outflow of the network is proportional to the average flow inside the region which is fortunate as the latter is easier to measure. Nonetheless, the realized MFD is *dependent* on the demand and is usually well below the idealized MFD. More specifically, related influential demand factors are its spatial distribution, often represented by origin-destination (OD) tables, and route choice [LECLERCQ & GEROLIMINIS, 2013; JI, DAAMEN, HOOGENDOORN, HOOGENDOORN-LANSER & QIAN, 2010]. Also, the temporal demand profile can impact the MFD as such can lead to non-stationary traffic states [AMBÜHL, LODER, LECLERCQ & MENENDEZ, 2021; LECLERCQ & PAIPURI, 2019].

Other reasons for the reduced observed performance are heterogeneous spatial density distributions [GEROLIMINIS & SUN, 2011b; MAZLOUMIAN, GEROLIMINIS & HELBING, 2010], which can occur due to non-regular network topologies. The specific role of the underlying network topology on the scatter in the MFD was investigated by BUISSON & LADIER [2009], which analyzed loop detector data (LDD) from Toulouse, France, from highways, urban center streets, and residential streets. They confirmed the strong impact of heterogeneity on the MFD's shape and degree of scatter. Nevertheless, procedures to substantially decrease the spatial heterogeneity of the density distribution were developed such as the partitioning of large-scale networks into regional sub-networks [e.g. TSUBOTA, BHASKAR & CHUNG, 2014]. Examples for corresponding partitioning methods can be found in SAEEDMANESH & GEROLIMINIS [2016] and AMBÜHL, LODER, ZHENG, AXHAUSEN & MENENDEZ [2019].

Additionally, the consideration of non-stationary traffic states in the measurements can increase the observed scatter in result in so-called hysteresis patterns as shown in a theoretical study by [LECLERCQ & PAIPURI, 2019]. This is further confirmed by AMBÜHL, LODER, BLIEMER, MENENDEZ & AXHAUSEN [2020] and WU, LIU & GEROLIMINIS [2011]

based on analysis of empirical data. Similar to the FD, the MFD was originally defined for stationary states. However, changes in such do not propagate instantaneously throughout the network, as the propagation speeds of shock waves are finite. This leads necessarily to a measurement of non-stationary states and therefore their appearance in the MFD. The resulting *loops* in the MFD are known as hysteresis patterns. The term 'hysteresis' describes the path dependency of the development of aggregated traffic states in urban networks, i.e. the different spatial density distribution during the onset and offset of congestion [KNOOP, LINT & HOOGENDOORN, 2015; SABERI & MAHMASSANI, 2013; MÜHLICH, GAYAH & MENENDEZ, 2014]. For example, higher flows can be observed during the loading of the network than during the unloading for a given density at a specific location [SHIM, YEO, LEE, HAMDAR & JANG, 2019]. Therefore, the accumulation-based MFD shall be used to predict traffic for merely slow-varying demands, as originally assumed by DAGANZO [2007]. MARIOTTE, LECLERCQ & LAVAL [2017] suggested that the trip-based formulation is more suitable for higher occurrences of non-stationary states than the accumulation-based one since it accounts for trip distances and does not assume the accumulation to remain constant when a user travels in the network. The consideration of such time-dependent aspects leads to the possibility to consider traffic dynamics and therefore non-stationary traffic states such as hysteresis loops.

Applications

The increasing knowledge about the MFD's properties and the growing empirical evidence make the concept a promising tool for a wide range of applications in urban traffic modeling, monitoring, and control. Indeed, there is an increasing body of literature that recognizes the potential of the MFD as a concept for traffic modeling, for the analysis of urban traffic phenomena, and for the design of transportation systems and traffic management and control schemes.

Per definition, MFD-based control systems refer to the network level. The first works concerned perimeter control and were conducted by KEYVAN-EKBATANI, KOUVELAS, PAPAMICHAIL & PAPAGEORGIOU [2012] and GEROLIMINIS, HADDAD & RAMEZANI [2013]. Since then, numerous studies investigated all types of traffic control for cities divided into multiple reservoirs and modeled aggregated traffic dynamics with differing details [e.g. ABOUDOLAS & GEROLIMINIS, 2013; KOUVELAS, SAEEDMANESH & GEROLIMINIS, 2017; ZHONG, CHEN, HUANG, SUMALEE, LAM & XU, 2018; NI & CASSIDY, 2020]. While most studies aimed at reducing congestion, INGOLE, MARIOTTE & LECLERCQ [2020] focused on perimeter traffic control to explicitly lower emissions while describing traffic dynamics based on the MFD. Other studies investigated the trade-off between local and global control [YANG, ZHENG & MENENDEZ, 2018], and combining perimeter control and congestion pricing [YANG, MENENDEZ & ZHENG, 2019]. BATISTA, INGOLE, LECLERCQ & MENÉNDEZ [2021b] revisited the assumption of constant trip lengths of the trip-based MFD formulation in the perimeter control application and found that an accurate trip length estimation is indeed crucial. For an overview of MFD-based perimeter control studies, we refer the interested reader to SIRMATEL, TSITSOKAS, KOUVELAS & GEROLIMINIS [2021], SIRMATEL & GEROLIMINIS [2019]

and HADDAD & ZHENG [2020].

Additional applications regard the effects of route guidance [e.g. YILDIRIMOGLU, SIRMATEL & GEROLIMINIS, 2018; AMINI, TILG & BUSCH, 2018]. For example, DANDL, TILG, ROSTAMI-SHAHRBABAKI & BOGENBERGER [2020] analyzed the differences in route guidance for autonomous mobility-on-demand vehicles based on a classical dynamic traffic assignment and an MFD-based approach. They found that the MFD-based travel time estimations are sufficiently detailed if the network is partitioned accordingly. This approach was then further applied in optimizing the design of a complex multi-modal and multi-stakeholder transportation system [DANDL, ENGELHARDT, HYLAND, TILG, BOGENBERGER & MAHMASSANI, 2021].

The introduction of the bi-modal MFD by GEROLIMINIS, ZHENG & AMPOUNTOLAS [2014] further enabled researchers to propose multi-modal traffic management strategies such as the integration of public transport priority into perimeter control [HAITAO, YANG, LIANG, MENENDEZ & GULER, 2019], or congestion pricing in a multi-modal system [ZHENG, RÉRAT & GEROLIMINIS, 2016]. Furthermore, the bi-modal MFD has been implemented as a performance indicator to allocate bus lanes [e.g. ZHENG, DANTSUJI, WANG & GEROLIMINIS, 2017]. In addition, it was utilized to design entire bus networks accounting for bi-modal interactions [TILG, UL ABEDIN, AMINI & BUSCH, 2020b], but also as an underlying traffic model to analytically quantify mode-specific space consumption [ROCA-RIU, MENENDEZ, DAKIC, BUEHLER & ORTIGOSA, 2020].

Next to the design of traffic management strategies, the MFD was exploited to study the behavior of urban traffic from an aggregated perspective. KNOOP, JONG & HOOGENDOORN [2014] investigated different network topologies and their impact on the MFD. ORTIGOSA, MENENDEZ & GAYAH [2015] examined the network configuration, that is the effects of one-way streets and left-turning possibilities. GIRAULT, GAYAH, GULER & MENENDEZ [2016] analyzed the impacts of signal coordination on the network performance. All these studies found significant impacts of examined characteristics on the MFD. LODER ET AL. [2019b] evaluated a large data set covering traffic data from more than 40 cities, and revealed significant dependencies of the MFD on topological factors such as road network density, intersection density, and others. In an empirical study for Zurich, Switzerland, AMBÜHL ET AL. [2021] identified several different temporal demand patterns of vehicular traffic and quantified their impact on the MFD. In summary, this non-exhaustive list of studies illustrates the wide range of MFD-related applications.

Estimation methods

The numerous applications ranging from traffic management and control to the analysis of urban traffic networks are based on the estimation of the MFD. Generally, corresponding estimation methods can be divided into empirical, simulation-based, and semi-analytical approximate ones. To our best knowledge, the terms ‘analytical’ and ‘semi-analytical’ are often used interchangeably and without precision in the field. Therefore, this dissertation will use the definition suggested by LECLERCQ & GEROLIMINIS [2013] who classified the

methods relevant to us as semi-analytical since they evaluated multiple analytical expressions throughout the process. Moreover, we will denote those semi-analytical methods as 'approximation' to highlight the existence of potential deviations between the estimated and the real MFD.

As illustrated in the sections above, many empirical studies exist and their contributions were valuable to understand the properties of the MFD, as well as the nature of urban traffic. The main limitation of empirical data-based estimations is that corresponding data sets are not always available and rarely cover the entire network of interest in sufficient detail. The collection, maintenance, and management of such data mean additional costs for local authorities, which are mostly the owner of loop detectors being the most important source of traffic data for MFDs. While floating car data (FCD) can also be utilized to estimate the MFD, such data is costly and often too aggregated to allow a detailed analysis. Another problem with the empirical estimation is that in case no massive data sets are available, it mostly fails to allow a systematic analysis of the MFD characteristics and important impact factors. Since the underlying network topology and demand patterns cannot be changed arbitrarily within an analysis study, their effects on the MFD often remain uncovered. Lastly, the data requires extensive pre-processing to identify faulty and missing data, and to account for biases and incomplete network coverage. After all, empirical data is only a snapshot of reality in urban traffic networks.

Some of these problems can be resolved by exploiting microscopic simulation tools. All kinds of data from each simulated vehicle or person can be extracted, and faulty, missing, and incomplete data are thus not a problem. Moreover, these tools allow a detailed analysis of the effects of supply and demand on the MFD. Difficulties arise, however, when a simulation model is not ready to use. The creation, calibration, and validation of such a model are very labor-intensive and costly processes, and thus often not feasible in practice. Moreover, substantial resources are necessary to meet computational power requirements. In addition to that, the computational cost of simulations is a clear obstacle to the analysis of a high number of hypothetical scenarios, real-time applications, or the implementation of the MFD estimation in a model-based optimization framework. Lastly, such simulations might lack explanatory value as the causality between driving behavior of single vehicles and macroscopic traffic phenomena is difficult to expose. On the contrary, semi-analytical approaches can shed light on the physics of urban traffic as cause and effect can be related more directly.

Hence, semi-analytical approximation methods enter the stage in case (i) no extensive, detailed, and well-maintained empirical traffic database exists; (ii) microscopic simulation models are absent, or the associated computational efforts too high; (iii) but an estimated MFD is desired for gaining a deeper insight into the physics of urban traffic, conducting hypothetical scenario analysis, performing real-time applications, or integrating it as part of an optimization framework.

The existing semi-analytical approximation methods all go back to the one described by DAGANZO & GEROLIMINIS [2008]. The authors introduced the method of practical cuts which finds an upper bound for the MFD of a homogeneous urban arterial road. Their method is

essentially based on the variational theory (VT) of traffic flow, which allows addressing KWT problems with precision and efficiency. Note that VT only applies to corridors where no inflows or outflows occur. As these methods and theories are of high importance to this dissertation, they are explained in detail in Chapter 2. A broader perspective has been adopted by LECLERCQ & GEROLIMINIS [2013] who extended the approach to account for more general arterials. In particular, some assumptions were relaxed, and more general supply conditions allowed, e.g. heterogeneous block lengths and signal settings. Other extensions worked on simplifying the mathematical problem formulation [DAGANZO & LEHE, 2016], and on including stochasticity in multi-modal corridors [DAKIC, AMBÜHL, SCHÜMPERLIN & MENENDEZ, 2020] or for corridors with varying topology [LAVAL & CASTRILLÓN, 2015]. The latter is labeled as stochastic approximation (SA) and is also studied in detail in this dissertation (see Chapter 2).

The lower requirements regarding existing data and computational power associated with such models make them a promising alternative to empirical and simulation-based methods. One of the limitations of semi-analytical methods is the potential estimation inaccuracy resulting from the reduced modeling complexity and the model-inherent aggregations. However, research has shown that semi-analytical MFD approximations are useful and lead to satisfactory results. For example, GEROLIMINIS & BOYACI [2012] analyzed the effects of supply-related input parameters on the MFD by utilizing the computational efficiency of semi-analytical approximation methods. Similar to that, DAGANZO, LEHE & ARGOTE-CABANERO [2018] semi-analytically investigated the impact of adaptive signal offsets on the MFD. Furthermore, AMBÜHL ET AL. [2020] introduced a functional form with a physical meaning for the MFD that exploits semi-analytical approximations to analyze urban traffic systems.

One of the greatest challenges for semi-analytical approaches to date is the approximation of the MFD for realistic *networks*. In particular, most existing approaches have failed to explicitly address the network level. They only apply to single corridors and, most importantly, do not account for inflows or outflows at intersections. Hence, they can only be implemented for small and artificially regular networks where the characteristics of the corresponding MFD are similar to those of one derived for a representative corridor. Moreover, some studies utilize a semi-analytical MFD approximation for networks even though they only refer to the corridor level and thus miss logical consistency [e.g. MARIOTTE ET AL., 2017]. While first approaches exist to account for average effects of turning flows on corridor MFDs [e.g. XU, YU & GAYAH, 2020], the literature, hitherto, lacks a thorough methodology to semi-analytically approximate MFDs for realistic urban networks.

1.2 Research questions, scope, and contributions

The review above has made it clear that the MFD is of interest for several practical and research topics and that its semi-analytical approximation at the network level is advantageous for numerous use cases. Moreover, it indicated the research gap this dissertation addresses. The following section states the corresponding research hypotheses in detail, describes the scope of the thesis, presents the contributions, and depicts the thesis outline.

1.2.1 Research questions

Whilst some research has been carried out on the semi-analytical MFD approximation at the corridor level, there have been few investigations into the explicit extension of which to the network level. To date, semi-analytical MFD approximation methods significantly simplify the underlying network topology. Moreover, they neglect the impacts of demand-related aspects and are hence not suitable to estimate the realized MFD. This thesis aims at closing this research gap. From this, the main research question can be formulated:

Can the realized MFD be semi-analytically approximated for urban networks?

Accounting for existing works on the MFD approximation, and their reliance on the concept of VT, which also only applies to corridors, this overarching question is subdivided into several more specific ones:

- **RQ1:** Of the existing semi-analytical MFD approximation methods, is a deterministic or stochastic one closer to reality for a realistic urban corridor?
- **RQ2:** How sensitive are these methods to input parameters reflecting the complexity of urban settings, e.g. signal control and public transport operation?
- **RQ3:** Given the role VT plays for the semi-analytical MFD approximation, how can the concept be extended to the network level?
- **RQ4:** Why is such an extension of VT advantageous compared to other numerical solution methods and microscopic simulation models for network KWT problems?
- **RQ5:** How can the method of cuts (MC) be extended to the network level while taking demand-related aspects such as turning ratios into account?
- **RQ6:** What is the role of assumptions regarding the network-wide propagation of traffic states for such an extension?

1.2.2 Scope

The research conducted within this dissertation focuses on the advancement of methodological aspects of the semi-analytical MFD approximation. While doing so, several assumptions are made and limitations occur. In the following, they are listed briefly. Additionally, they are discussed in more detail in Chapter 6.

- From a traffic modeling perspective, we fully rely on the KWT. This is mainly because existing methods, which our proposed models build upon, rely on the KWT as well. While there are alternative macroscopic flow models, e.g. higher-order models, studies confirmed that models based on KWT replicate realistic traffic dynamics with a reasonable accuracy, not only on freeways [e.g. POLSON & SOKOLOV, 2015], but also in urban areas [e.g. CHOW, LI, SZETO & WANG, 2015; HAN, PICCOLI & SZETO, 2015; LONG, GAO, ZHAO, LIAN & ORENSTEIN, 2011]. This is especially true on an aggregated level, and thus KWT is suitable for estimating network-wide traffic states.

- From the MFD perspective, we focus on the accumulation-based formulation. Note that this type of formulation can be an input to alternative ones, e.g. the trip-based one.
- We target the approximation of the MFD for urban networks. Freeway networks are of less concern, as the corresponding MFDs are usually not well-defined [GEROLIMINIS & SUN, 2011a]. The considered urban networks in this dissertation concern supply facilities such as road segments and signalized intersections. While all kinds of protected movements can be modeled endogenously, merges and conflicting streams are not included yet. They occur, for example, at unsignalized intersections or in roundabouts. For integrating merges into the models, the works of DAGANZO [1995] are a promising starting point. Conflicting streams can be included exogenously based on headway distribution models [e.g. HERZ, SCHLICHTER & SIEGENER, 1976]. Other network traffic-related assumptions are that the queueing discipline follows a first in first out (FIFO) rule. A potential way to relax this assumption is to split links in the model when a turning lane is introduced. Furthermore, we assume that turning ratios are constant over time. This can be relaxed by utilizing a moving average approach for describing turning ratios.
- We do not engage with the empirical validation of the proposed methods. Instead, we merely rely on validation with KWT simulations as we want to isolate the validity of our proposed approach according to KWT at the network level. Comparing KWT to empirical data lies beyond the scope as mentioned above.
- We concentrate on the realized MFD in contrast to the idealized one. Thus, we explicitly consider the impact of heterogeneous spatial demand patterns represented by turning ratios at intersections. However, aspects of temporal dynamic urban traffic such as hysteresis phenomena cannot be modeled at this stage.
- The proposed methodology applies to unimodal traffic albeit certain aspects of multi-modality can readily be incorporated. For example, public transport vehicles can be modeled as moving or stationary bottlenecks and their effects can be quantified. However, we do not include public transport or other modes in the MFD approximation.

1.2.3 Contributions

The findings of this dissertation should make an important contribution to the field of urban traffic flow research. The contributions refer to the semi-analytical MFD approximation but even go beyond that. They are listed below:

- **Experimental analysis of existing semi-analytical MFD approximation methods (RQ1 and RQ2)**
  We conduct an in-depth analysis of the existing semi-analytical MFD approximation methods, i.e. the MC and the SA, by comparing the approximations to empirical data from an urban corridor in Munich, Germany. The analysis includes the examination of an important measurement bias in empirical data and the sensitivity of both semi-

analytical methods to input data related to traffic signal control and public transport operation. To our best knowledge, no other study with such a rich data set including signal data was conducted in the context of MFD approximation. Our results illustrate the ability of the MC to approximate the MFD reasonably well. Thus, the deterministic approximation method outperforms the stochastic one for the given case. Moreover, the sensitivity analysis reveals that the most important input parameters are related to signal control.

- **Extension of variational theory to networks (RQ3 and RQ4)**
  We develop a method, the network variational theory (nVT) that extends VT to apply to urban signalized networks and demonstrate it in a case study. While nVT is relevant for the network MFD approximation, as VT itself is the methodological pillar of MC, the domain of applications goes far beyond that since more general KWT problems can be solved at the network level. Moreover, nVT inherits important properties of the original formulation such as the numerical accuracy and the ability to model complex intra-link bottlenecks. This makes the extended method indeed superior compared to alternative methods for corresponding use cases.

- **Development of a semi-analytical network MFD approximation framework (RQ5 and RQ6)**
  We develop a framework inspired by the MC to semi-analytically estimate the realized network MFD. This framework, the network method of cuts (nMC), is able to account for different spatial demand patterns while deriving the upper bound of the network-wide average flow. Next to the developed nVT, it integrates three more methods to estimate the network-wide capacity and maximum density. This flexibility allows to trade-off estimation accuracy, computational cost, and modeling complexity. Moreover, it allows us to shed light on the role of assumptions regarding traffic state propagation in the MFD approximation, as well as on the impact of spatial demand patterns on the MFD.

1.3 Thesis outline

The remaining thesis is structured as described below. Figure 1.2 visualizes the overall methodology.

Chapter 2 begins by laying out the state of the art of theories and methods relevant for this dissertation. First, a review on KWT as one of the most important macroscopic traffic flow models is given. Based on this, the concept of VT and corresponding solution methods are introduced and set into context. Last, we explain the background on MC which utilizes VT to semi-analytically estimate the corridor MFD. In addition, we explain the SA as a stochastic extension, as well as attempts to apply the MC to the network level.

Chapters 3 to 5 represent the main block of the dissertation. Chapter 3 provides an in-depth experimental comparison of the MC and the SA, displayed as solid and dashed grey curve in the figure. Both methods are evaluated for a corridor in Munich, Germany, and analyzed

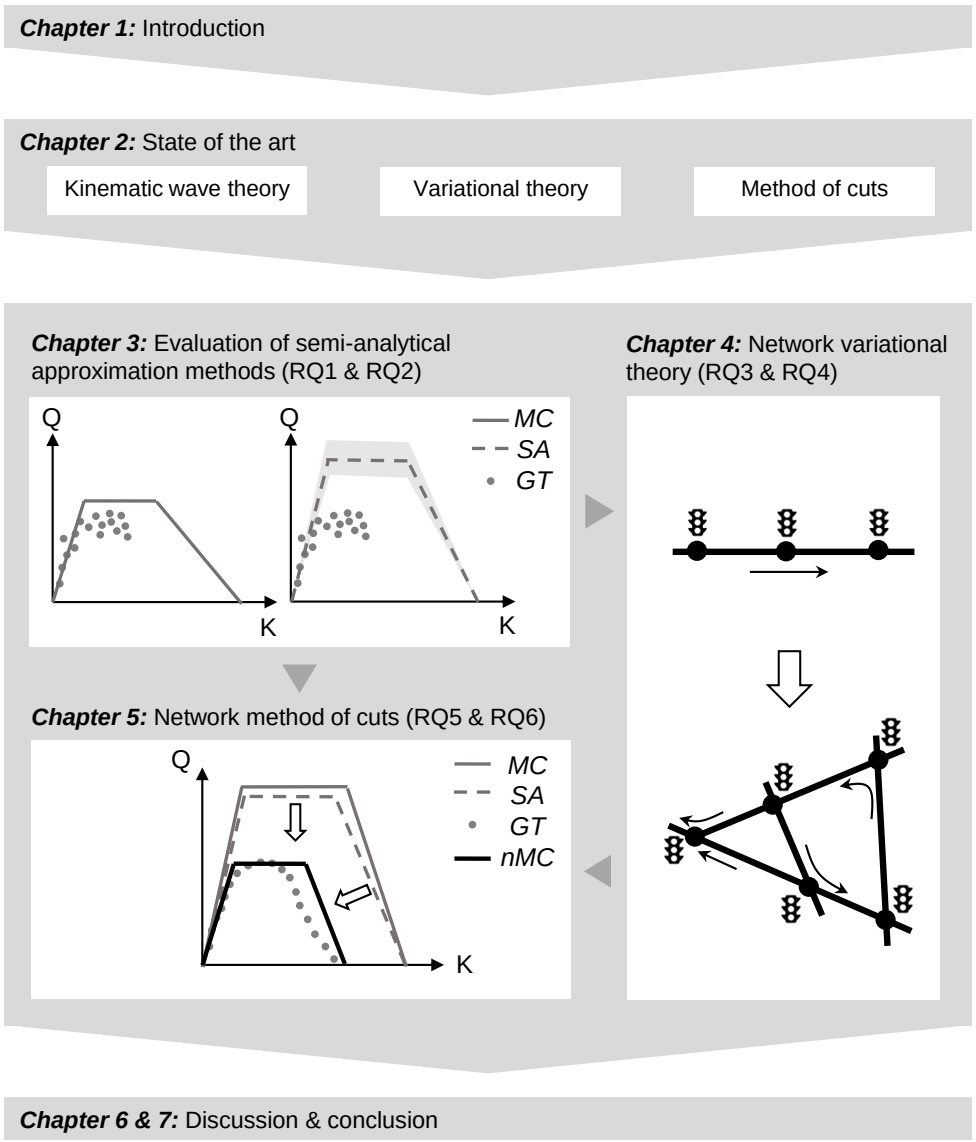


Figure 1.2: Thesis outline. In the course of the dissertation, the method of cuts (MC) and the stochastic approximation (SA) are evaluated by comparing it to a ground truth (GT). The novel developed frameworks are the network variational theory (nVT) and the network method of cuts (nMC).

in light of an empirical data set as ground truth (GT) which is schematically shown as grey circles in the figure. Additionally, a Monte Carlo Simulation-based sensitivity analysis reveals the role of input parameters for both methods.

Motivated by the fact that both investigated semi-analytical MFD approximation methods are based on VT, chapter 4 proposes an extension of VT from the corridor to the network level as illustrated in Figure 1.2. In this chapter, we develop a model, the nVT, which introduces source terms to consider turning flows at signalized intersections within a corridor. This is key to solving a network KWT problem based on the concept of VT. The nVT contributes

to the nMC introduced in Chapter 5 but also applies to problems beyond the pure MFD approximation.

Inspired by the original MC, Chapter 5 develops the nMC framework to semi-analytically estimate the realized MFD at the network level. Thereby, we not only incorporate the developed extension of VT, but also propose two additional methods to estimate the network capacity, and one to approximate the network-wide maximum density. These approximations have a lower estimation accuracy but involve reduced modeling complexity, as well as less computational expense. In a case study, we show that the approximated MFD is substantially closer to the ground truth compared to the MC and the SA as shown in the figure.

Chapter 6 discusses important aspects of the previous chapters. The structure of this chapter follows the logical process of designing, validating, and eventually applying a model. The focus lies on the developed methodologies, i.e. the nVT and nMC. They are critically assessed in the context of these process steps and with regard to existing literature.

Chapter 7 summarizes the thesis including the implications and limitations. Additionally, future work and next steps are outlined.

Chapter 2

State of the art

This chapter explains the theories and methods relevant to this dissertation. First, we present the KWT as the most prominent first-order macroscopic traffic flow model. Then, we explain the concept of VT and a corresponding solution method. Last, we describe the method of cuts including its extensions.

2.1 Kinematic waves: Theory and numerical solution methods

2.1.1 Basic concept

The KWT, also called LWR theory in the literature and proposed in the seminal work by LIGHTHILL & WHITHAM [1955] and RICHARDS [1956], describes the macroscopic evolution of flow $q(x,t)$ and density $\kappa(x,t)$ across time t and space x. The term *macroscopic* stems from the fact that those quantities are modeled in analogy to continuum fluids. The theory is based on the partial differential equation (PDE) of conservation laws (see eq.(2.1)), also known as continuity equation [TREIBER & KESTING, 2013], and on the FD, $q = H(k)$.

$$\frac{\partial \kappa}{\partial t} + \frac{\partial q}{\partial x} = 0, \tag{2.1}$$

As FDs with different functional forms can be utilized within the theory, one rather speaks of a class of LWR models than a single model. The theory is considered a first-order PDE as only one dynamic equation, the continuity equation, is included [TREIBER & KESTING, 2013]. In contrast, higher-order models add additional complexity by introducing another equation to model dynamic speeds instead of relying on an FD [e.g. PAYNE, 1971; WHITHAM, 2011].

SEO, BAYEN, KUSAKABE & ASAKURA [2017] summarized the advantages and disadvantages of these LWR models. The main strength of this theory is that basic traffic phenomena such as the transition between free-flow and congested traffic states can be reproduced, while its simplicity allows applying efficient calibration and solution methods. Substantial limitations are that only stationary traffic states according to the FD can be modeled, and vehicles obtain infinite acceleration. These limitations lead to the inability to model traffic phenomena such as the capacity drop, hysteresis effects, and stop-and-go waves, which are important especially for modeling traffic on freeways.

NEWELL [1993] revisited the LWR theory from the perspective of the cumulative vehicle count $N(x,t)$ in his simplified KWT. It represents a reformulation of the conservation law-based theory into a Hamilton-Jacobi PDE. Such PDE types are well-known and established mathematical solution methods exist [SEO ET AL., 2017; EVANS, 2010]. The simplified KWT describes the cumulative number of vehicles N that have passed point x by time t. The cumulative count N across all (x,t) forms a surface that was introduced by MOSKOWITZ [1965] and is accordingly called the Moskowitz function. For a given point $P(x,t)$, the partial derivatives of this surface with respect to t and x are the flow $\frac{\partial N}{\partial t} = q(x,t)$ and the density $-\frac{\partial N}{\partial x} = \kappa(x,t)$. The KWT as Hamilton-Jacobi PDE is formulated as:

$$\frac{\partial N}{\partial t} - H\left(-\frac{\partial N}{\partial x}\right) = 0. \tag{2.2}$$

Based on the Moskowitz function, the theory can approximate vehicle trajectories and thereby represents an essential connection between macroscopic traffic dynamics and microscopic vehicle behavior. Figure 2.1 highlights this by showing an example of the Moskowitz surface. The x-axis displays the space x, the y-axis the time t, and the z-axis the cumulative count N. Note that we let space and time increase from right to left, to obtain a better view of the surface. Moreover, we display some iso-cumulative count curves to highlight that they indeed can be interpreted as vehicle trajectories.

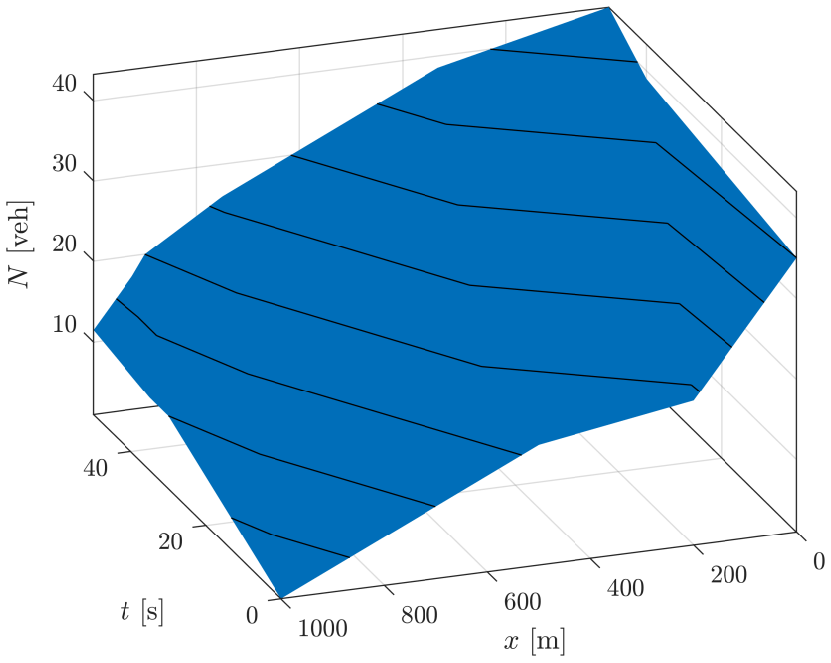


Figure 2.1: Moskowitz surface.

2.1.2 Solution methods

Over the past decades, numerous approaches have been developed to numerically solve KWT PDEs in both conservation law or Hamilton-Jacobi form. LAVAL & LECLERCQ [2013] showed that many established traffic models can be cast under the Hamilton-Jacobi umbrella, and are related to two-dimensional coordinates either of the Eulerian (x, t), the time-Lagrangian (t, N), or the space-Lagrangian (N, x) type.

The majority of solution methods refers to Eulerian coordinates. Examples are the cell transmission model (CTM) [DAGANZO, 1992; DAGANZO, 1995], the link transmission model (LTM) [YPERMAN, 2007; TAMPÈRE, CORTHOUT, CATTRYSSE & IMMERS, 2011; HAN ET AL., 2015; JIN, 2015], the analytical grid-free method by MAZARÉ, DEHWAH, CLAUDEL & BAYEN [2011], and methods based on VT which is an equivalent to KWT [DAGANZO, 2005b; DAGANZO, 2005a; DAGANZO & MENENDEZ, 2005; CLAUDEL & BAYEN, 2010a; CLAUDEL & BAYEN, 2010b]. Substantially less research has explored the time- and space-Lagrangian coordinate systems for solving the PDEs. However, some works based on the VT approach were reported [e.g. MAHUT, FLORIAN & TREMBLAY, 2003; LECLERCQ, LAVAL & CHEVALLIER, 2007; LECLERCQ & BECARIE, 2012; DURET & YUAN, 2017]. A more detailed overview of numerical schemes to solve KWT problems was provided by SEO ET AL. [2017]. For more details on the mathematical representation, we refer to LAVAL & LECLERCQ [2013].

Given a large number of alternative solution methods for KWT problems, one may ask: Which one shall I choose? While the variety of methods is huge, each solution method has its own strengths (and weaknesses). Thus, not surprisingly, the selection of an appropriate KWT solution method is problem-specific. For scenarios with multi-class and multi-commodity traffic traveling through entire networks, the CTM is certainly a popular method. It refers to Eulerian coordinates and belongs to the class of finite difference methods. However, it is known to entail discretization errors which lead to the phenomenon of numerical viscosity [e.g. TREIBER & KESTING, 2013]. This type of error results in shock waves which are 'smoothed' as time evolves. On the one hand, the error converges to zero with decreasing time-step sizes. On the other hand, the resulting finer computational grid increases the computational effort. Another well-known KWT model is the LTM, which is, in contrast to the CTM, numerically precise and very efficient for large-scale networks. For example, networks of the size of Australia's Gold Coast with a length of 2'800 km were modeled with LTM-based frameworks [see RAADSEN, BLIEMER & BELL, 2016]. However, this model has difficulties in modeling complex heterogeneous KWT problems as it does not explicitly account for corresponding aspects. KWT problems become complex and heterogeneous once the underlying FD is space-time dependent, and stationary and/or moving bottlenecks within links occur.

Furthermore, solution methods based on VT have been proposed in both Eulerian and Lagrangian coordinate systems. They calculate the cumulative count $N(x, t)$ with accuracy for bilinear FDs [LAVAL & LECLERCQ, 2013]. In contrast to the Eulerian approach, the Lagrangian one is able to model flows at intersections, vehicle-specific characteristics, and multi-class traffic flow with parsimony. Challenges appear in representing multi-lane highways with changing FDs, bottlenecks, and lane changes [LECLERCQ ET AL., 2007].

More importantly, as confirmed by LAVAL, COSTESEQUE & CHILUKURI [2016], it does neither allow to endogenously nor allow to exogenously consider continuous inflows and outflows within links, i.e. source terms. On the contrary, Eulerian VT can take such source terms into account, albeit merely as exogenous input. On the one hand, this VT formulation only applies at the corridor level due to flow conservation. On the other hand, it enables the solution of complex and heterogeneous KWT problems [DAGANZO & MENENDEZ, 2005]. The solution methods are based on dynamic programming [DAGANZO, 2005b; DAGANZO, 2005a], which derives the cumulative count precisely. In addition, MAZARÉ ET AL. [2011] proposed analytical solutions based on VT which also calculates flow and density exactly. However, to our best knowledge, the effects of source terms cannot be evaluated with this analytical method. As VT is of special interest to this thesis, its basic theory and numerical solution method are explained in more detail below.

2.2 Variational theory

In this thesis, we utilize the Eulerian VT as it is the most suitable method for the problem at hand. Moreover, it is equivalent to KWT from the conceptual perspective [DAGANZO, 2005b; DAGANZO & KNOOP, 2016]. Hereafter, we refer to the formulation in these coordinates as 'VT' and skip the term 'Eulerian' for the sake of brevity.

2.2.1 Basic concept

DAGANZO [2005b] introduced the VT concept, which was known in the fields of PDE since ANGEL & BELLMAN [1972], in the context of traffic flow to solve complex and heterogeneous KWT problems. It was inspired by the three-detector problem [NEWELL, 1993]. To solve this problem, the author proposed a method to intersect the cumulative counts from two different positions to find the cumulative count at any arbitrary position in between. Thereby, he accounted for the speeds of kinematic waves and then applied a minimum operation to find the actual cumulative count at the position of interest. In simple words, he stated that traffic states at the intermediary position are determined by upstream states if free-flow conditions prevail, or by downstream conditions if congestion occurs.

DAGANZO [2005b] generalized this procedure. VT allows to determine the cumulative number of vehicles $N(x,t)$ for a given point $P(x,t)$. The cumulative number forms a continuous surface across all points in the solution space, the Moskowitz surface (see Figure 2.1). It is not differentiable where shock waves occur. The theory assumes the conservation of flows, and also relies on an FD as input. Generally, this FD can be inhomogeneous and can vary across time and space, i.e. $H(\kappa, x, t)$, but is required to be concave and piecewise differentiable. As VT becomes exact for triangular FDs, this thesis only considers such in the following. These FDs are characterized by the free-flow speed u, the backward wave speed w, and the maximum density κ_{max}. The maximum flow, i.e. the capacity, q_{opt} can be calculated at the optimal density κ_{opt}, i.e. $q_{opt} = H(\kappa_{opt})$. Figure 2.2 illustrates the triangular FD and the corresponding parameters.

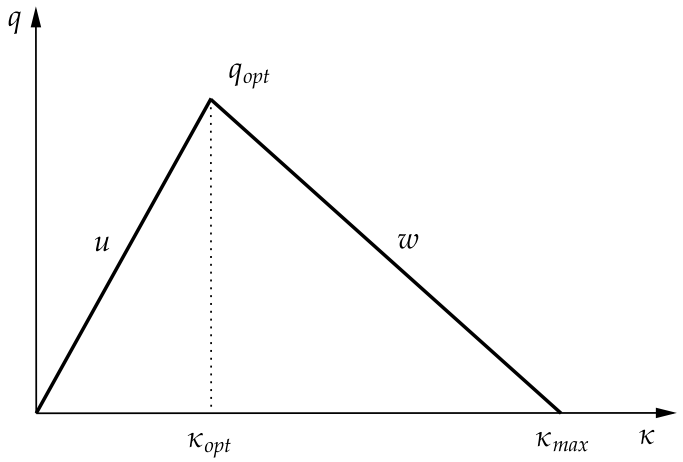


Figure 2.2: A triangular fundamental diagram.

Let us now explain the methodological framework of VT. To facilitate the following explanations, we illustrate the basic concept in Figure 2.3. The overall procedure aims at finding N at

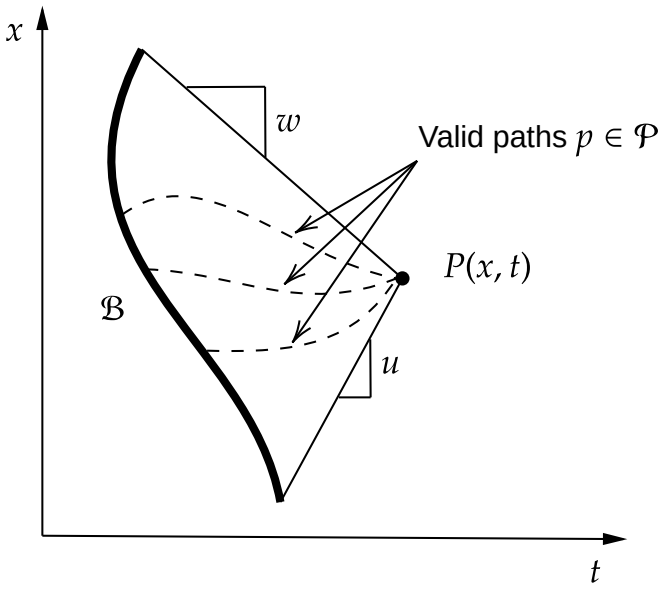


Figure 2.3: Concept of VT.

the generic point $P(x, t)$ as shown in the figure. It requires the specification of a well-posed initial or boundary value problem. This boundary data in form of a cumulative count N_B is specified along a curve $\mathcal{B}$, as illustrated by the thick black line in Figure 2.3. For example, the curve would satisfy $t = 0$ for all x in an initial value problem. First, we find the set of valid paths $p \in \mathcal{P}$ that start at the boundary $\mathcal{B}$ and end at P. A *valid* path p requires its slope to lie between the extremal speeds $v \in [w, u]$ according to the FD. Then, a cost value z is

assigned to each path p depending on its slope. This cost z corresponds to the maximum traffic flow that can pass a moving observer traveling along p. Once all valid paths $p \in \mathcal{P}$ are identified, and the associated cost values z_p and initial or boundary values $N_{\mathcal{B},p}$ assigned, the cumulative count N_P can be found by formulating a least-cost problem:

$$N_P = \min_{p \in \mathcal{P}} \{N_{\mathcal{B},p} + z_p\}. \tag{2.3}$$

In other words, to eventually determine N_P we need to find the path p which results in the minimal increase of N from the boundary $\mathcal{B}$ to the point P.

2.2.2 Solution methods

The first solution methods for VT are based on dynamic programming and were introduced along with the concept itself by DAGANZO [2005a] and DAGANZO & MENENDEZ [2005]. The solution space is defined as the time-space region for which the evolution of traffic according to KWT shall be found. This solution space is then covered with a digraph where the nodes represent points at which N is evaluated, and edges are valid paths between those nodes. Each edge has an associated cost z as described above. DAGANZO [2005a] showed that for piecewise linear FDs there are graphs in which shortest paths are equal to the continuous shortest paths between all valid node pairs. Therefore, these graphs are defined as 'sufficient'. A sufficient graph provides exact results for N_P, if there is no sampling error in the boundary data $N_{\mathcal{B}}$. Furthermore, DAGANZO [2005a] proved that for triangular and homogeneous FDs, all valid paths have extremal slopes, are optimum, and all validly connected networks sufficient. They further stated that the procedure is exact if the FD is triangular and homogeneous, the graph is validly connected, the boundary data are linear at the origins of the graph, and the edges are straight with speeds either equal to u or w. Additionally, DAGANZO & MENENDEZ [2005] showed that bottlenecks can be introduced in the solution space as paths. Such a path denotes when and where the bottleneck is active in the time-space plane. The associated costs reflect the bottleneck capacity. For example, a red phase of a traffic signal would be represented by a straight path with constant x, a temporal length equal to the phase duration, and costs z equals zero.

The definition of a graph that covers the solution space allows us to determine the cumulative count N based on eq.(2.3) for each node. In general, multiple possibilities for building such a graph depending on the problem were introduced by DAGANZO & MENENDEZ [2005]. One of them is the 'lopsided' graph where nodes in the graph align horizontally. They are connected horizontally and by slanted edges with slopes equal to the extremal wave speeds. To obtain such a graph, the absolute value of the ratio of u and w is required to be integer-valued. We define this parameter as $\theta = |\frac{u}{w}|$. Note that VT does not require θ to be integer-valued in general. This graph represents a discretized grid with the temporal and spatial distances Δt and $\Delta x = u\Delta t$, respectively. Hereafter, we denote this lopsided variational graph as G. The cumulative number $N(x,t)$ where (x,t) are the coordinates of nodes on G, can then be

found as follows [DAGANZO & MENENDEZ, 2005; LECLERCQ & PAIPURI, 2019]:

$$\begin{aligned} N(x,t) = \min\{&N(x-\Delta x, t-\Delta t), \\ &N(x+\Delta x, t-\theta\Delta t) + \Delta x \kappa_{max}, \\ &N(x, t-\Delta t) + \beta\}. \end{aligned} \tag{2.4}$$

The terms in this equation can directly be related to the traffic waves or an existing stationary bottleneck. More specifically, the first term refers to those traffic states which travel with free-flow speed u from upstream locations, the second one to congested states which travel with the backward wave speed w from downstream locations, and the third term enables to account for active stationary bottlenecks and the associated capacity constraints at the same location as the target node. For example, a traffic signal can be represented by this term. Note that moving bottlenecks need to be integrated manually in G when applying eq.(2.4).

Recall that the cost z of an edge corresponds to the maximum rate at which traffic can pass a moving observer traveling at a speed corresponding to the slope of that edge. Therefore, the cost of edges representing the travel of free-flow states is $z = 0$, as no one can overtake a corresponding moving observer. The edge related to the propagation of congested traffic states has costs of $z = \Delta x \kappa_{max}$. The parameter β equals edge cost z representing stationary bottlenecks. For example, if we want to model a traffic signal at position x, we set β according to the capacity during each phase. That is, $\beta = 0$ if it is red during the interval $[t - \Delta t, t]$ and $\beta = H(\kappa_{opt})$ if it is green and the capacity is determined by the FD. In summary, eq.(2.4) allows solving a KWT problem representing any signalized or unsignalized corridor with a triangular FD, where θ is integer-valued, according to VT. Note that the trajectory of any bottleneck needs to be fully mapped in G to be considered in the solution.

To further illustrate eq.(2.4), we present an excerpt of G in Figure 2.4. As described above, the graph G is represented by a grid where the nodes are spatially and temporally spaced with Δx and Δt, respectively. For this example, we set the absolute ratio of the free-flow speed and the backward wave speed as $\theta = 2$. Additionally, the figure depicts the three 'from nodes' and the corresponding edges to $P(x,t)$: the upstream edge representing free-flow propagation, the downstream edge representing congestion propagation, and the horizontal edge representing stationary bottleneck constraints.

Applications and methodological extensions of VT are shown in Chapter 4, as they are primarily relevant for the development of the model as presented in that chapter.

2.3 Method of cuts

2.3.1 Basic concept

Based on the tenets of VT, DAGANZO & GEROLIMINIS [2008] introduced the MC as a semi-analytical approximation of the MFD. To develop the theory, they focused on a ring road as a simple network where flow conservation holds per definition. Representing this ring road

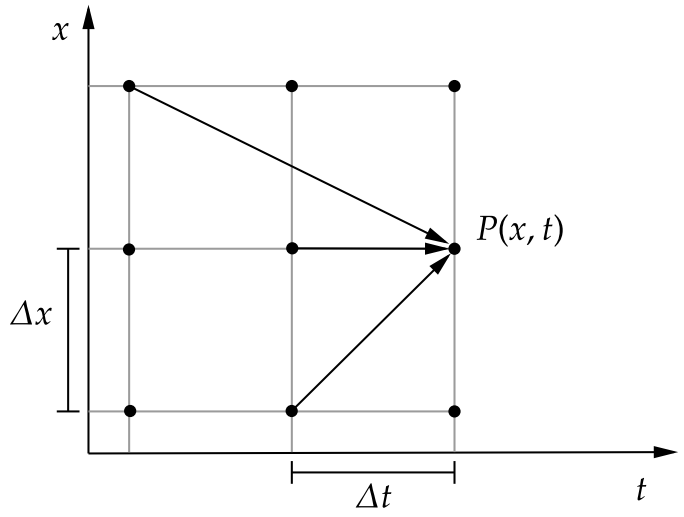


Figure 2.4: Excerpt of the variational graph [TILG, AMBÜHL, BATISTA, MENENDEZ & BUSCH, 2021b].

in the time-space plane, i.e. with a one-dimensional space axis, implies the definition of an upstream and downstream end of the road. As these ends are equivalent in the actual topology, the observed flows must always match. Also, the average density of the road remains constant. Hence, only stationary traffic states exist considering a sufficiently long time horizon. Moreover, the authors assumed a homogeneous FD, uniform block lengths, and equal signal control settings for all intersections within the considered road.

Then, the authors posed an initial value problem of an extended version of the considered road. They connected infinite copies of the original time-space representation of the road in space to generate an infinite street. This led to a well-posed KWT problem and therefore allowed them to apply VT to find the cumulative count N_P at a point $P(x_0, t_0)$ with $t_0 \to \infty$. Such cumulative vehicle count can be found by solving eq.(2.3). It requires the initial vehicle number and the costs along the chosen path. DAGANZO & GEROLIMINIS [2008] assumed the initial number at $(x_0, t = 0)$ as zero. Given the average density, the initial number $N_{\mathcal{B}}$ for all points apart from x_0 at the boundary $\mathcal{B}$ with $t = 0$ could be found. As all moving observers reach $P(x_0, t_0)$, their starting location x on the boundary $t = 0$ depends on their speed v. Following this, DAGANZO & GEROLIMINIS [2008] defined the initial values as $\kappa \cdot v \cdot t_0 \pm \kappa \cdot l$ where l is the length of the original street. The costs $R(v_m)$ are an upper bound for the flow that passes, on average, the moving observer m traveling along p:

$$R(v_m) = \lim_{t_0 \to \infty} \inf_p \{z_p : v_p = v\} / t_0 \tag{2.5}$$

Having defined both, the initial values on the boundary $\mathcal{B}$ and the costs for a path p, the cumulative count N_P can be found based on eq.(2.3) as $N_P = \inf_v \{\kappa \cdot v \cdot t_0 \pm \kappa \cdot l + R(v_m) \cdot t_0\}$. Dividing both sides by t_0 leads to the final formulation for the flow:

$$q(\kappa) = N_P / t_0 = \inf_{v_m} \{\kappa \cdot v_m + R(v_m)\}, \tag{2.6}$$

This flow approaches a location-independent limit for the stationary state and a given average density.

Let us attempt to provide more intuition for the method. Recall the well-known concept of moving observers of traffic flow theory [DAGANZO, 1997]. The observed flow of a moving observer is $q_m = q - \kappa \cdot v_m$ where v_m is the observer's speed. Given the large size of the initial value problem posed by DAGANZO & GEROLIMINIS [2008], the represented street appears to be in a steady-state despite the existence of signals, etc. Thus, for a known average speed v_m of the moving observer, and the observed flow q_m, we can bound the actual flow q given the density κ with the moving observer formula. Many moving observers travel at different speeds through the solution space and each finds an upper limit for the flow. These limits can be illustrated as straight lines in the (q, κ) plane by the function $q = \kappa v_m + R(v_m)$ with the observer speed v_m as the parameter [DAGANZO & GEROLIMINIS, 2008]. Such a straight line is the 'cut' where the method's name stems from. The lower envelope of all cuts is the approximated MFD.

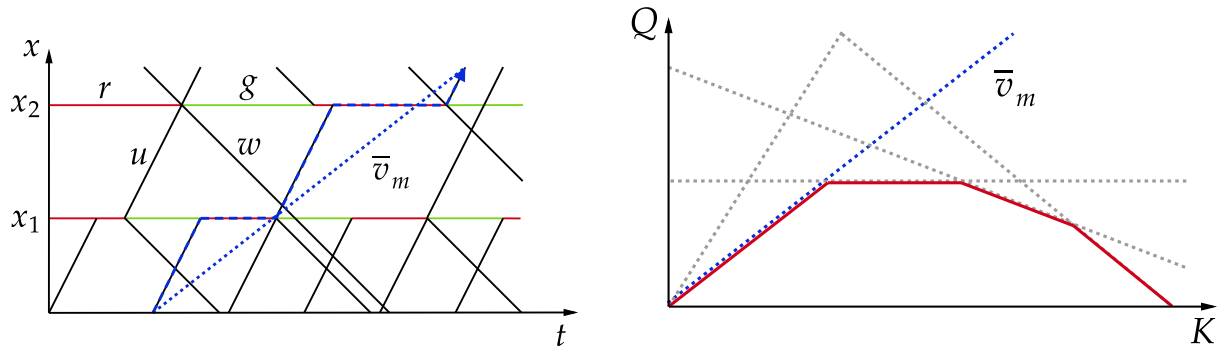
2.3.2 Solution methods

Evaluating eq.(2.6) for all v is tedious. Thus, DAGANZO & GEROLIMINIS [2008] proposed to only evaluate three families of 'practical' cuts to approximate the MFD. They restricted the possible observer speeds to $v = \{u, 0, w\}$, and the observers could not reverse directions while traveling in the variational graph. They stated that the resulting MFD is well approximated but might not be 'tight', i.e. more constraining flow limits might exist.

Later, LECLERCQ & GEROLIMINIS [2013] proposed to properly define a graph in the solution space to encompass all relevant moving observer paths, the so-called global variational graph. This graph lets observers reverse their directions and hence find the corresponding shortest paths in general. The related input parameters refer to links and intersections in the considered corridor. Of the links, the lengths, the number of lanes, and the FDs are required. Of the intersections, the signal control parameters are needed. In contrast to the method of practical cuts by DAGANZO & GEROLIMINIS [2008], fewer regularities are necessary, i.e. signal timings and block lengths were allowed to vary between intersections. However, green times and cycle lengths are not allowed to vary within a specific intersection, i.e. only fixed-time signal control can be modeled. The overall problem setting for the given road is referred to as *hyperlink*. Based on it, similar to VT, the variational graph in the time-space plane is constructed with nodes at the location of intersections and edges connecting those nodes. The edges have slopes equal to $v = \{w, 0, u\}$. In the graph, the nodes are created at the time of a signal phase change for each intersection. Slanted edges start at these nodes and end at the next horizontal edge representing a red phase at the downstream or upstream intersection. As in VT, costs are associated with each edge representing the number of vehicles that can pass a moving observer traveling along the edge. The costs of horizontal edges depend on the signal phase. In case the MFD is approximated for irregular corridors, multiple 'time windows' in the variational graph are considered. This leads to several cuts with the same slope, but potentially varying costs. The average costs are then considered for the actual cut. LECLERCQ & GEROLIMINIS [2013] showed that this set of rules

ensures the generation of a *sufficient* graph in the sense of VT, even for irregular corridors, and therefore the MFD approximation can become tight.

For illustration purposes, we show the procedure to derive a single cut as an example in Figure 2.5. The base for a cut is a global variational graph, see Figure 2.5a. This graph corresponds to a corridor with two intersections located at x_1 and x_2. The horizontal edges represent green and red phases, g and r. The slanted edges start at bottleneck termini, as described above. They intersect with horizontal edges representing green phases and end at those representing red phases. The path of a moving observer m is shown as a blue



(a) Example of the variational graph including the path of observer m.

(b) MFD defined by a family of cuts, including the one corresponding to observer m.

Figure 2.5: Variational graph including a moving observer's path and its translation to the MFD [TILG, AMBÜHL, BATISTA, MENENDEZ, LECLERCQ & BUSCH, n.d.].

dashed line in the graph. This observer only travels on edges that represent red phases and slanted ones that have a slope equal to the free-flow speed. Therefore, the costs $R(v_m)$ are zero as no flow can overtake the observer on its path. Furthermore, the average speed $\overline{v}_m$ is displayed as a blue dotted curve. Based on the average speed and the total cost, we can define a cut in the (Q, K) plane as shown in Figure 2.5b. Interpreting a cut as a linear equation in the (Q, K) plane, the costs are the intercept and the average speed equals the slope. This procedure is then repeated for many different paths in the graph, and thereby generating a family of cuts, shown as the grey dotted lines in Figure 2.5b. These cuts represent the upper limit for the average flow Q. Finally, the MFD approximation is then the lower envelope of all cuts.

2.3.3 Stochastic formulation

Accounting for stochasticity in signal timings and block lengths in urban corridors, LAVAL & CASTRILLÓN [2015] proposed the SA to approximate the MFD. It builds upon the concept of the MC but only evaluates three different types of cuts which are derived based on a variational graph. However, in contrast to the MC, the related moving observers start at a common point in the solution space and travel backward in time. More specifically, the cuts are derived based on the following strategies:

- Strategy s_0 corresponds to a stationary observer with a mean moving observer speed equal to zero, $v_m = 0$.
- Strategy s_1 corresponds to moving observers traveling at a speed $v_m = [u, w]$. Additionally, they stop whenever they reach an edge in the variational graph which represents a red phase. Then, they follow the corresponding edge as it represents a shortcut in the context of VT. At the beginning of the red phase, they resume their previous speed until they reach the next edge representing a red phase.
- Strategy s_2 differs from s_1 only by the fact that moving observers stop at every horizontal edge in the variational graph, independently of it representing a green or red phase.

Furthermore, the SA applies a shear transformation to the FD leading to an symmetric and normalized diagram. Such a transformation of the FD was also exploited in other studies [e.g. DAGANZO & KNOOP, 2016]. This transformation does not alter the flow q. Only the density κ is affected, and the transformed density is denoted as κ'. This procedure leads to the fact that the FD is only a required input parameter if the resulting MFD shall be shown in the usual flow and density plane. The general application of the SA does not require the FD as an input. The necessary input parameters refer to the considered corridor, and are the mean green μ_g and mean red μ_r times, the mean block length μ_l, and the related coefficients of variation δ which are assumed to be equal for simplicity. Based on the three strategies, the transformed FD, the supply-related input parameters, and the renewal reward theory, LAVAL & CASTRILLÓN [2015] formulated analytical equations to approximate the MFD. These equations correspond to the cumulative distribution function for each cut, which is assumed to follow a Gaussian distribution. Consequently, certain percentiles of the MFD for a given density can be specified. Hence, the MFD can be represented stochastically and calculated as:

$$F_{q(\kappa')}(q) = 1 - (1 - F_{s_0}(q))^B \prod_{s \in \Omega} (1 - F_{s,\kappa'}(q)) \tag{2.7}$$

where $F_{q(\kappa')}(q)$ denotes the cumulative distribution function of the MFD, q the flow, κ' the transformed density, s_0 and $\Omega = \{s_1, s_2\}$ the different strategies, and B the number of intersections. Compared to the MC, the fundamental difference is that signal timings are not required to be constant. Thus, the SA does not only apply to scenarios with fixed-time control but also with adaptive or actuated signal control.

2.3.4 Extensions to the network level

The MC estimates a corridor MFD which is independent of demand given the assumptions that no turning flows at modeled intersections exist. Therefore, only one OD pair exists, and demand in this context only refers to the temporal evolution, i.e. the loading and unloading of the corridor. Given the assumption of DAGANZO & GEROLIMINIS [2008] that such temporal changes are slow, only stationary traffic states exist in the corridor, and these are all represented by the MFD. Thus, the approximation by the MC indeed is independent

of the demand under these circumstances. However, it also becomes clear that those assumptions can easily be violated in reality, e.g. when the network MFD is estimated. A fast-varying *temporal* demand profile represents such a violation and research has shown that deviations from the semi-analytical MFD approximation can occur [LECLERCQ & PAIPURI, 2019]. Also, heterogeneous *spatial* demand patterns in networks can substantially affect the shape of the realized MFD. The approximation of the idealized MFD becomes even more intricate, as the optimal demand for given network supply has to be identified. These considerations illustrate that the application of MC to networks implies a substantial reduction of the network's complexity since it can only be represented based on corridors without transferring flows between them. The accompanying simplification of actual demand and supply conditions inevitably introduces a bias of unknown scale.

To date, several attempts to apply MC at the network level were reported in the literature. DAGANZO & GEROLIMINIS [2008] proposed to identify a representative corridor of the network and evaluate the practical cuts to approximate the MFD. They stated that if the postulated regularity conditions regarding supply and demand apply, the estimated MFD is expected to be close to the real one, even though their methodology represents a substantial reduction of the real network's complexity. As explained above, LECLERCQ & GEROLIMINIS [2013] extended the practical cuts to more irregular corridors. The authors applied their method to a network, which consisted of three parallel corridors only interacting at the upstream and downstream ends. GIRAULT ET AL. [2016] studied bidirectional grid networks and estimated the corresponding network MFD. They proposed to derive four corridors that represent the average network topologies in cardinal directions. Then, they utilized the MC and averaged the resulting corridor MFDs to derive the network one. Of course, this approach neglects spatial demand patterns and resulting inter-dependencies of traffic states across the corridors. For example, spillback propagation between different corridors could not be modeled. Additionally, their approach implied the intersection with the smallest green-to-cycle ratio as the most constraining one and that it is saturated once the capacity is reached. However, this is not necessarily the case if substantial turning flows exist. As described above, LAVAL & CASTRILLÓN [2015] proposed the SA by integrating the aspect of stochasticity into the philosophy of the deterministic MC. They conjectured their approximation also to apply to the network level due to the ability to account for stochasticity although they did not consider spatial demand patterns.

Nevertheless, this aspect is receiving more attention from researchers recently, and initial attempts to include the effects of turning flows into the concept of MC were reported. GEROLIMINIS & BOYACI [2012] did not estimate a network MFD, but took into account the effects of inflows to a limited extend in their MFD approximation. In particular, they estimated the effects such flows could have for vehicles traveling on a certain route. Thus, they did not account for inflows at intersections in the MFD estimation, but only their bottleneck effects. In other words, they estimated a route MFD instead of a corridor one. Recall that we defined a corridor as an ordered sequence of links. In distinction to that, we define a route as an ordered sequence of links from the perspective of a vehicle trip given its origin and destination. Consequently, certain routes correspond to a corridor, others do not.

Building upon the SA, AGHAMOHAMMADI & LAVAL [2021] proposed an approach that utilizes

a maximum likelihood estimation to improve the approximated MFD in case networks are considered. While the first results look promising, the logic behind the model still refers to a single corridor. Also based on the SA, an extension to consider the effects of turning flows was provided in XU ET AL. [2020]. For a theoretical two-ring model, the authors developed an approach based on a Markov process. Turning flows were considered by a global ratio which was assumed to be equal across all intersections within the network. This again represents a strong reduction of the network's complexity. Further approaches to account for the effects of turning flows refer to the two-ring model [GAN, JIN & GAYAH, 2017; JIN, GAN & GAYAH, 2013]. Again, these analytical methods are strongly limited to very regular and simplified topologies, and cannot easily be extended to realistic urban networks.

In conclusion, no comprehensive semi-analytical approach to estimate the realized MFD applying to realistic urban networks exists. While promising research activities exist, these are seemingly not able to account for interdependencies between corridors. The most widely used methods are the deterministic MC by LECLERCQ & GEROLIMINIS [2013] and the SA by LAVAL & CASTRILLÓN [2015]. Therefore, in the next chapter, we compare them based on a thorough experimental analysis. As the concept of VT is the methodological pillar for the MC, we aim at extending its application to networks in Chapter 4. Finally, we investigate the extension of the semi-analytical MFD approximation to networks in Chapter 5.

Chapter 3

Evaluation of semi-analytical approximation methods

This chapter is based on the following publications:

- G. TILG, S. AMINI & F. BUSCH [2020a]: Evaluation of analytical approximation methods for the macroscopic fundamental diagram. In: *Transportation Research Part C: Emerging Technologies* 114, pp. 1–19. ISSN: 0968-090x. DOI: doi.org/10.1016/j.trc.2020.02.003
- G. TILG, S. AMINI & F. BUSCH [2019]: Arterial macroscopic fundamental diagram: A comparison of analytical approximations and empirical data from Munich. In: *Presented at the 98th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA

3.1 Introduction

Semi-analytical approximation methods are appealing for use cases where empirical data are absent, and no microscopic simulation model is available, or the application of such is not expedient (see Section 1.1). Even though these approximations currently only apply to the corridor level, the numerous applications highlight their relevance for MFD-based modeling, monitoring, and control of urban traffic. Naturally, this leads to the question of how well empirically derived MFDs can be approximated. Prominent semi-analytical approximation methods are the original method of practical cuts [DAGANZO & GEROLIMINIS, 2008] and the SA by LAVAL & CASTRILLÓN [2015] which have both been validated with data from Yokohama, Japan. However, to our best knowledge, no such validation has been reported for the MC for inhomogeneous corridors, nor has the SA been directly compared to the MC for such a corridor. Additionally, many advances on empirical MFD estimation methods have been achieved to date that were not considered in the original validation studies. Signal data is an input for both semi-analytical approximation methods. Yet, no information on detailed signal data is available for the case of Yokohama. Therefore, it remains unclear to what extent the existence of such data can improve the approximation accuracy of semi-analytical methods. Furthermore, extensions to consider the effects of public transport operation

exist for both methods [XIE, CHIABAUT & LECLERCQ, 2013; CASTRILLON & LAVAL, 2017]. Nonetheless, no reports about any evaluation of them based on comparison to empirical data are known.

This chapter addresses these research gaps by comparing the MC and the SA including their most recent extensions to consider public transport operation to an empirically derived corridor MFD. Even though the SA applies technically seen to networks, conceptually it refers to corridors. Also, the MC was originally developed for corridors. Consequently, we choose a corridor for this study as a topology to be evaluated. The empirical data at hand consists of LDD and signal data. The LDD were recorded at an inhomogeneous corridor segment in Munich, Germany, with multi-modal traffic, active actuated signal control, and transit signal priority. The occurring modes include private cars and buses, trams on physically separated tracks, and bicycle traffic on bike lanes. This setting allows investigating the effects of infrastructure complexity on differences between empirically and semi-analytically derived MFDs. The three dimensional passenger MFD [e.g. GEROLIMINIS ET AL., 2014; LODER, AMBÜHL, MENENDEZ & AXHAUSEN, 2017] extends the idea of the MFD to multi-modal systems. It relates the passenger production to the network bus and car accumulation. However, such an analysis requires additional data and lies beyond the scope of this chapter. Thus, we focus on the MFD approximated by the MC and the SA including their public transport extensions. Thereby, we compare the corresponding results to empirical data from the chosen corridor in Munich. The results of this comparison are thoroughly analyzed. We set the focus on the measurement bias inevitably included in LDD, the signal data, and the impact of public transport operation. No comparison of the semi-analytical methods including the public transport extension has been conducted so far. Also, no investigation of the role and benefit of a rich signal data set to accurately estimate the MFD has been reported. Lastly, to our best knowledge, no attempts to quantify the effects of these elements on the difference between empirical and semi-analytically approximated MFDs have been presented.

Interestingly, the results of this chapter indicate that applying the MC for a wide range of signal data leads to an aggregated traffic pattern close to the empirical MFD. This is the main contribution of this chapter. As an overall result, this study sheds light on the performance of the studied semi-analytical approximation methods, as well as on the importance of corresponding assumptions and simplifications.

The remainder of the chapter is structured as follows. Section 3.2 describes the current state of research regarding the empirical MFD estimation. Section 3.3 specifies the general methodology of the analysis including a case study, the applied methods, the investigation of the measurement bias, and the sensitivity analysis of input parameters related to signal control and public transport operation. The results of the case study and their discussion are presented in Section 3.4. Furthermore, the section includes the results of the measurement bias analysis based on a microscopic simulation in SUMO [LOPEZ, BEHRISCH, BIEKER-WALZ, ERDMANN, FLÖTTERÖD, HILBRICH, LÜCKEN, RUMMEL, WAGNER & WIESSNER, 2018], and of the sensitivity analysis of the semi-analytical approximations. Finally, Section 3.5 draws closing conclusions, highlights limitations, and briefly outlines possible future research.

3.2 Empirical estimation of the macroscopic fundamental diagram

The empirical estimation of the MFD is based on data from fixed sensors or probe vehicles. This section presents related estimation techniques and drawbacks associated with various data sources. However, potential ways of dealing with such drawbacks are not listed here. Instead, we refer to the corresponding original studies.

Loop detectors measure vehicle counts and the duration they are occupied. The data are often aggregated to 3- or 5-minute intervals. Thus, average flows during these intervals, as well as occupancy values in percent can be derived. LDD might be faulty since malfunctions can occur during operation. However, data recording systems apply basic algorithms which help to exclude the majority of technology-induced measurement errors. The first MFD based on LDD was shown in GEROLIMINIS & DAGANZO [2008]. It was derived by averaging the link-weighted network density and flow. Even though a well-defined MFD was found for that case, empirical MFDs have to be treated with care as the literature identified several biases associated with LDD. For example, WU ET AL. [2011] showed that high scatter in the MFD for a corridor can result from the fact that loop detectors also capture non-stationary states. AMBÜHL ET AL. [2017] described the loop detector placement bias. It refers to the fact that loop detectors only measure traffic states at a specific position, which is then often regarded as representative for the entire link. In addition, COURBON & LECLERCQ [2011] argued that loop detector positions need to be distributed uniformly across the links of the network to represent average traffic states on a global level. ZOCKAIE, SABERI & SAEDI [2018] found that not only the position of loop detectors is crucial, but also the link selection plays an important role in the estimation of the MFD. This is because the measured traffic on selected links is related to OD pairs. For a valid MFD estimation, these ODs have to represent the overall OD distribution in the network. Another bias associated with the link selection was described in AMBÜHL ET AL. [2017]. The authors argued that loop detectors are usually installed on links more prone to congestion. Thus, this might result in an overestimation of network-wide congestion. Despite these biases, LDD are a commonly utilized data type for the MFD estimation due to their high availability compared to other data types.

Probe vehicles can be an alternative data source to estimate the MFD [e.g. GEROLIMINIS & DAGANZO, 2008]. These data come either from GPS devices, mobile phones, or are the result of processed Bluetooth data [e.g. TSUBOTA ET AL., 2014]. They are often referred to as FCD. They are considered to represent a better spatial coverage of traffic states in the network [NAGLE & GAYAH, 2014]. LECLERCQ, CHIABAUT & TRINQUIER [2014] showed that the MFD estimation based on probe vehicle data is more accurate than based on LDD. However, a crucial element is the penetration rate. It describes the relative number of probe vehicles in the total flow which is difficult to estimate from empirical data but essential for the MFD estimation. A potential way to cope with this drawback was shown by NAGLE & GAYAH [2014] who fused FCD with limited LDD. However, they assumed that probe vehicles are uniformly distributed throughout the network. This might not be always true since these data are often related to certain vehicle types (e.g. taxi data) which might drive only on specific routes, and thus biases might be introduced. Consequently, this leads to

a heterogeneous distribution of trajectories. DU, RAKHA & GAYAH [2016] investigated the effects of such heterogeneously distributed FCD on the MFD including data from probe vehicles with varying trip lengths. More analyses of the fusion of LDD and FCD for MFD estimation were reported in ZOCKAIE ET AL. [2018] and AMBÜHL & MENENDEZ [2016]. Furthermore, DAKIC & MENENDEZ [2018] showed that data from automated vehicle location devices installed in public transport vehicles can contribute to the empirical MFD estimation.

Another kind of situation occurs in the case of microscopic simulation data. Under these circumstances, all vehicle trajectories are known and can be extracted. This enables the calculation of the average flow, density, and speed based on Edie's definition [EDIE, 1963]. In such a case, no measurement biases are included in the data. SABERI, MAHMASSANI, HOU & ZOCKAIE [2014] showed the existence of the MFD for a microscopic simulation of Chicago, Illinois, and Salt Lake City, Utah. They found that even without any measurement biases, the MFD still shows a range of flows for a given density. More specifically, they presented a hysteresis pattern in the MFD which occurs due to inhomogeneous congestion in the network. This indicated that a data-based estimation of the MFD will always show considerable scatter. This is true for both empirical and bias-free simulated data. Potential reasons for such, e.g. heterogeneous density distribution and the hysteresis phenomenon, were discussed in Chapter 1.1.

3.3 Methodology

This chapter performs an up-to-date evaluation of the existing methods to semi-analytically approximate the MFD. A preliminary analysis for a data set of one working day of a corridor segment in Munich, Germany, showed that the occurring differences cannot be explained by a simple scenario-specific evaluation [TILG ET AL., 2019]. Thus, a more sophisticated approach is proposed in this section. The framework for the general methodology is shown in Figure 3.1.

First, we conduct a case study based on a segment of the urban corridor Leopoldstraße, located in the city of Munich, Germany. The choice of a corridor seems to be appropriate for this study since the MC was originally developed for corridors, and the SA conceptually also refers to such. Thus, the choice of a corridor minimizes possible error sources. The semi-analytical approximation and the empirical estimation methods are applied based on data recorded on this road. The purpose of this case study is to investigate the fit of the semi-analytical MFD approximations to the empirical MFD. Furthermore, it is the starting point for a systematic study of the appearing differences between the resulting MFDs.

As mentioned before, LDD include a measurement bias that can affect the empirical MFD and thus also influence the difference to semi-analytically approximated MFDs. Therefore, a second step is to investigate this measurement bias. This is done by the means of microscopic simulation. The calibration of the simulation is carried out based on empirical and virtual LDD. To estimate the bias-free MFD, virtual trajectory data are used to calculate average flow and density according to Edie's method as shown in LECLERCQ ET AL. [2014].

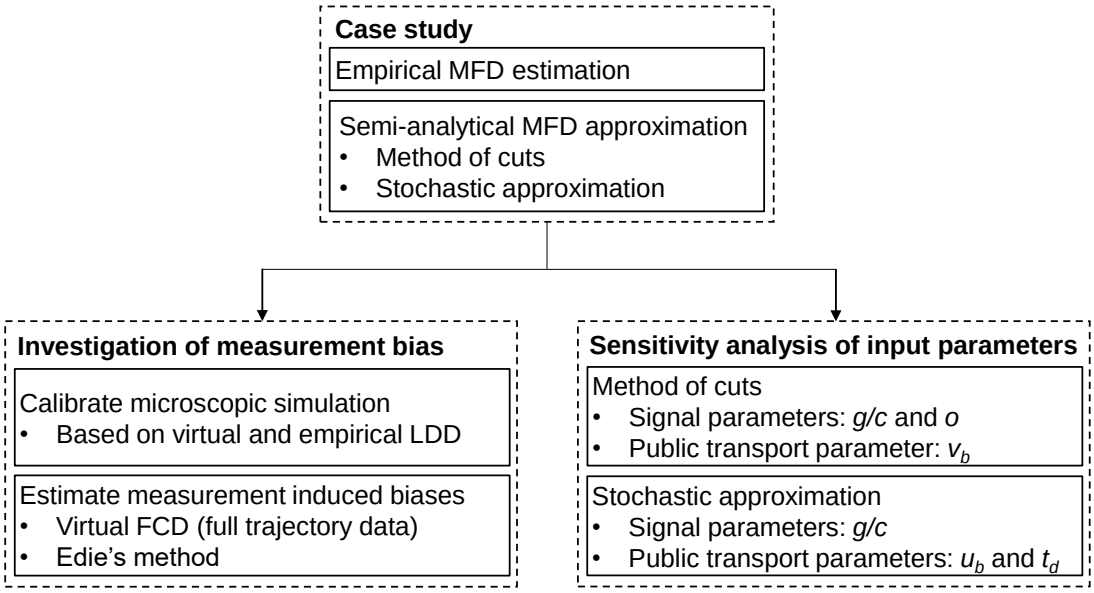


Figure 3.1: General methodological framework for the analysis of the MC and the SA.

The resulting bias-free MFD enables us to analyze the nature of the measurement bias for the presented case.

Last, we investigate the role of signal data and public transport operation on the approximated MFD. The semi-analytical methods and related input parameters imply certain simplifications of reality. The explicitly studied parameters are explained below. The study is conducted as a sensitivity analysis based on a Monte Carlo Simulation for both semi-analytical approximation methods. This study allows investigating the role of input parameters and analyzing the corresponding results in comparison to the empirical MFD.

The remaining section describes the layout of the case study, the application of the MC and SA in this context, the empirical MFD estimation, the investigation of the measurement bias, and the structure of the sensitivity analysis.

3.3.1 Case study

The Leopoldstraße leads from the center to the very north of Munich and serves as a critical corridor that connects the city center to the city's ring road. The empirical data containing raw LDD and signal data were collected from a segment of the northbound Leopoldstraße on a typical working day in October 2017. Comparable results for the time series of occupancies are found for other weekdays at the Leopoldstraße. Additionally, the distribution of green times g and cycle lengths c in a second available data set does not substantially differ from the one shown in this chapter. Thus, we assume that analyzing the data presented here is representative for the given road segment. Figure 3.2 shows the layout of the studied road segment with a total length of 1.1 km. It includes five signalized intersections, labeled from one to five. The studied segment mostly consists of two lanes except for some of the

intersections where a dedicated turning lane exists for right or left-turning vehicles. The loop detectors on the turning lanes are stop line detectors, which cannot be directly used to derive the MFD, and therefore, are not considered in this study. As it can be seen in Figure 3.2, there are some minor unsignalized intersections along the corridor from which no data, i.e. number of turning and incoming vehicles, were available. For the sake of simplicity, the impact of these intersections is neglected.

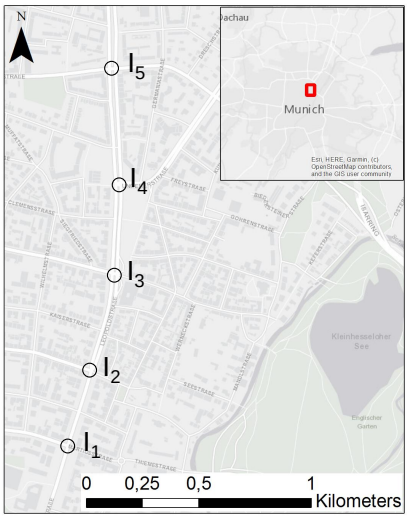


Figure 3.2: The urban corridor Leopoldstraße in Munich, Germany. The variables I_1 to I_5 refer to the examined intersections and the index increases in direction of travel.

Each of the five intersections is controlled by actuated traffic signals, i.e. variable green time and cycle lengths as well as prioritized public transit vehicles. Right at the north-east corner of the intersection I_3 an important mobility hub *Münchner Freiheit* is located that is a transfer point for public transport users between bus, tram, metro, and car-sharing vehicles. Moreover, another on-street bus stop is located between intersection I_1 and I_2. Two bus lines operate on the corridor, each with a 10-minute headway during most times of the day. Additionally, a tram line operates from intersection I_3 to I_5. However, the tram tracks are physically separated from the other lanes and thus do not interact with traffic flow on the neighboring lanes.

3.3.2 Estimation of the macroscopic fundamental diagram

An empirical MFD is estimated as described below. Furthermore, the MC and SA are applied on the described segment of the Leopoldstraße. The quantitative comparison is made based on the root mean square error (RMSE), which was successfully applied in the terms of MFD-related research [e.g. AMBÜHL ET AL., 2018].

Empirical estimation

The estimation of the empirical MFD for the case study is based on data recorded from loop detectors. In total, data from 10 loop detectors for the northbound travel direction were evaluated. All considered loop detectors are installed approximately 35 m upstream of the stop line. The average length of a loop detector is 1.5 m. Each observation consists of a time-stamp and an occupancy duration. A set of plausibility tests such as searching for duplicate detections, negative headways, and extreme occupancy values is performed. Subsequently, the data are aggregated in 5-minute intervals. The result is a vehicle count and an occupancy value per detector and interval. In order to reduce noise, outliers, and errors occurring due to aggregation, we smooth the data using time series analysis as suggested in AMBÜHL ET AL. [2018]. Then, the obtained occupancies are used to calculate densities assuming an average vehicle length of 5 m. The density is calculated based on the sum of the mean vehicle and detector length [see AMBÜHL ET AL., 2018].

The available signal data includes the time-stamp for each phase change and the phase type (green, amber, or red) for all signals. Since the traffic signals are not fixed-time controlled, the cycle lengths c, green times g, and red times are strongly skewed. Figure 3.3 shows box-and-whisker diagrams for green times g and cycle lengths c of the northbound leg at each intersection for the measurement day. The y-axis shows the duration of the corresponding parameter in seconds. The intersection labels are displayed on the x-axis. As it is shown in

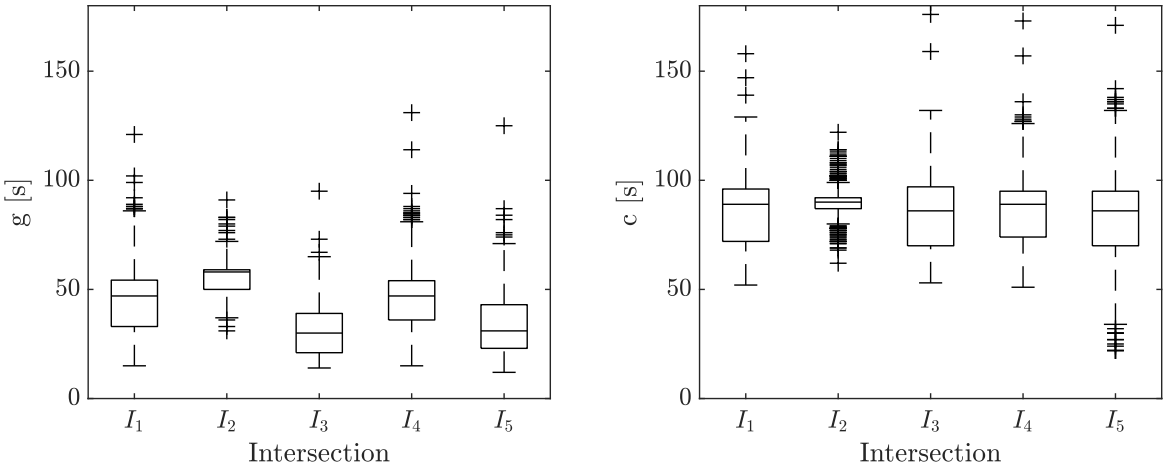


Figure 3.3: Green times g and cycle lengths c for the studied intersections at Leopoldstraße (October, 2017).

the plots, a substantial variance in both parameters can be observed. The reasons for this variance are the actuated traffic signal control, the transit signal priority, and the fact that the green times and cycle lengths of the entire day, and thus different signal plans, are included in the illustration. Additionally, it can be observed that the scatter is higher for longer green times. This is possibly due to the required minimum green times at the studied intersection legs. On the contrary, the distribution of cycle lengths seems to be symmetric. Moreover, the majority of green times of intersection I_3 and I_5 are substantially lower than the ones at

the other intersections. Regarding intersection I_3, this can be explained by the fact that it is located next to the *Münchner Freiheit* mobility hub, where many bus and tram lines stop. Thus, the effect of transit signal priority is very strong. Possible reasons for a lower mean green time at intersection I_5 are a tram and bus line that cross the intersection. Both are prioritized and thus green times are reduced for private vehicles.

Furthermore, limited historical FCD are available. The data are too scarce to benefit from fusing it with LDD for the empirical MFD estimation. However, it can be used to estimate the free-flow speed u as described below.

Method of cuts

The first semi-analytical approximation for the MFD which is evaluated is the MC. As described in Section 2.3, a global variational graph needs to be constructed to apply this method. For this purpose, the link FD, the corridor topology, and the signal timings are required. XIE ET AL. [2013] further extended the MC to approximate the effects of public transport vehicles on the capacity of an arterial. The additional input parameters consist of the average bus speed v_b and the headways. The effects of buses are modeled based on the numerical representation of moving bottlenecks. A moving bottleneck is approximated by reducing the capacity of a certain link. That is, the costs of the corresponding green phase edges at the upstream and downstream intersection are adjusted. Thus, the effects of public transport can be included in the variational graph and the remaining procedure to derive cuts does not need to be adjusted.

The link FD plays an important role for the MC. To define it, three out of the four related parameters, the free-flow speed u, the capacity q_{opt}, the backward wave speed w, and the jam density κ_{max} have to be specified. It has a crucial impact on the approximation of the MFD. For the presented case, the free-flow speed is chosen according to FCD which were available for the entire analyzed segment of the Leopoldstraße, and equals $u = 45$ km/h. The jam density is assumed to be $\kappa_{max} = 150$ veh/km. This is based on visual inspection of a number of queues captured by high-resolution aerial pictures. The backward wave speed is assumed as $w = 15$ km/h. LECLERCQ [2005] reported similar values for urban roads in Toulouse, France. In their study, however, the jam density is found to be substantially higher. Thus, we choose a slightly larger value for w to calculate realistic values for the link capacity q_{opt}. Given these three parameters and assuming a triangular FD, the link capacity is calculated as ca. $q_{opt} = 1690$ veh/h.

The block lengths are measured from OpenStreetMap® [OPENSTREETMAP CONTRIBUTORS, 2019] and are defined as the road segment length between two consecutive signalized intersections. The effects of the minor intersections along the corridor are neglected for this study. This is due to the fact that the semi-analytical approximation methods are not capable of considering the impact of such intersections, i.e. they cannot account for inflows or outflows within the modeled corridors. Also, the corresponding streets are of a minor hierarchical level and partly only one-lane streets. Furthermore, no flow or occupancy data from these streets are available. As shown in Figure 3.3, the green times and the cycle

lengths of all intersections are strongly skewed. However, the MC allows only for one value per intersection. In other words, only fixed-time signal control can be modeled. Thus, we choose the average for both green times and cycle lengths for the case study. For the sensitivity study, a different approach is chosen as described in a subsection further below.

The last set of input parameters with regards to public transport operation. The multi-modal extension of the MC proposed by XIE ET AL. [2013] only requires an exogenous average bus speed v_b and headways as input parameters. As no corresponding data is available, we make an assumption based on a study of free-flow speeds of different vehicle classes for Great Britain [DEPARTMENT FOR TRANSPORT, 2015]. There, a ratio of car and bus free-flow speeds in urban areas with a speed limit of about 50 km/h is reported as 90 %. Taking into account that there exists one on-street bus stop in the studied segment of the Leopoldstraße, and assuming a dwell time of 30 s, a bus speed v_b of 75 % of the car free-flow speed u seems to be appropriate. This results in an absolute bus speed of $v_b = 34$ km/h. The impact of this assumption is investigated as part of the sensitivity study. The dwell time of 30 s on average are chosen according to limited data from other bus lines recorded with GPS trackers.

Based on these input parameters a variational graph can be constructed. It allows deriving cuts to estimate an upper bound for the corridor MFD. Since the Leopoldstraße represents a highly inhomogeneous corridor, an appropriate calculation time window is chosen to generate a sufficiently large solution space [LECLERCQ & GEROLIMINIS, 2013]. Also, all possible moving observers are taken into account. For cuts with the same speeds, the ones with the lowest costs are chosen. The large size of the calculation time window ensures that all possible paths, and thus all constraining cuts are found. Finally, taking the lower envelope of all cuts leads to an estimate of the upper bound for the MFD of the given corridor.

Stochastic Approximation

The SA is the second method being evaluated. Again, the input parameters are derived based on the road topology, signal data, and available public transport data.

The FD is not necessarily required to apply the SA. However, it is indeed needed to compare the resulting MFD to the empirical and the MC-based one. The same values for the free-flow speed u, the backward wave speed w, the jam density κ_{max}, and consequently the capacity q_{opt} are applied.

Furthermore, the block length is an input for the SA as well. The specific input parameter is the mean dimensionless block length $\lambda_B = 6.8$. The number of blocks in the studied road segment is $B = 5$. Theoretically, the SA allows considering distributed green times and cycle lengths. The related input parameters are the average green time $\mu_g = 42$ s, and the long-run red to green ratio $\rho = 1.2$. Averaging the coefficient of variation of the green times, the red times, and the block length results in an overall average coefficient of variation equals $\delta = 0.21$. No offset parameter is necessary as input for the SA.

The public transport extension of the SA was described in CASTRILLON & LAVAL [2017] for homogeneous corridors. In this case, homogeneity refers to equal signal parameters and

block lengths amongst the modeled intersections. The authors modified the SA to integrate these impacts and introduced three new parameters, the average bus free-flow speed u_b, the stop probability p_s, and the average bus dwell time t_d. CASTRILLON [2015] presented an extension of this method for heterogeneous corridors. This is done by adjusting model parameters to account for the stationary bottleneck effect of buses as well as for their role as a moving bottleneck.

In the contrary to the MC, the bus free-flow speed is not assumed to account for potential stops. Such stops are modeled explicitly via the stop probability p_s. Since there is an on-street bus stop only on one of the five segments, we set $p_s = 1/5$. The average free-flow bus speed u_b is chosen as 90 % of the car free-flow speed, see the explanation in the section above. Again, the average dwell time is set to $t_d = 30\,\text{s}$.

Based on these parameters the mean and coefficient of variation for the cuts related to three different strategies can be calculated (see Section 2.3.3). Assuming normally distributed cuts, the cumulative distribution function for each cut is defined. This enables to compute the cumulative distribution function and thus the percentiles of interest for the MFD for the herein studied corridor.

3.3.3 Investigation of the measurement bias

A measurement bias inevitably exists in LDD since the data are recorded at fixed positions. Certain procedures to minimize this bias were described in Section 3.2. For example, measurements from several fixed sensors on different positions could be projected to a virtual link. These projected and then accordingly weighted measurements could lead to a reduced bias in the data. However, this is infeasible for the case studied in this chapter since all loop detectors are positioned at nearly the same distance from the stop line. Another way to exclude measurement-induced biases is to gain complete knowledge about the traffic conditions, i.e. about all trajectories. Since this is not possible for the empirical data of the Leopoldstraße, a microscopic simulation environment is built in SUMO for this purpose. The layout of the simulation assembles the real corridor. In total, 129 OD pairs are defined and the route choice is designed based on recorded turning rates at each intersection. Additionally, two public transport bus lines each with a headway of 10 minutes are represented in the simulation. The estimation of OD pairs and loading curves is based on the measured travel times. More details regarding the simulation are described in GRIGOROPOULOS, KELER, KATHS, KATHS, SPANGLER, HOFFMANN & BUSCH [2018].

In order to investigate the measurement bias, a calibrated microscopic simulation is needed. This is achieved by adapting the Krauss car-following model [KRAUSS, 1998] parameters with the objective of minimizing the difference between the two MFDs from real loop detectors and virtual detectors in the simulation model. For this purpose, flow and density measurements are recorded at positions that correspond to the loop detector positions at the real Leopoldstraße. We apply a locally estimated scatter plot smoothing (LOESS) regression [CLEVELAND, 1979] based on these virtual loop detector measurements. Additionally, a LOESS regression is applied to the empirical data. The comparison of both curves is the

base for the calibration procedure which is designed as an RMSE minimization. This would not be possible without using regression techniques. Secondly, the calibrated simulation is further used to estimate the MFD based on full trajectory data following Edie's definitions [EDIE, 1963]. By doing so, the impact of measurement-induced biases on the MFD for the Leopoldstraße can be evaluated.

3.3.4 Sensitivity analysis of signal and public transport parameters

We conduct a sensitivity analysis to study the impact of important input parameters of the semi-analytical approximations on the resulting MFDs. The MC does not allow for implementing varying green phases and cycle lengths within an intersection. The SA assumes green phases and cycle lengths which are independent and identically distributed. Both assumptions are violated by using the signal data at hand. Furthermore, the Leopoldstraße is frequented by buses of the public transport operator. To estimate the impacts of multi-modal interactions on the MFD by applying the approximation methods, additional data on public transport operation are required. However, no reliable data for the specific input parameters are currently available. Consequently, realistic assumptions were made.

We conduct a Monte Carlo Simulation to study the violation of assumptions concerning the signal-related input parameters, as well as to analyze the assumption regarding public transport-related input parameters. In order to replicate the real-world scenarios in the variational graph, all possible input parameter combinations have to be considered. However, this is infeasible due to the high computational effort. To overcome this challenge, we apply a Latin Hypercube sampling to generate a representative range of values for each input parameter. Several scenarios are studied. In each scenario, a different number and type of input parameter are varied, which are related to signal control and public transport operation. The signal control-related input parameters are drawn from the distribution of measured values. For the MC, this includes the green time and cycle length ratio g/c, as well as the offset o. For the SA, the parameters considered for sampling are the average green time μ_g, the long-run red to green ratio ρ, and the corresponding standard deviations. In order to derive corresponding mean and standard deviation values, small samples of five consecutive cycles are considered. This sample size is chosen as a trade-off between having a sufficiently large number of mean and standard deviation values for the Monte Carlo Simulation and avoiding extreme values for each mean and standard deviation. For the sake of simplicity, these parameters are also labeled g/c hereafter, as they essentially refer to green times and cycle lengths. The public transport-related input parameters consist of the bus frequency, the average speed v_b of buses for the MC, as well as the average bus free-flow speed u_b and dwell times t_d for the SA. However, no reliable data for these specific input parameters are available. Thus, the bus frequency is assumed to follow the schedule and is therefore not further investigated within the sensitivity study. Moreover, v_b and u_b are varied between 50 % and 100 % of the car free-flow speed u. In addition, t_d is varied between 10 s and 60 s, which are set based on a limited number of on-site measurements.

For each input parameter, a sample with a size of $n = 10000$ is drawn. Based on a set of input parameters, the MFDs can be approximated for each scenario. The effects of

varying input parameters are studied for both semi-analytical methods, the MC and the SA. Table 3.1 gives an overview of the conducted scenarios by showing the investigated parameter combinations. As mentioned above, g/c represents the green phase and cycle length parameters for both the MC and the SA. The input parameters o and v_b are only required by the MC. The average bus free-flow speed u_b and average dwell time t_d are only required by the SA.

Table 3.1: Input parameter configuration for the sensitivity study

	Method of cuts							Stochastic approximation						
Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14
g/c	✓			✓	✓		✓	✓			✓	✓		✓
o		✓		✓		✓	✓							
v_b			✓		✓	✓	✓							
u_b									✓		✓		✓	✓
t_d										✓		✓	✓	✓

3.4 Results and discussion

This section shows and discusses the results for the case study, the investigation of the measurement bias, and the sensitivity analysis. Many results are presented as MFDs, such as in Figure 3.4a, and Figure 3.5a – 3.8. If not stated otherwise, the x-axis shows the corridor-average density K in veh/km/ln and the y-axis shows the corridor-average flow Q in veh/h/ln.

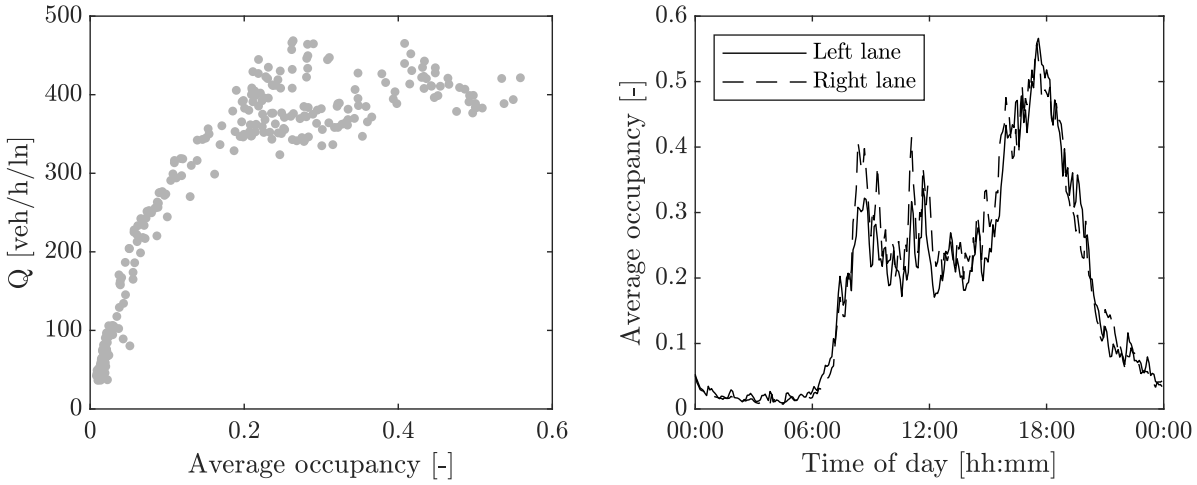
3.4.1 Case study

The results for the case study include the empirical MFD and the MFDs approximated by the MC and the SA, respectively.

Empirical MFD

The empirical MFD is estimated based on the methods explained and the data described in the previous section. The result is shown in Figure 3.4a. The x-axis shows the average dimensionless occupancy per lane ranging from 0 to 1. The figure shows a well-defined MFD with a clear shape. Especially, the data related to free-flow states are only a little scattered. Maximum flows are observed around 480 veh/h/ln. These observations seem to be plausible as they are similar to MFDs reported for other cities [e.g. AMBÜHL ET AL., 2017]. The occupancy reaches values up to 0.6. At occupancies between 0.2 to 0.4, a clockwise hysteresis loop can be observed. Its investigation reveals that the upper part of the loop occurred during the loading period, while the lower part occurred during the unloading

period in the evening. These periods correspond to the onset and offset of congestion. This pattern can be explained as follows. The direction of travel points towards the outer parts of the city, including the city's ring road. In other words, commuters are traveling along the Leopoldstraße to leave the city. This explains the time of the onset and offset of congestion. Similar patterns were observed in other studies [SABERI & MAHMASSANI, 2013; BUISSON & LADIER, 2009]. More details on this phenomenon are presented in Chapter 1.



(a) Empirically estimated MFD. **(b)** Average lane-specific occupancies.

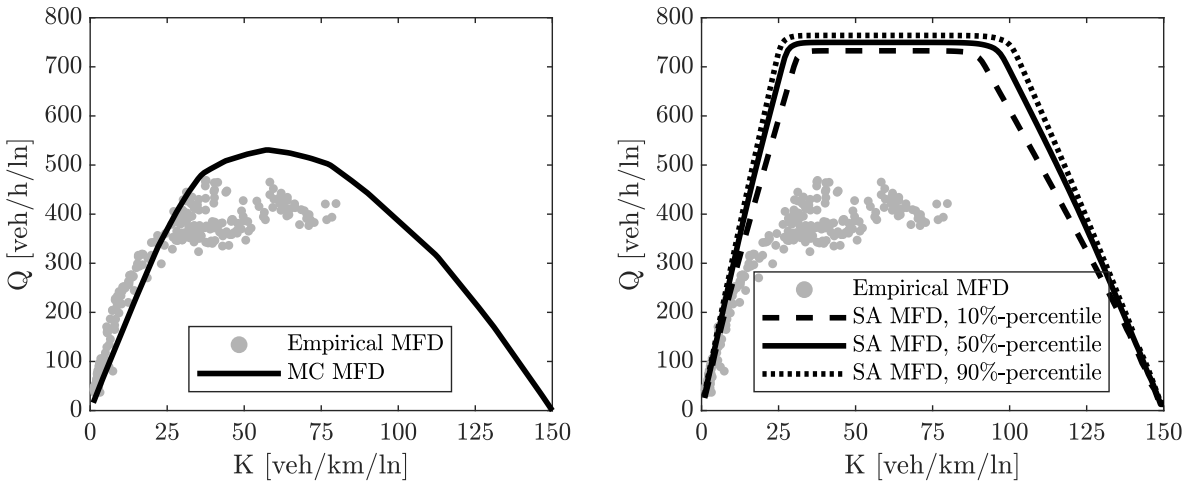
Figure 3.4: Empirical results for the Leopoldstraße on October 17, 2017.

The capacity of the corridor could be reduced due to constraining effects on the supply side (such as signals), but also due to demand-related matters. For example, if road users strongly favor one lane, e.g. due to turning possibilities, the other lane's capacity might be underutilized and thus the MFD would not show the effective capacity for the whole corridor segment. To investigate this issue, we plot the lane-specific average of occupancies for the analyzed corridor. The results are displayed in Figure 3.4b. The x-axis and the y-axis show the time of day and the dimensionless occupancy per lane, respectively. The corresponding values for the left lane in direction of travel are shown as a solid black curve, and as a dashed black line for the right lane. The figure demonstrates that there are only minor differences between the two lanes. In general, occupancy values for the right lane are slightly higher than the ones for the left lane. Reasons for that might be the potentially higher frequency of flow-disturbing elements such as buses leaving a bus bay, cars looking for parking spots, and logistic vehicles loading and unloading goods. However, the differences are minor and thus we conclude that the capacity of both lanes is utilized to a similar degree. Another reason for low flows apart from bottlenecks within the studied road segment can be constraining conditions upstream and downstream of the corridor. We suppose there are no upstream constraining conditions since we observe a congested branch in the empirical MFD. If there would be any constraining conditions upstream, the observed congested branch would be rather limited. This is not the case, as can be seen in Figure 3.4a. Additionally, traffic flow could be constrained by spillbacks from downstream intersections. Unfortunately, the available data does not include measurements from such

locations. However, qualitative on-site evaluation from several days suggests that spillbacks are rare and do not systematically impact the capacity of the most downstream intersection of the studied segment of the Leopoldstraße. Comparable time series of occupancies are found for other weekdays. Additionally, the distribution of green times and cycle lengths in a second available data set does not substantially differ from the one shown in this chapter. Thus, we assume that analyzing the presented data is indeed representative for the given road segment. In summary, we conjecture that the observed capacity in the MFD is due to bottlenecks within the investigated system and not due to underutilized lanes or constraining impacts from up- or downstream.

Semi-analytical approximations

The evaluation of the semi-analytical approximations for a given set of input parameters leads to an estimate for the upper bound of the MFD, shown in Figure 3.5. The curves show the MFDs estimated based on the MC and the SA. For comparison, we display the scatter plot representing the empirically derived MFD with occupancies converted to densities as described in Section 3.3. The general trapezoidal shape results from the fact that the MC and SA approximate the MFD based on different families of cuts, as explained in Chapter 2.



(a) The MFD based on the MC. **(b)** The MFD based on the SA.

Figure 3.5: Comparison of semi-analytically approximated and empirically derived MFDs.

A quantitative comparison of the semi-analytical MFD based on the MC (see Figure 3.5a) and the empirical MFD leads to the following observations. The semi-analytically derived free-flow cut matches the data rather well, even though the flows are in general underestimated for densities ranging up to 25 veh/km/ln. For densities greater than that, flows are overestimated by the MC. Moreover, the maximum flow is reached around 470 veh/h/ln in reality. The estimated capacity of the MC is around 530 veh/h/ln. Therefore, the MC overestimates the capacity of the corridor by roughly 13 %. Possible reasons for such differences between the empirically estimated and the semi-analytically approximated MFD are the assumptions

related to the KWT such as instantaneous acceleration and stationary traffic flow. Short green times and cycle lengths could lead to an enhancement of the effects of violating these assumptions. Furthermore, taking the average of green times and cycle lengths as input to the MC might decrease its accuracy. Moreover, the MC does not allow for consideration of signal actuation or transit signal priority, which are in operation in reality and affect traffic flow dynamics. Recall that the MC represents the idealized MFD, while the empirical corresponds to the realized one. Thus, the observed differences could also occur due to turning flows. Last, the empirical MFD is based on LDD and thus includes a measurement bias. This bias is analyzed within this chapter as well, and the respective results are presented in the next section. All these points potentially contribute to the observed discrepancies in the empirically estimated and semi-analytically approximated MFD.

Note that the SA computes a cumulative distribution function for the MFD, i.e. a distribution of flows Q for a given density K. The results of the evaluation of the SA for the 10 %-, the 50 %-, and the 90 %-percentile are shown in Figure 3.5b as dashed, solid, and dotted curves, respectively. Additionally, the figure shows the empirical data as grey scatter for comparison. Generally, the free-flow branch of the semi-analytical approximations fits the empirical data well up to a flow of about 250 veh/h/ln. For higher flows, differences arise. The largest discrepancy occurs for the capacity which is about 300 veh/h/ln, or about 63 % in relative terms. This discrepancy clearly exceeds the one observed for the MC. Moreover, the difference between the percentiles of the semi-analytical approximations is minor and cannot represent the scatter in the empirical data. The largest differences between them occur close to capacity and for the congested branch. We suppose that there exist several possible reasons for the large differences in the estimated and observed capacities. The actuated traffic signal control and transit signal priority violate the assumption of an independent distribution of green and red times. Moreover, there exist substantial differences in green-time-to-cycle-length ratios between intersections as can be concluded from Figure 3.3. More specifically, intersection I_3 has the lowest g/c ratio and, thus, is determining the overall capacity of the corridor. However, the SA derives the capacity cut based on the average of capacities related to each intersection. In the present case of a corridor with substantial differences of capacities within the intersections, this might lead to a wrong estimation of the overall capacity. Similar to above, the fact that the SA estimates an idealized MFD instead of a realized MFD can contribute to the observed differences. Last, it is again noteworthy that the empirical data includes a measurement bias that impacts the general fit. All in all, these aspects could lead to an overestimation of the capacity based on stationary cuts.

To further quantify the comparison of the semi-analytical methods to the empirical MFD, the corresponding RMSE values are computed. This results in an RMSE of 83.0 for the MC, and one of 271.2 for the SA. These values indicate that the overall approximation based on the MC is superior to the one based on the SA for the input parameter configuration of the case study. As mentioned above, probable reasons are related to the measurement bias included in empirical data, the representation of signal control and public transport operation in the models, as well as assumptions inherited from the KWT. The results of a thorough analysis of these aspects are investigated in the next subsections.

3.4.2 Investigation of the measurement bias

Since only limited empirical data are available, the associated measurement bias can only be estimated based on a microscopic simulation, as described in Section 3.3. Figure 3.6 shows the results of this investigation. Three different MFDs are presented. First, the MFD based on empirical loop detector data (ELDD) as presented in Figure 3.4a is displayed again as grey scatter plot for comparison. Additionally, the curve resulting from the LOESS regression based on this data is shown as a black solid line. Second, the curve resulting from a LOESS regression based on virtual loop detector data (VLDD) is shown as a black dotted line. The curve suggests flows $q > 0$ veh/h/ln for densities $\kappa = 0$ veh/km/ln. Note that this is a result of the regression method, and is physically impossible. Last, the LOESS regression curve based on trajectories from the simulation, i.e. virtual floating car data (VFCD), is presented as a black dashed line.

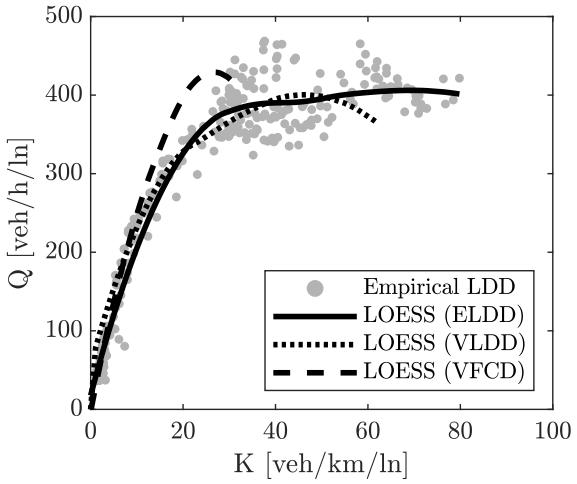


Figure 3.6: Illustration of the measurement bias estimation.

The regression results based on ELDD and VLDD were used for calibration. The RMSE of the LOESS (ELDD) and LOESS (VLDD) is 22 veh/h/ln. Considering the impact of day-to-day variance in traffic flow, measurement errors, and other factors, we adjudge this value sufficiently low to claim the microscopic simulation being calibrated well enough for our purposes.

Since the trajectories contain all information, the curve based on virtual trajectories represents an estimate for an MFD without the measurement bias. It can be seen that the free-flow branch is slightly shifted towards higher speeds and the capacity is increased by roughly 40 veh/h/ln. This means that the measurement bias could lead to an underestimation of the actual measured capacity, and thus the error of the semi-analytical approximation could be smaller as observed by comparison with the empirical LDD based MFD. Furthermore, the results indicate that the average free-flow speed in the network is underestimated by simply relying on empirical data.

3.4.3 Sensitivity analysis of signal and public transport parameters

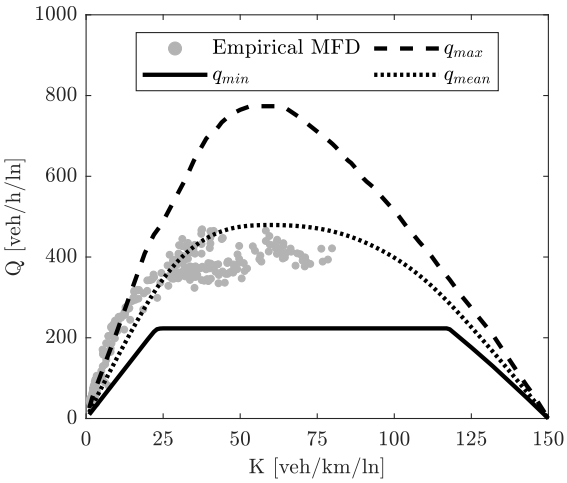
As discussed based on the results from the case study, the representation of signal control schemes and public transport operation in the semi-analytical approximations are potential reasons for the differences of the corresponding MFDs and the empirical MFD. In order to study the impact of the related input parameters, a Monte Carlo Simulation with $n = 10000$ draws is conducted as described in Section 3.3. Thus, for each density value, a range of 10000 flow values is drawn. Based on these results, a corresponding minimum q_{min}, mean q_{mean}, and maximum flow q_{max} can be derived. For the whole range of densities, these flows can be plotted as curves. Each curve is compared to the empirical MFD and an RMSE is calculated. The following subsections show and discuss the results for both investigated semi-analytical approximations. The data plotted in Figure 3.7 and Figure 3.8 show the empirical MFD as scatter plot, and q_{min}, q_{mean}, and q_{max} as dashed, solid, and dotted curve, respectively.

Method of cuts

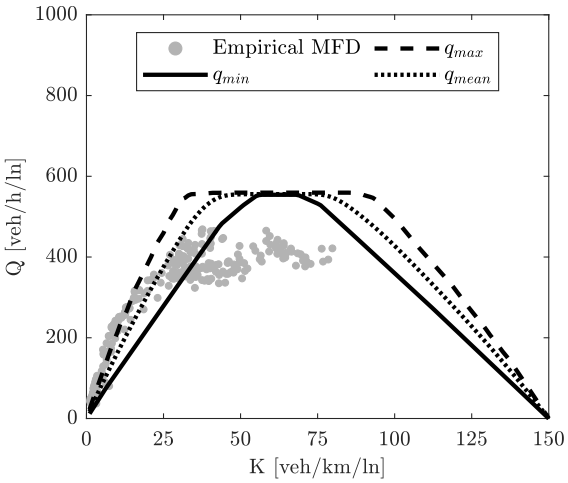
Three input parameters of the MC are related to signal control and public transport operation. This is the green time and cycle length which is represented by g/c, the offsets o, and the average bus speed v_b. To investigate all parameters and the possible combinations of them, seven scenarios are studied (see Section 3.3.4). Presenting all plots within this chapter is not expedient since the generally observed patterns can be explained with the three diagrams shown in Figure 3.7.

As expected, varying the parameter g/c related to cycle length and green times has a large impact on the resulting MFDs (see Figure 3.7a). The capacity ranges between ca. 200 veh/h/ln and 800 veh/h/ln. Additionally, the free-flow and congested branches of the MFD are varying strongly. Interestingly, the mean flows q_{mean} approximate the empirical MFD well. Figure 3.7b shows the results of the sensitivity analysis of the input parameter related to the offset o. As expected, the impact on capacity is negligible. However, the free-flow and congested branches are highly impacted by the offset. In general, the offset is challenging to specify, since it is adaptive in reality and this cannot be reproduced by the MC. Lastly, Figure 3.7c presents the results for the scenario related to the average bus speed v_b. It can be seen that the impact is minor. In other words, for the current representation of public transport in the MC, this parameter plays only a minor role for the given case. Thus, assuming a corresponding value instead of measuring it is not crucially affecting the estimated MFD.

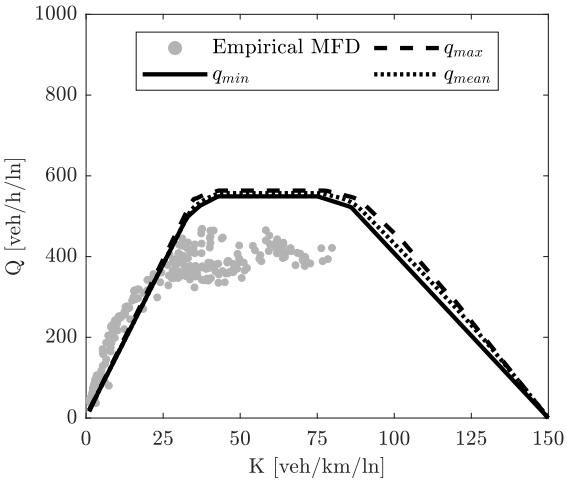
The figures for scenarios 4 to 7 are not presented here but included in the RMSE-based analysis below. Generally, visual inspection shows that the parameter g/c has the main influence also when it is combined with the other studied input parameters. The combination of o and v_b reveals that their impact on the MFD is still minor in comparison to g/c. Thus, the resulting MFDs have a shape being similar to the ones shown in Figure 3.7. For the sake of completeness, the numerical results in the form of RMSE values for all scenarios are presented in Table 3.2.



(a) Scenario 1: Sensitivity to g/c.



(b) Scenario 2: Sensitivity to o.



(c) Scenario 3: Sensitivity to v_b.

Figure 3.7: Results of the sensitivity analysis of the MC.

These results reveal further insights. First, it can be seen that the calculated RMSEs for q_{mean} are the smallest except for scenario 2 and 6. q_{mean} for scenarios 1, 4, 5, and 7 have a lower RMSE than the one calculated for the MFD of the case study. Thus, using q_{mean} would lead to an improved estimation of the empirical MFD for these scenarios. Another interesting result is that the RMSE of q_{mean} differs from the one related to the case study. In the case study, the MC was applied with the average of each input parameter. Comparing the RMSE of the MFD resulting from the case study to the RMSE of q_{mean} resulting from the sensitivity study reveals a difference. That is that the average input values do not lead to the mean output values. This finding confirms the purposefulness of the sensitivity study. Moreover, the range between the RMSE of q_{min} and q_{max} differs greatly. It is especially large when the parameter g/c is varied, such as in scenarios 1, 4, 5, and 7. This lets us conclude that the impact of g/c is the largest on the semi-analytical approximation by the MC. This is in line

Table 3.2: RMSE results for the sensitivity study of the MC

Scenario	RMSE		
	q_{min}	q_{mean}	q_{max}
1	138.1	59.2	224.8
2	88.1	104.3	123.8
3	106.5	111.2	117.7
4	141.1	59.3	228.6
5	138.8	58.5	225.0
6	83.7	104.8	127.7
7	140.8	58.7	230.1

with what was concluded from Figure 3.7a. Last, the low impact of the average bus speed v_b is evident when the corresponding RMSE values are studied for scenario 3. Additionally, the values for scenario 5 are similar to the ones of scenario 1, so the interactive effect of v_b with g/c is low. The same applies to scenarios 2 and 6, where the interactive effect of v_b with o is examined.

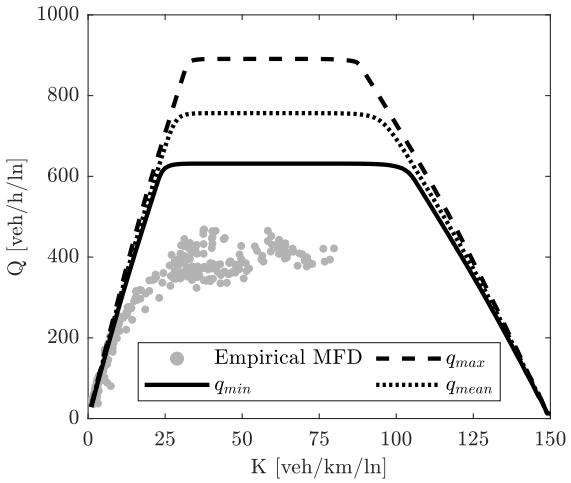
In summary, it can be observed that the approximated capacity is highly sensitive to g/c. This is an expected result. However, averaging the flows for each density based on the computed MFD approximations reveals a well-fitting curve to the empirical MFD. This is interesting since possible g/c values can be estimated based on signal control programs and thus an ex-ante prediction of the corridor capacity is feasible, even though actuated traffic signal control cannot be evaluated explicitly within the studied methods. The high sensitivity to g/c is due to its direct relation to the stationary cut which affects the capacity for the present case. The choice of offsets o has a substantial impact on the free-flow and congested branch but has no effect on the capacity of the resulting MFDs. The reasons are that varying offsets o do not affect the g/c ratio and thus not the capacity, as long as the stationary cuts are constraining. However, different offsets change the variational graph in such a way that the impact on shortest paths within the framework of the MC leads to differing free-flow and backward wave branches in the resulting MFDs. The fit of these branches to empirical data or potentially bias-free data is limited. However, this is not surprising since only fixed-time control can be implemented into the MC, whereas in reality traffic signal actuation and transit signal priority are in operation. Additionally, it is clearly shown that the public transport operation, as currently modeled, has nearly no effect on the resulting MFD. This can be due to the low impact on the variational graph for the given headway and average bus speed.

Stochastic approximation

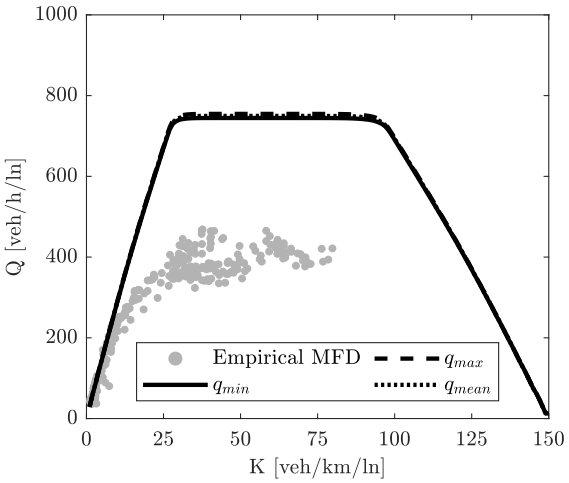
Similar to the MC, three input parameters of the SA are related to signal control and public transport operation. Again, g/c reflects the green time and cycle length, the dwell times of buses t_d, and the average bus free-flow speed u_b. Seven different scenarios are studied to investigate all parameters and their possible combinations as described in Section 3.3.4. Since the SA calculates a cumulative distribution function of the MFD, a percentile has to be

specified to derive a single MFD for each run within the sensitivity analysis. We select the 50 %-percentile as the most representative.

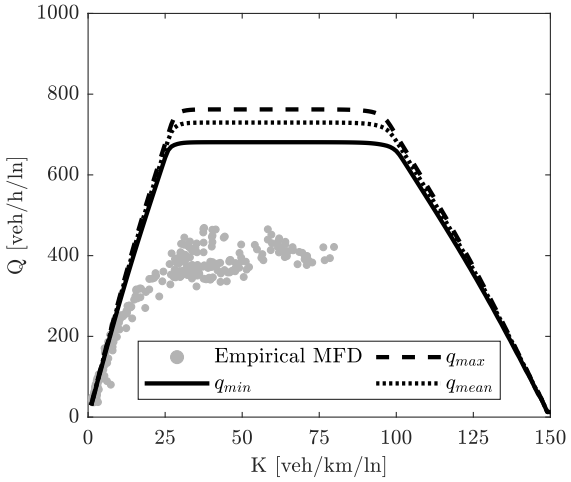
Visual inspection of all plots reveals again that it is sufficient to present only the diagrams of the first three scenarios in Figure 3.8. The main results can be explained through these figures.



(a) Scenario 8: Sensitivity to g/c.



(b) Scenario 9: Sensitivity to t_d.



(c) Scenario 10: Sensitivity to u_b.

Figure 3.8: Results of the sensitivity analysis of the SA.

Figure 3.8a shows that the parameter g/c related to cycle length and green times has again a large impact on the resulting MFDs. The capacity ranges roughly between 600 veh/h/ln and 900 veh/h/ln. Contrarily to the results of the MC analysis, the free-flow and congested branches of the MFD are barely affected. Figure 3.8b presents the results for varying the bus dwell times. It becomes clear that this input parameter has a negligible influence on the MFD for the investigated case. Thus, it seems that the choice of this parameter is not important for applying the SA for the case of the Leopoldstraße. Last, Figure 3.8c displays the results for

the scenario related to the average bus free-flow speed u_b. The parameter has a substantial influence on the shape of the MFD. More specifically, it can be seen that the capacity ranges between 650 and 750 veh/h/ln. For all figures, it is observed that the free-flow and congested branches are substantially less impacted by varying the input parameters compared to the MC. Since the free-flow cut matches the empirical data well, this is an advantage of the SA.

The figures for scenarios 10 to 14 are not shown. Visual inspection reveals that no unexpected interaction effects between the studied input parameters exist. The consideration of multiple variable input parameters results in adding their impacts. For example, varying t_d has still very limited effects when it is combined with varying g/c or u_b. The resulting MFDs are similar to the ones shown in Figure 3.8. For the sake of completeness, the numerical results in the form of RMSE for all scenarios are presented in Table 3.3.

Table 3.3: RMSE results for the sensitivity study of the SA

Scenario	RMSE		
	q_{min}	q_{mean}	q_{max}
8	267.5	271.2	274.8
9	187.2	275.7	367.5
10	221.6	256.8	280.2
11	204.3	256.6	293.7
12	148.9	261.0	369.7
13	174.6	275.5	372.0
14	133.7	260.9	369.7

These results reveal further insights. First, it can be seen that the calculated RMSE values for q_{min} are always the smallest. Additionally, all q_{min} have a lower RMSE than the one calculated for the MFD of the case study. Thus, using q_{min} would lead to an improved estimation of the empirical MFD for all scenarios. Another interesting result is that the RMSE of q_{mean} differs from the one related to the case study. Similar to the results for the MC, the mean input to the SA does not lead to the mean output. Moreover, the ranges between the RMSE of q_{min} and q_{max} differ greatly. They are especially large when the parameter g/c is varied, such as in scenarios 8, 11, 12, and 14. This lets us conclude that the impact of g/c is the largest on the semi-analytical approximation by the SA. This is in line with what was concluded from Figure 3.8a and with similar results for the MC. Last, the low impact of the bus dwell times t_d is evident when the corresponding RMSE values are studied for scenario 10. The values for scenario 12 are similar to the ones of scenario 10, so the interactive effect of u_b with g/c is low. The same applies to scenarios 10 and 13, where the interactive effect of u_b with t_d is examined.

Generally, the SA's MFD shows a good and robust fit to the free-flow branch of the empirical one. However, the capacity is clearly overestimated, even when considering the minimum of all computed MFDs in the corresponding scenarios. Reasons for such an overestimation are discussed in Section 3.4.1. Nevertheless, the method seems to capture public transport operation more effectively, since the related parameters have a substantial impact on the

resulting MFDs. This could be due to the endogenous consideration of effective bus speeds in the model [CASTRILLON & LAVAL, 2017]. Extending the SA to account for single intersections which dominate the overall capacity might be an interesting line of research. Such an extension could lead to a better fit to empirical data for cases similar to the Leopoldstraße, i.e. corridors with a comparably low number of intersections and high variability in signal timings.

This study aimed at investigating the differences between the empirically estimated and semi-analytically approximated MFDs. The results shed light on possible reasons, such as the measurement bias, model assumptions, and simplified considerations of signal control and public transport operation. The remaining unexplored reasons for discrepancies are likely the assumptions of the KWT and the hysteresis phenomenon. For example, the role of infinite acceleration and deceleration implied by KWT could be investigated. First analyses conducted by the authors showed that more effort is necessary to understand the effect of these parameters on the MFD. Moreover, considering only the mean of green-to-cycle ratios in the SA clearly leads to an inaccurate estimation of the stationary cut. This needs to be further investigated in the future. The hysteresis phenomenon can be considered with the trip-based MFD approach as shown in MARIOTTE ET AL. [2017]. However, both the MC and SA are methods to derive accumulation-based MFDs and thus such an investigation lies beyond the scope of this chapter. Overall, the conducted analysis helps to gain a better understanding of several potential error sources in the MFD approximation.

3.5 Conclusion

This chapter compares two different methods which semi-analytically approximate the corridor MFD. The first method is the MC which is deterministic and approximates the MFD based on the link FD, road topology, signal settings, and public transport operation parameters. The second method is the SA, which considers the stochasticity of input parameters and estimates a cumulative distribution function for the MFD. Within a case study, both methods are evaluated by comparing the approximated MFDs against the MFD derived from empirical data collected at five intersections of an urban corridor, namely the Leopoldstraße in Munich, Germany. The empirical data set consists mainly of LDD and time-stamps for signal phase changes. The case study reveals differences between the empirical and semi-analytical MFDs. Subsequently, the chapter investigates potential reasons for occurring discrepancies between the empirical estimation of the MFD and corresponding semi-analytical approximations. This includes the analysis of the measurement bias which is induced by LDD. The results suggest that the actual capacity could be larger than the empirical MFD proposes for the studied case. Moreover, the free-flow branch of the estimated MFD appears to be steeper without the bias. Eventually, an in-depth investigation of the input parameters regarding signal control and public transport operation for both methods and their impacts on the resulting MFD approximation is conducted. The analysis is performed as Monte Carlo Simulation based on a Latin Hypercube sampling. The results indicate that the capacity predicted by both methods is highly sensitive to the green-to-cycle ratio-related

input parameters. Interestingly, the mean flow resulting from calculating a large sample of MFDs based on the MC by varying the green-to-cycle ratio fits the empirical data very well. Moreover, the offsets have a strong impact on the free-flow branch of MFDs predicted by the MC, which is generally underestimated by this method. On the contrary, the SA overestimates the capacity in all cases. However, this method predicts the free-flow branch accurately. Input parameters related to public transport operation have a comparably low impact. Also, both models are only able to estimate the idealized MFD, even though the ground truth in form of empirical data represents the realized one. In theory, this problem could be addressed by taking the effects of spatial demand patterns such as turning ratios into account. Currently, this is not possible with the MC or the SA. Overall, the MC seems to be more promising for the case studied as indicated by the corresponding RMSE results. By knowing possible green-to-cycle ratios, a general shape of the MFD seems to be predictable for highly inhomogeneous urban corridors.

Future research regards the analysis of additional urban corridors to generalize the findings for the MC and the SA. Furthermore, extending the MC to be applicable to road networks and corridors with traffic actuated signal control would be an interesting task. This could lead to a fast and efficient way to analyze the capacity of a corridor with actuated signal control. Thus, it could be used to study different signal control strategies regarding maximum corridor capacity. However, it is unclear if the mathematical framework holds for hyperlinks that represent varying signal timings between intersections and at each intersection to represent more complex signal control schemes. Thus, completely new approaches might be necessary. Additionally, more accurate implementation of public transport can be incorporated. This might further improve the accuracy of the methods. Last, only the SA is conjectured to apply to networks. No deterministic method such as the MC is available to semi-analytically approximate the MFD for networks in general. In the past, the MC was only applied to simplified networks, which can be approximated by a single corridor. Therefore, corresponding developments could open the door to a semi-analytical approximation of the MFD.

Chapter 4

Network variational theory

This chapter is based on the following publication:

- G. TILG, L. AMBÜHL, S. F. BATISTA, M. MENENDEZ & F. BUSCH [2021b]: On the application of variational theory to urban networks. In: *Transportation Research Part B: Methodological* 150, pp. 435–456. DOI: 10.1016/j.trb.2021.06.019

4.1 Introduction

The literature related to (Eulerian) VT reveals its wide range of applications apart from pure traffic modeling even though it only applies to corridors. This includes multi-modal traffic analysis [DAKIC ET AL., 2020; GULER, GAYAH & MENENDEZ, 2016; WU, GULER & GAYAH, 2017; WU & GULER, 2018; GAYAH, ILGIN GULER & GU, 2016; SAEEDNIA & MENENDEZ, 2016], approximate MFD estimation [DAGANZO & GEROLIMINIS, 2008; LECLERCQ & GEROLIMINIS, 2013; LAVAL & CASTRILLÓN, 2015; TILG ET AL., 2020a], and traffic data imputation [MEHRAN, KUWAHARA & NAZNIN, 2012; SUN & BAN, 2013]. Additionally, there are extensions for consideration of higher-order traffic models [LEBACQUE & KHOSHYARAN, 2013; COSTESEQUE & LEBACQUE, 2014], stochasticity [WADA, USUI, TAKIGAWA & KUWAHARA, 2018; DAKIC ET AL., 2020], and continuous source terms [LAVAL ET AL., 2016]. The wide range of applications, the existing model extensions, as well as the ability to consider complex intra-link heterogeneities make the VT a valuable framework to numerically solve KWT problems. As a result, extending its applicability from single corridors to networks benefits all these applications, and allows us to evaluate the impact of complex intra-link heterogeneities at the network level with numerical precision and efficiency. To the best of our knowledge, none of the previous studies have started to explore the application of VT at the network level accounting for its complexities. To address this gap, we propose an extension of VT that permits to model traffic dynamics at the network level. The contributions of this chapter are fourfold:

1. We include inflows and outflows in the mathematical framework of VT. Contrary to LAVAL ET AL. [2016], these source terms accept any values from the set of real numbers, most importantly, they also include zero.

2. We model the propagation of spillbacks across the network. This allows us to track the evolution of network-wide congestion patterns.

3. We evaluate the proposed methodology and compare it with a microscopic KWT model implemented in SUMO [LOPEZ ET AL., 2018]. This enables us to show a proof of our concept and to demonstrate the inheritance of useful properties from the original VT, including the numerical precision and the ability to model complex bottlenecks.

4. We propose an algorithm to apply our model to any road network and showcase its applicability using the well-known example of the Sioux Falls network.

This chapter is organized as follows. Section 4.2 presents a brief background on limitations, applications, and methodological extensions of VT. The basic theory, as well as solution methods, were described in Chapter 2. Section 4.3 proposes a VT-based methodology to estimate traffic state propagation at the network level. This includes the consideration of inflows and outflows at intersections and the correct propagation of spillbacks. Section 4.4 provides a proof of concept. Section 4.5 demonstrates, with the aid of small networks that the proposed model inherits the original VT properties, which have proven to be rather useful for a wide range of applications. Section 4.6 presents the applicability of our methodology to larger networks. Finally, Section 4.7 highlights the conclusion of this study and outlines potential future research topics.

4.2 Limitations, applications, and extensions of variational theory

4.2.1 Limitations

While any type of bottleneck can be implemented, its trajectory needs to be fully mapped in G in order to be considered [DAGANZO & MENENDEZ, 2005]. For example, nodes in the variational graph G for single corridors and triangular FDs with non-integer θ do not align horizontally. Thus, additional manual efforts are required to include traffic signals. While this poses a limitation from a practical perspective, the general application of VT is still feasible for such cases.

Another limitation is that no continuous multi-modal, multi-commodity traffic streams can be modeled. However, hybrid approaches to model multi-modal aspects were successfully developed in previous studies where slower modes (e.g. trucks or buses) are represented as moving bottlenecks [DAKIC & MENENDEZ, 2018; SAEEDNIA & MENENDEZ, 2016].

The main limitation of VT so far is that it only applies to single corridors, i.e. where the conservation equation holds. To relax this limitation is the main purpose of the presented theory extension.

4.2.2 Applications

The concept of VT has been applied in several contexts so far, with a few studies advancing the existing VT modeling techniques. FRIESZ, HAN, NETO, MEIMAND & YAO [2013] developed a dynamic user equilibrium framework for networks. Traffic dynamics were modeled based on the LWR theory and numerically represented with VT. Yet, they did not consider spillbacks across intersections in their study, and thus underestimated delays for heavily congested scenarios. Similarly, LI & ZHANG [2015] described the performance of queuing systems with multiple sequential and parallel bottlenecks based on VT. However, their framework did not account for spillover effects. CHOW ET AL. [2015] utilized the ability of VT to model complex traffic dynamics at the corridor level on urban arterials in London (UK) and compared it to the CTM. They confirmed the high-quality results of VT for such settings, especially for platoon dispersion and moving bottlenecks. HANS, CHIABAUT & LECLERCQ [2015] developed a mesoscopic model based on VT to exactly estimate travel times for urban arterials. While the application of VT was still limited to the link and corridor levels, these studies show that VT has been recognized as a powerful numerical scheme for solving KWT problems.

VT is not restricted to the study of car traffic only. It has also been utilized for multi-modal traffic analyses. GULER ET AL. [2016] investigated innovative transit signal priority designs and estimated their effects on intersection performance with VT. Similarly, WU ET AL. [2017] analyzed the impacts of bus stop locations and transit signal priority on intersection operations. WU & GULER [2018] examined the optimal locations of transit signals in the context of mixed traffic with cars and buses. GAYAH ET AL. [2016] studied the impacts of general obstructions on the capacity of an isolated intersection and computed the capacity losses using VT. SAEEDNIA & MENENDEZ [2016] evaluated the effect of truck platoons on freeway traffic also using VT.

Another field of application is traffic state estimation and data fusion [MEHRAN ET AL., 2012; SUN & BAN, 2013]. This line of research has explored the application of VT to reconstruct vehicle trajectories based on data from fixed and/or mobile sensors, as well as data fusion algorithms. MEHRAN & KUWAHARA [2013] further extended this approach to predict vehicle trajectories based on real-time and historical data. DURET & YUAN [2017] proposed a traffic state estimation framework based on Eulerian and Lagrangian observations. Their framework includes the concept of VT. CHEN, WEI, MENG & ZHENG [2020] reconstructed trajectories of vehicles using a hybrid approach that integrates VT and Kalman filtering. More works on developing data fusion frameworks exploiting the concept of VT concern real-time applications [KAWASAKI, HARA & KUWAHARA, 2017] and measurements from vehicles running on opposite lanes [KAWAI, TAKENOUCHI, IKAWA & KUWAHARA, 2019; TAKENOUCHI, KAWAI & KUWAHARA, 2019].

Additionally, VT has been applied for the approximation of the MFD. As described in Chapter 2, the MC introduced by DAGANZO & GEROLIMINIS [2008] estimates the MFD for a homogeneous ring road based on VT. Accordingly, related extensions and studies either refer to or utilize VT [BOYACI & GEROLIMINIS, 2011; LECLERCQ & GEROLIMINIS, 2013; LAVAL & CASTRILLÓN, 2015; TILG ET AL., 2020a; DAGANZO & KNOOP, 2016; LECLERCQ & PAIPURI, 2019; GIRAULT ET AL., 2016; AMBÜHL ET AL., 2020; LODER, DAKIC, BRESSAN, AMBÜHL,

BLIEMER, MENENDEZ & AXHAUSEN, 2019a]. Again, these approximation methods apply VT only at the corridor level, even when trying to estimate traffic conditions at the network level. Early works have addressed the inclusion of turning ratios into the MFD approximation, see TILG, AMBÜHL, BATISTA, MENENDEZ, LECLERCQ & BUSCH [2021c] and XU ET AL. [2020]. A comprehensive approach for this purpose is presented in Chapter 5.

4.2.3 Methodological extensions

Not only has VT been applied in several contexts, but a number of attempts to extend the original framework have also been reported. For example, LEBACQUE & KHOSHYARAN [2013] showed that generic second-order models admit a Hamilton-Jacobi and variational formulation as an optimal control problem. COSTESEQUE & LEBACQUE [2014] numerically investigated the VT formulation for higher-order traffic models. Additionally, the deterministic nature of VT was questioned, and by the inclusion of stochastic shortest path algorithms, new fields of application were made accessible. WADA ET AL. [2018] applied VT for coordinated traffic signal control for both deterministic and stochastic demands. DAKIC ET AL. [2020] applied a stochastic shortest path search within VT to estimate the capacity of bi-modal corridors.

Another interesting line of research is the impact of source terms in VT, i.e. the existence of inflows and outflows in corridors. Eq.(2.3) in Chapter 2 implies there are no inflows nor outflows within the corridor. This is because the occurrence of such source terms violates flow conservation along the corridor, which is the main assumption of the conservation equation (see eq.(2.1) in Chapter 2). While the assumption of zero net inflows might hold for singular corridors, it represents a major limitation for applying VT at the network scale. LAVAL ET AL. [2016] studied the incorporation of such source terms into the VT framework. They developed a framework to consider continuous inflows and outflows (i.e. source terms) which could represent trips starting and ending throughout the link. However, their method is not able to handle source terms at intersections, as they are discrete in space and time, and include non-zero values during green and values equal to zero during red phases.

The wide range of applications as well as the methodological extensions of the basic concept shows the general interest in VT. These works indicate that extending its applicability from corridors to networks is a valuable contribution. In the next section, we introduce a framework to apply VT at the network level. Thereby, we account for source terms in VT at intersections and model spillbacks. We refer to our framework as ‘nVT’ as an abbreviation for network VT, and as ‘original VT’ to the formulation introduced by DAGANZO [2005b].

4.3 Generalizing variational theory to networks

In this chapter, we modify VT and establish a comprehensive model that allows us to numerically solve complex heterogeneous KWT problems at the network level. We make the following assumptions:

1. *Fundamental diagram*: The application of nVT requires a concave and piecewise differentiable FD. This assumption is inherited from the original VT. Triangular FDs with θ being integer-valued facilitate the modeling of traffic in signalized networks.
2. *Variational graph*: In order to consider the effects of bottlenecks as well as source terms, their trajectories need to be fully mapped in G.
3. *Queueing discipline:* Traffic flow follows a first-in-first-out (FIFO) rule.
4. *Network topology*: The proposed framework applies merely to signalized urban networks without any partially conflicting traffic streams. Therefore, the focus lies on networks with signalized intersections with dedicated phases for the conflicting traffic streams.
5. *Turning ratios*: The framework assumes constant turning ratios, which are an exogenous input.

The relaxation of these assumptions is discussed later in Chapter 6. Below, we describe the overall framework which consists of three steps.

Step 0 initializes the problem by defining the infrastructure and the demand. This consists of the definition of the network topology, the signal control settings, and the temporal and spatial demand patterns. The latter aspect includes the origins as well as turning ratios at each intersection. Note that this implies an indirect definition of destinations. This allows generating the network $\mathcal{N}$ and the data at the boundaries.

Step 1 involves the decomposition of the network $\mathcal{N}$ into a set of corridors $C \in \mathcal{C}$. This enables us to define a multi-dimensional variational graph in order to solve the given KWT problem. The graph has $(|\mathcal{C}| + 1)$ dimensions, consisting of the corridors in $\mathcal{C}$, and one temporal dimension. We provide more details on this step in subsection 4.3.1.

Step 2 applies our VT framework, the nVT, taking the multi-dimensional variational graph as input. Our framework builds an extension that incorporates source terms into the original VT concept. We treat turning flows as source terms at the location of intersections that are discrete in time and space. Moreover, the model replicates spillbacks across intersections in the network. This step is explained in detail in subsection 4.3.2.

4.3.1 Step 1: Decomposition of networks into corridors

The problem initialization results in a network $\mathcal{N}$ consisting of intersections $I \in \mathcal{I}$ and links $L \in \mathcal{L}$. In this chapter, we define a link as the road segment between two intersections. Additionally, we specify signal control settings, i.e. red and green times, and turning ratios $\alpha \in A$ at each intersection $I \in \mathcal{I}$, as well as origin flows. As discussed above, the definition of the variational graph requires the decomposition of the network $\mathcal{N}$ into a set of corridors $\mathcal{C}$. We define a corridor as an ordered sequence of links in the same direction of travel. Thereby, the topological order including turning ratios α_{ij} between each pair of corridors C_i and C_j as well as control settings for $I \in \mathcal{I}$ have to be stored. Turning ratios are assumed to be constant. The implications of this assumption are discussed in Chapter 6.

There are two main requirements for decomposing the network $\mathcal{N}$ into a variational graph:

1. The set $\mathcal{C}$ includes all links $L \in \mathcal{L}$.
2. Each link L exists only once in $\mathcal{C}$, i.e. the corridors do not have any overlapping segments.

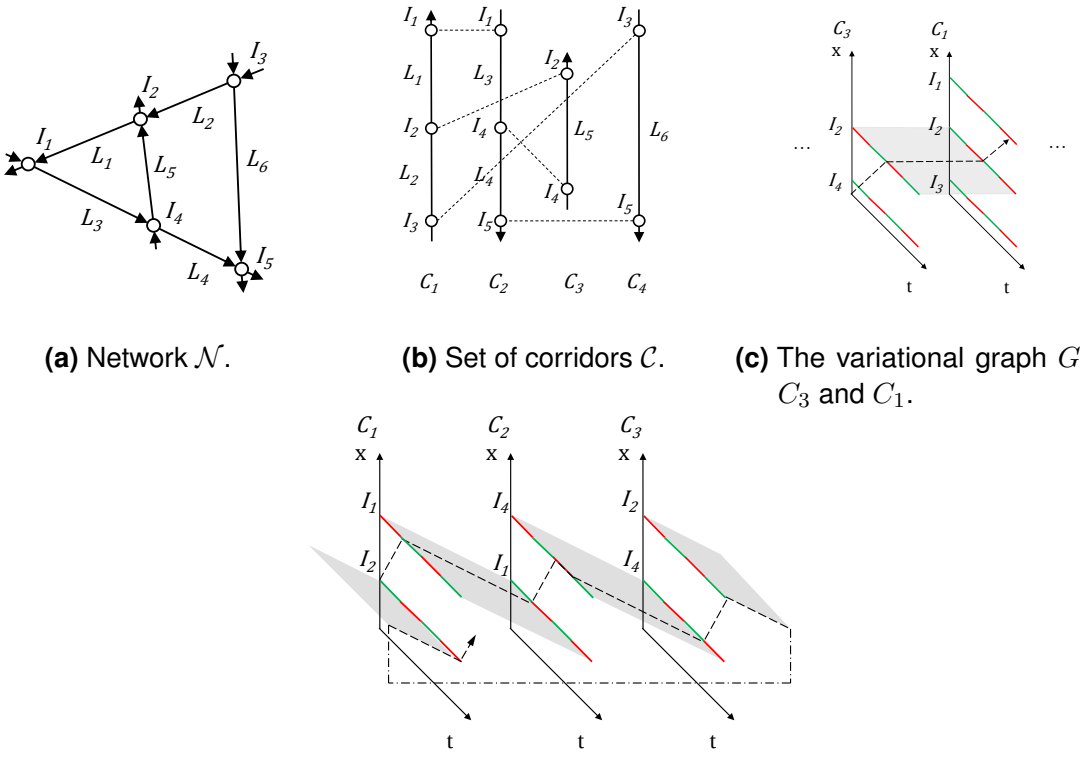
For small toy networks, we can determine the set $\mathcal{C}$ manually from $\mathcal{N}$. For the most interesting case of real networks, this decomposition can be performed according to the actual layout of roads. From a practical perspective, it is very convenient to decompose real networks according to the network structure as defined by local municipalities. Each link in a real network is part of a 'street', defined by its name, and such streets are non-overlapping. Each street can be represented as a corridor C and as such can be incorporated in G. This will always satisfy both conditions mentioned above. Note that the resulting decomposed network, i.e. the set $\mathcal{C}$, will not affect the result. The KWT solution is determined by the demand (origin flows and turning ratios), as well as the supply (link FDs, link lengths, etc.). As these parameters are not affected by assigning links L to corridors C, the KWT and therefore the VT solution are unaffected as well. In other words, the KWT solution is independent of the decomposition method.

Figure 4.1 schematically illustrates the process with an example. Figure 4.1a depicts a network $\mathcal{N}$ consisting of five intersections, $\mathcal{I} = \{I_1, I_2, I_3, I_4, I_5\}$, and six unidirectional links, $\mathcal{L} = \{L_1, L_2, L_3, L_4, L_5, L_6\}$. The direction of travel on each link is indicated with arrows. Note that our method is not limited to unidirectional links, but this assumption simplifies the example. The intersections connecting the different corridors play an important role, as they serve as interfaces where flow is transferred from one corridor to the other, i.e. where source terms apply. We denote such intersections as inter-corridor connections to distinguish them from intersections without any turning flows (e.g. pedestrian crossings). For our example, we assume that $\alpha > 0$ for all intersections. That is, they are all inter-corridor connections.

We manually decompose $\mathcal{N}$ into a set $\mathcal{C} = \{C_1, C_2, C_3, C_4\}$ as shown in Figure 4.1b. This set contains only non-overlapping corridors covering the entire network. Both requirements stated above are then fulfilled. The inter-corridor connections are highlighted as dashed lines. As observed, the network's topology is fully retained in this decomposition.

In Figure 4.1c one can see an excerpt of the multi-dimensional variational graph G for corridors C_3 and C_1. Red and green phases are represented by the corresponding colors at the inter-corridor connections. The grey plane illustrates the inter-corridor connection I_2, and a single trajectory from a vehicle changing from one corridor to the other is shown as a dashed line for illustration purposes.

Figure 4.1d shows the graph G for corridors C_1, C_2, and C_3 including inter-corridor connections represented as grey planes. These corridors include circular route dependencies. This is highlighted by a trajectory, shown as a bold dashed line, and illustrating a vehicle that starts to drive at intersection I_2, changes the corridor several times, and ends up at intersection I_2 again, but some time later. The figure demonstrates that such trajectories are still feasible within the decomposed network. These trajectories will be reflected by the corresponding cumulative counts. In other words, our framework works independently of demand patterns



(a) Network $\mathcal{N}$.

(b) Set of corridors $\mathcal{C}$.

(c) The variational graph G for C_3 and C_1.

(d) The variational graph G for C_1, C_2, and C_3 including circular routes.

Figure 4.1: Schematic illustration of the network decomposition and definition of a multi-dimensional variational graph G.

whether phenomena such as circular route dependencies exist or not. Such spatial demand aspects are implicitly included in the turning ratios, which are considered as input in our framework, as described in the following sections.

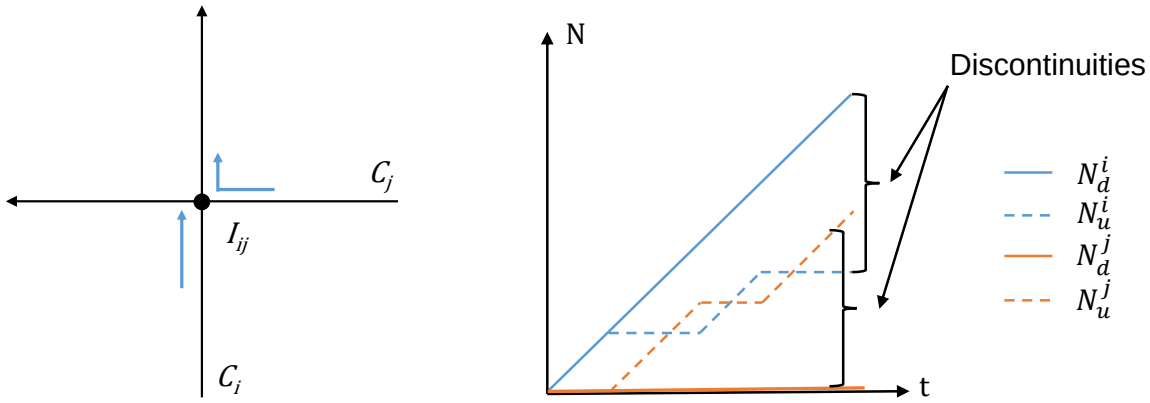
4.3.2 Step 2: Integration of inter-corridor connections

Once the multi-dimensional variational graph G is defined, we apply nVT to solve the KWT problem. The original VT formulation, introduced in Chapter 2, does not apply as it does not account for discrete source terms. The implications of source terms for variational theory and the underlying concept of the Moskowitz function are shown first. Then, mathematical formulations for upstream and downstream traffic state propagation across inter-corridor connections are presented.

Discontinuous Moskowitz function

In this section, we highlight the effects of source terms on the Moskowitz function on corridors by means of two examples, a ‘free-flow case’ and a ‘congested case’.

Consider two uni-directional corridors C_i and C_j that intersect each other at the signalized inter-corridor connection I_{ij} as displayed in Figure 4.2a. As there exist no bottlenecks on the corridors upstream of I_{ij}, we refer to this example as the 'free-flow case'. Let us denote the Moskowitz function at the position right downstream of I_{ij} on each corridor as N_d^i and N_d^j, respectively. The Moskowitz function for positions right upstream of I_{ij} are N_u^i and N_u^j, respectively. We further assume a high demand on both corridors, such that the inter-corridor connection is saturated. Moreover, we consider a case where all vehicles from C_j merge onto C_i, i.e. the turning ratio $\alpha_{ji} = 1$, and all vehicles on corridor C_i continue straight at the inter-corridor connection, i.e. $\alpha_{ij} = 0$. Figure 4.3b depicts possible time series of N_d^i, N_u^i, N_d^j, and N_u^j.



(a) Two corridors C_i and C_j and a simple inter-corridor connection.

(b) Moskowitz function N for the upstream and downstream positions.

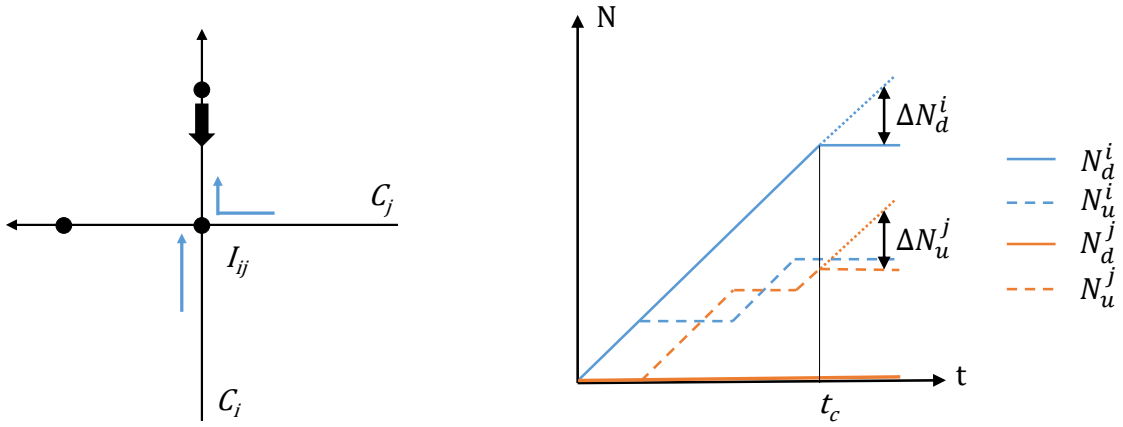
Figure 4.2: Free-flow case: Network and Moskowitz function.

The Moskowitz functions N_u^i and N_u^j at the upstream positions show discharge flows during green phases, and zero flows during red phases. The downstream Moskowitz function for C_i is the sum of both upstream functions, i.e. $N_d^i = N_u^i + N_u^j$, assuming that the travel time between x_u and x_d is negligible. The Moskowitz function downstream of the inter-corridor connection on corridor C_j equals to zero, i.e. $N_d^j = 0$, since no vehicles enter from C_i and all flow on C_j turns at the inter-corridor connection I_{ij}.

Recall that the Moskowitz function is in general continuous and respects the flow conservation principle (see Section 2.2). Since this assumption also applies to the original VT, it requires a continuous surface across space x and time t. However, in our example, a discontinuity of the function occurs at the inter-corridor connection I_{ij}, as N on the same corridor jumps drastically from upstream to downstream of I_{ij} (see Figure 4.2b). In other words, no flow conservation applies considering each corridor separately although flow is indeed conserved at the network level, i.e. when flows are aggregated across the two corridors. Note that the discontinuous change in N is positive when a net inflow occurs (e.g. at corridor C_i), and negative when a net outflow occurs (e.g. at corridor C_j). Additionally, the absolute value of the discontinuity increases with time depending on the volumes of the transfer flows.

The second example is the 'congested case' which is depicted in Figure 4.3. We examine

a simple network with active bottlenecks somewhere downstream of the inter-corridor connection I_{ij}, from which congestion propagates upstream, as schematically illustrated in Figure 4.3a. The bottlenecks on each corridor are displayed by the black points, while the bold black arrow represents the back-propagation of congestion. Figure 4.3b shows possible time-series of the Moskowitz functions on both corridors, upstream and downstream of I_{ij}, i.e. N_d^i, N_u^i, N_d^j, and N_u^j.



(a) Two corridors C_i and C_j including two bottlenecks and a simple inter-corridor connection.

(b) Moskowitz function N for the upstream and downstream demands.

Figure 4.3: Congested case: Network and Moskowitz function.

In the figure, the occurrence of the spillback at t_c appears as a reduction in N_d^i growing with time which we denote as ΔN_d^i. To ensure the correct propagation of spillbacks across inter-corridor connections, we have to account for the discontinuity at I_{ij}. Moreover, we have to define how to apportion ΔN_d^i to the N_u of each corridor. As discussed in the previous section, N_d^i is the sum of N_u^i and N_u^j. In order to divide ΔN_d^i into the summands, we have to consider its time-dependency. Flows from the connected corridors can never occur simultaneously in the case of a signalized intersection without partially conflicting flows. This implies that congestion only propagates to the currently discharging corridor. Thus, only the Moskowitz function of this corridor is adapted. In our example, congestion is propagated from N_d^i to N_u^j, i.e. N_u^j is reduced by $\Delta N_u^j = \Delta N_d^i$. Therefore, ΔN_d^i is transferred to N_u^j, while N_u^i stays constant until corridor C_i gets the green light at the inter-corridor connection.

These examples for the ‘free-flow case’ and the ‘congested case’ highlight the existence of a discontinuity of the Moskowitz function at inter-corridor connections. Moreover, they show that the propagation of spillbacks necessitates the knowledge of the current signal phase. We have to extend the concept of VT to a discontinuous Moskowitz function to be able to solve the KWT represented by the variational graph G. In the following, we propose a framework to cope with this discontinuity for both the downstream and upstream traffic state propagation. This ensures a complete model including network-wide effects of spillbacks. For the sake of clarity, we avoid using the indices i, j as much as possible for the explanations below, and only use them when strictly necessary. We show that our methodology inherits the numerical exactness from the original VT for triangular FDs.

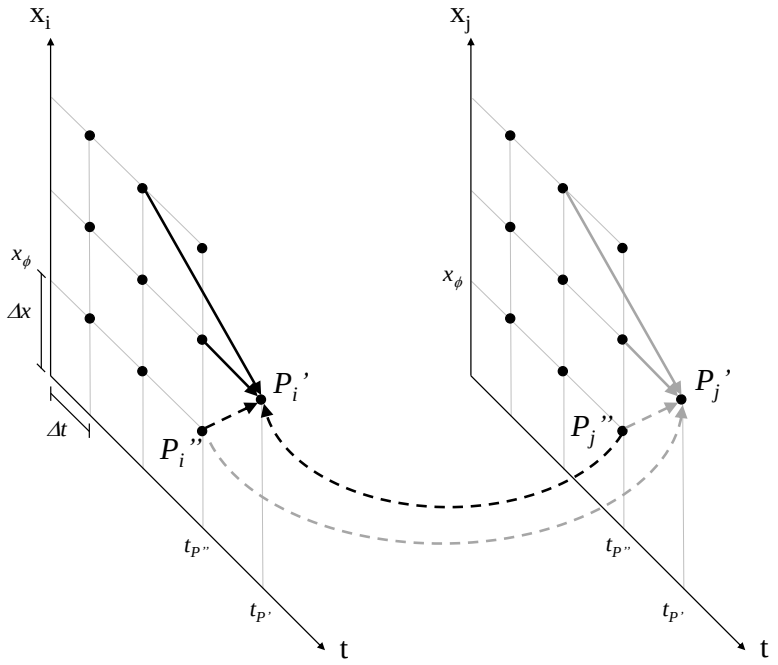
Multi-dimensional variational graph at inter-corridor connections

To further examine the effects of the discontinuous Moskowitz function, the analysis of corresponding excerpts of the variational graph G is convenient. We define the location of an inter-corridor connection I_{ij} in G as x_Φ. At this position, the turning ratio α_{ij} determines which portion of the flow is transferred from one corridor to another. Therefore, $x = x_\Phi + \Delta x$ is the first location in G for a specific corridor, where N is influenced by transfer flows. In other words, the discontinuity is located between x_Φ and $x_\Phi + \Delta x$. For transfer flows to be considered, they need to be fully included in G. In other words, the respective beginning and end times need to be mapped in G.

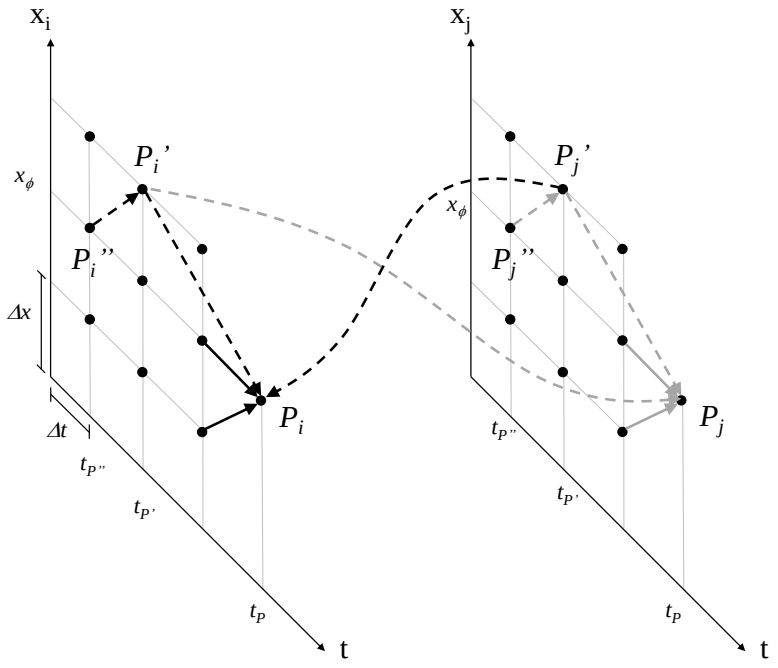
We recall that the original VT formulation (see Chapter 2) is valid as long as the Moskowitz function is continuous. It is then valid for all (x, t), where x is not in the vicinity of the inter-corridor connection, i.e. $x \neq x_\Phi$ and $x \neq x_\Phi + \Delta x$. Between those positions, a discontinuity exists and the original formulation has to be modified as per the following two criteria. First, recall that the graphical interpretation of eq.(2.4) consists of three from-nodes and the corresponding edges (see Figure 2.4 in Chapter 2). In order to account for transfer flows in the case of multiple corridors, eq.(2.4) should also reflect the traffic states from the adjacent corridor and the turning ratios. Second, eq.(2.4) cannot be evaluated at x_Φ and $x_\Phi + \Delta x$ because the edges in G related to the terms of the minimum operation cross the discontinuity which violates the underlying assumptions of VT. This is shown in Figure 4.4. Thus, eq.(2.4) must be modified for $x = x_\Phi$ and $x = x_\Phi + \Delta x$.

Figure 4.4 depicts two excerpts of the multi-dimensional variational graph G for corridors C_i and C_j around the inter-corridor connection located at x_Φ. Edges which cross the discontinuity are shown as dashed lines. The terms in the minimum operation of eq.(2.4) corresponding to these dashed edges need to be modified in order to account for the existence of the discontinuity. The black color represents edges which are evaluated for corridor C_i, and the grey those being evaluated for corridor C_j. We label the points which have to be treated differently from the original formulation. These points are P'' at $(x_\Phi, t_{P''})$, P' at $(x_\Phi + \Delta x, t_{P'})$, and P at (x_Φ, t_P). Note that $t_{P''} = t_P - (1 + \theta)\Delta t$, $t_{P'} = t_P - \theta\Delta t$, and $t_{P''} = t_{P'} - \Delta t$. The following explanations refer to these points.

Figure 4.4a illustrates the propagation of traffic states from upstream nodes (i.e. free-flow traffic states). This propagation is affected by the discontinuity when the corresponding edges start at x_Φ and end at $x_\Phi + \Delta x$. Thus, the first term in eq.(2.4) has to be modified when calculating $N(x, t)$ for $x = x_\Phi + \Delta x$. Figure 4.4b shows the propagation of traffic states from downstream nodes (i.e. congested traffic states). It is affected by the discontinuity when the corresponding edges start at $x_\Phi + \Delta x$ and end at x_Φ. Consequently, the second term in eq.(2.4) has to be modified when calculating $N(x, t)$ for $x = x_\Phi$. The following subsections describe this extension of eq.(2.4) for the propagation of traffic states from upstream and downstream nodes in the vicinity of the inter-corridor connection, i.e. at $x = \{x_\Phi, x_\Phi + \Delta x\}$. Note that downstream and upstream propagation of traffic states do not necessarily refer to the same pair of corridors. The propagation depends on the specific intersection layout and turning possibilities. Consider the following example: At an inter-corridor connection of three corridors A, B, and C, flow can transfer from corridor A to corridor B, and from corridor B to



(a) Downstream propagation of traffic states.



(b) Upstream propagation of traffic states.

Figure 4.4: Excerpts of the multi-dimensional variational graph G at the inter-corridor connection located at x_Φ.

corridor C. That implies that upstream traffic states propagate from corridor A to corridor B, and downstream traffic states propagate from corridor C to corridor B. Such examples are included in the case study shown in Section 4.6.

Propagation of traffic states from upstream nodes at $x = x_\Phi + \Delta x$

This section explains how to include source terms when traffic states propagate from upstream nodes across the discontinuity of the Moskowitz function. In this case, the first term in the minimum operator in eq.(2.4) becomes decisive, as it refers to the free-flow states traveling from upstream. We modify this term at $x = x_\Phi + \Delta x$ to account for the upstream demand and incorporate inflows and outflows based on turning ratios at inter-corridor connections.

Recall that each term in eq.(2.4) consists of a known N and the costs z along a path. In order to modify the equation, we adapt these two elements for its first term. First, we focus on the known N-value at the upstream node, which equals to $N_{P''}$ following the designations in Figure 4.4a. We assume that N is known for all $t < t_{P'}$. This is feasible because we can solve for N moving from left to right in the variational graph. The turning ratio α_{ij} specifies the transfer flow from $x_{i,\Phi}$ to $x_{j,\Phi} + \Delta x$. Similarly, the turning ratio α_{ji} specifies the transfer flow from $x_{j,\Phi}$ to $x_{i,\Phi} + \Delta x$. The demand at P'_i can be calculated as the sum of the flow which stays on the corridor C_i, i.e. $N^i_{P''} \cdot (1 - \alpha_{ij})$, and the inflow from corridor C_j, i.e. $N^j_{P''} \cdot \alpha_{ji}$. This is depicted by the black dashed lines in Figure 4.4a for C_i.

Second, we examine the costs z associated with the path from P'' to P'. Note that $z = 0$ for the propagation of traffic states from upstream nodes, as such states correspond to free-flow conditions. This fact does not change by considering inflows, and therefore the costs for these edges remain equal to zero.

Eventually, we determine the term $N^i_{P'}$ by evaluating the following equation:

$$\begin{aligned} N^i_{P'}(x_\Phi + \Delta x, t_{P'}) = \min\{ & N^i_{P''}(x_\Phi, t_{P''}) \cdot (1 - \alpha_{ij}) + N^j_{P''}(x_\Phi, t_{P''}) \cdot \alpha_{ji}, \\ & N^i(x_\Phi + 2 \cdot \Delta x, t - \theta \Delta t) + \Delta x \kappa_{max}, \\ & N^i(x_\Phi + \Delta x, t - \Delta t) + \beta \}. \end{aligned} \tag{4.1}$$

The term $N^j_{P'}$ can be determined analogously. For both corridors, it can be seen that this equation only requires to consider $N_{P''}$ at the position of the inter-corridor connection x_Φ. All other N are known from the original VT formulation. Note that signal phases do not have to be considered explicitly as they are already reflected in the values of $N^i_{P''}$ and $N^j_{P''}$, i.e. they are implicitly taken into account.

Propagation of traffic states from downstream nodes at $x = x_\Phi$

In order to complete our modeling framework, we have to ensure the correct propagation of traffic states from downstream nodes across the inter-corridor connections where disconti-

nuities of the Moskowitz function appear. Recall that only congested traffic states can be propagated from downstream. Thus, we modify the second term in eq.(2.4) for $x = x_\Phi$ which is decisive when congestion occurs. In other words, in this section, we explain how to model spillbacks. Without the loss of generality, the formulation in the following refers to corridor C_i only. The formulation applies analogously for evaluating the traffic conditions on corridor C_j. For the sake of clarity, we avoid using the coordinates of the points P'', P', and P in the following formulations except for eq.(4.7). They are described above and illustrated in Figure 4.4.

Again, the second term of the minimum operation consists of two elements. The first element is the known Moskowitz function value $N_{P'}$ at the downstream node P' at $(x_\Phi + \Delta x, t - \theta\Delta t)$. The second one refers to the costs z, i.e. the number of vehicles that can pass a moving observer traveling from the downstream node P' to the point P at (x_Φ, t).

We first focus on the value $N_{P'}$. Note that any changes in $N_{P'}$, i.e. $\Delta N_{P'}$ (see Figure 4.3b) should first be apportioned to each of the corridors upstream of I_{ij} according to the signal phases and turning ratios. In the absence of congestion the portion of $N_{P'}$ which corresponds to the Moskowitz surface related to N_P would simply be $N_{P''} \cdot (1 - \alpha_{ij})$ as per the first term in eq.(4.1). The effect of congestion can be described as a reduction of N, denoted as ΔN in Figure 4.3b. Since we know $N_{P'}$ and $N_{P''}$, we can calculate ΔN. Note that corresponding spillbacks can originate on both corridors C_i and C_j. Thus, we have to consider two different cases for each corridor. For corridor C_i we can write:

$$\Delta N^{ii}_{P'} = \left(N^i_{P''} \cdot (1 - \alpha_{ij}) + N^j_{P''} \cdot \alpha_{ji}\right) - N^i_{P'}, \tag{4.2a}$$

$$\Delta N^{ij}_{P'} = \left(N^i_{P''} \cdot \alpha_{ij} + N^j_{P''} \cdot (1 - \alpha_{ji})\right) - N^j_{P'}. \tag{4.2b}$$

Recall that VT is built upon the concept of the moving observer. Keeping this notion in mind, one can interpret $\Delta N_{P'}$ as the difference in the number of vehicles a moving observer traveling from $N_{P''}$ to $N_{P'}$ passes due to congestion compared to free-flow conditions. The indices i and j correspond to the corridors where flows originate from and propagate to, respectively. For example, $\Delta N^{ij}_{P'}$ relates to vehicles in congestion coming from corridor C_i with destination in corridor C_j. $\Delta N^{ii}_{P'}$ relates to vehicles in congestion coming from C_i with destination in corridor C_i. Note that the portion of $N_{P'}$ related to the upstream Moskowitz functions on each corridor is determined based on the respective turning ratios α, as shown for the free-flow case in eq.(4.1). This has to be considered when the change in the vehicle number $\Delta N_{P'}$ is apportioned. We do so by dividing $\Delta N_{P'}$ by the turning ratio associated with each corridor. As an example assume a turning ratio of α_{ij} = 50 % and a $\Delta N^{ij}_{P'} = 5$ vehicles. That would mean that 5 vehicles coming from P''_i did not reach P'_j due to congestion. Moreover, this $\Delta N^{ij}_{P'} = 5$ would mean that 10 vehicles departed P''_i as only every second vehicle wanted to travel to P'_j. To find then the portion of $N_{P'}$ corresponding to the Moskowitz function related to N_P, denoted as $\widehat{N}_{P'}$ in eq.(4.3), we subtract from $N_{P''}$ the total number of vehicles leaving P'' being blocked by the congestion at P'. This lets us formulate the effects of congestion at P' related to the Moskowitz function at x_Φ.

$$\widehat{N}^{ii}_{P'} = N^{i}_{P''} - \Delta N^{ii}_{P'} \cdot \frac{1}{1-\alpha_{ij}}, \tag{4.3a}$$

$$\widehat{N}^{ij}_{P'} = N^{i}_{P''} - \Delta N^{ij}_{P'} \cdot \frac{1}{\alpha_{ij}}. \tag{4.3b}$$

To find the upper bound for N_P due to congestion, we further have to consider the costs along the path from P' to P which corresponds to the second element of the term related to congestion in eq.(2.4). Again, we consider the case of congestion propagation only on corridor C_i. Note that the costs are $\Delta x \kappa_{max}$ in the original formulation. In our case, these costs only apply when the signal phase does not change during the interval $[t-(1+\theta)\Delta t, t]$, i.e. in the time interval between P'' and P. However, when a signal phase change occurs during that interval, flows from both corridors might occur at different times within the interval. Thus, the costs z include vehicles originating from both corridors. We need to apportion z according to the vehicles' origin for correct spillback propagation. In other words, we need to ensure that costs only reflect the vehicles related to the corridor where congestion is propagating to. Otherwise, we would overestimate the costs of the edge, and consequently the second term in eq.(2.4). Ultimately, it could result in an overestimation of N_P and thus in an erroneous traffic state propagation from downstream. Assuming reasonable backward wave speeds and that there is only one signal phase change during the considered time interval, there are only two possibilities for such a phase change to occur:

- *Change from red to green:* In this case, a green phase is active at $t = t_P$. The correct maximum number of vehicles from C_i that a moving observer traveling from P'_i to P_i and from P'_j to P_i would count has thus to be reduced by the vehicles coming from C_j. Let the variable Δr denote the duration of the red phase for C_i starting at $t = t_{P''} = t_P - (1+\theta)\Delta t$. This is equivalent to the green time for C_j for that same period. Then, the maximum number of vehicles discharging from C_j and going to C_i is $\Delta x \kappa_{max} \cdot \frac{\Delta r}{(1+\theta)\Delta t} \cdot \alpha_{ji}$. Consequently, the maximum number that can discharge from C_i and stay on that corridor is $\Delta x \kappa_{max} - \Delta x \kappa_{max} \cdot \frac{\Delta r}{(1+\theta)\Delta t} \cdot \alpha_{ji}$. As for $\Delta N_{P'}$, we consider that a part of the total inflow on corridor C_i stays on that corridor and a part turns into corridor C_j, so we divide by the appropriate turning ratios, i.e. $1-\alpha_{ij}$ for flow staying in C_i and α_{ij} for flow from C_i to C_j. The costs z can then be formulated as:

$$z^{ii} = \frac{1}{1-\alpha_{ij}} \cdot \left(\Delta x \kappa_{max} - \Delta x \kappa_{max} \cdot \frac{\Delta r}{(1+\theta)\Delta t} \cdot \alpha_{ji} \right), \tag{4.4a}$$

$$z^{ij} = \frac{1}{\alpha_{ij}} \cdot \left(\Delta x \kappa_{max} - \Delta x \kappa_{max} \cdot \frac{\Delta r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right). \tag{4.4b}$$

- *Change from green to red:* In this case, a red phase is active at $t = t_P$. Let the variable Δg denote the duration of the green phase starting at $t = t_{P''} = t_P - (1+\theta)\Delta t$. Then, the maximum number of vehicles discharging from C_i and staying on this corridor is $\Delta x \kappa_{max} \cdot \frac{\Delta g}{(1+\theta)\Delta t} \cdot (1-\alpha_{ij})$. Analogously to the previous case, an formulation for the costs z can be found. However, for the case of an active red phase at $t = t_P$, the term

related to capacity constraints in eq.(2.4) becomes decisive. Therefore, and for the sake of clarity, we do not explicitly state the corresponding mathematical formulation in this chapter.

The sum of $\widehat{N}_{P'}$ from eq.(4.3) and the costs z from eq.(4.4) is an important step for the final formulation of the spillback-induced upper bound for N_P. This sum can be written as:

$$\widehat{N}^{ii}_{P'} + z^{ii} = N^i_{P''} + \frac{1}{1-\alpha_{ij}} \cdot \left[\Delta x \kappa_{max} \cdot \left(1 - \frac{\Delta r}{(1+\theta)\Delta t} \cdot \alpha_{ji} \right) - \Delta N^{ii}_{P'} \right], \tag{4.5a}$$

$$\widehat{N}^{ij}_{P'} + z^{ij} = N^i_{P''} + \frac{1}{\alpha_{ij}} \cdot \left[\Delta x \kappa_{max} \cdot \left(1 - \frac{\Delta r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right) - \Delta N^{ij}_{P'} \right]. \tag{4.5b}$$

Note that the term in the squared brackets can become negative. This is for example the case when $\Delta N_{P'} = \Delta x \kappa_{max}$, i.e. the space between P'' and P' is fully congested, and both Δr and α are unequal to zero. Since the Moskowitz function is a monotonically increasing function, such values are impossible. We avoid them by applying a maximum function on the term in the squared brackets. Then, we can describe the influence of congestion from corridors C_i and C_j on the Moskowitz function N^i_P as:

$$N^i_P = N^i_{P''} + \frac{1}{1-\alpha_{ij}} \cdot \max \left\{ 0, \Delta x \kappa_{max} \cdot \left(1 - \frac{\Delta r}{(1+\theta)\Delta t} \cdot \alpha_{ji} \right) - \Delta N^{ii}_{P'} \right\}, \tag{4.6a}$$

$$N^i_P = N^i_{P''} + \frac{1}{\alpha_{ij}} \cdot \max \left\{ 0, \Delta x \kappa_{max} \cdot \left(1 - \frac{\Delta r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right) - \Delta N^{ij}_{P'} \right\}. \tag{4.6b}$$

This equation formally corresponds to the intuition that we have to consider the different Moskowitz functions upstream and downstream of the discontinuity, as well as the current signal phases. Eq.(4.6a) becomes constraining when spillbacks come from corridor C_i, and eq.(4.6b) becomes constraining for spillbacks from corridor C_j. The general formulation for $N_P(x_\Phi, t)$ includes then eq.(4.6), as well as the free-flow and the capacity related terms. Note that this model conceptually corresponds to a FIFO diverge.

$$\begin{aligned} N^i_P(x_\Phi, t) = \min\{ & \\ & N^i(x_\phi - \Delta x, t - \Delta t), \\ & N^i_{P''}(x_\phi, t - (1+\theta)\Delta t) \\ & + \frac{1}{1-\alpha_{ij}} \cdot \max \left\{ 0, \Delta x \kappa_{max} \cdot \left(1 - \frac{\Delta r}{(1+\theta)\Delta t} \cdot \alpha_{ji} \right) - \Delta N^{ii}_{P'} \right\}, \\ & N^i_{P''}(x_\phi, t - (1+\theta)\Delta t) \\ & + \frac{1}{\alpha_{ij}} \cdot \max \left\{ 0, \Delta x \kappa_{max} \cdot \left(1 - \frac{\Delta r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right) - \Delta N^{ij}_{P'} \right\}, \\ & N^i(x_\phi, t - \Delta t) + \beta \}. \end{aligned} \tag{4.7}$$

4.3.3 Implementation

The proposed equations enable us to numerically derive the KWT solution for a network with turning flows. The corresponding pseudo-algorithm is shown in algorithm 1. It describes the application of the proposed equations for a network $\mathcal{N}$. Based on this network, the set of corridors $\mathcal{C}$ is derived. It includes all turning ratios A, the control settings, and the inflow at origins. For a total simulation period T, one iterates through all t for each corridor C and evaluates eq.(2.4), and eq.(4.1) - eq.(4.7).

Algorithm 1 Application of nVT to a network $\mathcal{N}$

Step 0: Initialize network $\mathcal{N}$ including turning ratios A, control settings, and inflows at origins
Step 1: Derive the set of corridors $\mathcal{C}$ from $\mathcal{N}$
Step 2: Apply the nVT framework on $\mathcal{C}$:
for t in T **do**
  for $C \in \mathcal{C}$ **do**
    Evaluate eq.(2.4) for all $x \neq x_\Phi$ and $x \neq x_\Phi + \Delta x$
    Evaluate eq.(4.1) for all $x = x_\Phi + \Delta x$
    Evaluate eq.(4.7) for all $x = x_\Phi$
  end for
end for

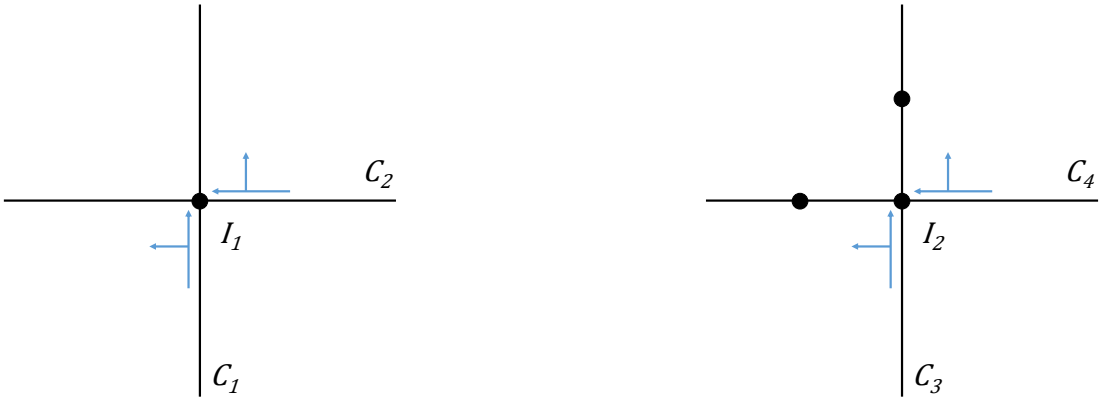
The nVT model allows to solve complex heterogeneous KWT problems related to networks with signalized control. Such problems involve source terms that are discrete in time and space. As VT allows the incorporation of any type of bottlenecks, nVT inherits this ability. The required inputs are the FD, turning ratios, origin flows, the network topology, and control settings. In the next sections, we will further evaluate our model, and compare it to a microscopic KWT simulation in a case study.

4.4 Proof of concept

In this section, we investigate the performance of our nVT model on a small toy network. This proof of concept includes the evaluation of the incorporation of source terms and the propagation of congestion across inter-corridor connections.

We test nVT for a ‘free-flow scenario’ and a ‘congested scenario’ on the networks depicted in Figure 4.5. The network on the left consists of two corridors, C_1 and C_2, that intersect each other at the inter-corridor connection I_1. Recall that we defined inter-corridor connections as intersections with turning ratios $\alpha > 0$. While the right-hand network is similar, two additional intersections exist, one on each corridor. These intersections act as active bottlenecks downstream of the inter-corridor connection I_2 and therefore lead to congestion. The corridors are labeled as C_3 and C_4, respectively. The designation of the two scenarios refers to the occurring traffic states downstream of the inter-corridor connection. We assume

possible turning movements for both inter-corridor connections I_1 and I_2, as illustrated by the blue arrows in the figure. We set a cycle length of 90 s, and a red and green phase of 45 s for all approaches at all intersections. All offsets are set to zero. We assume a triangular FD with jam density $\kappa_{max} = 150$ veh/km, a free-flow speed $u = 10$ m/s and a backward wave speed $w = -5$ m/s. The inflow demand is set to a volume-to-capacity ratio of $VOC = 1.0$, which refers to the intersection capacity, for both scenarios.



(a) Free-flow scenario.

(b) Congested scenario.

Figure 4.5: Test networks.

In the following, we first evaluate the traffic dynamics for the network displayed in Figure 4.5a. We discuss our implementation of source flows for the case where no congestion is propagated from downstream of I_1 (Section 4.4.1). We then analyze the results for the congested case (see Figure 4.5b), where the bottlenecks downstream of I_2 are active (Section 4.4.2). This will confirm that our proposed nVT model is able to capture traffic dynamics in its full range including spillbacks, which occur due to the active bottlenecks. Additionally, we verify our results by reproducing the traffic conditions on the same networks (given its topology and control settings) with the microscopic simulator SUMO [LOPEZ ET AL., 2018] and an implementation of Newell's car-following model. This verification is crucial as it allows us to show that our proposed nVT is indeed in line with classical KWT at the network level. As other studies [e.g. DAGANZO & LEHE, 2016; DAGANZO, GAYAH & GONZALES, 2011; GAYAH & DAGANZO, 2011], our implementation uses Newell's car-following model with unbounded acceleration/deceleration. Yellow intervals, dilemma zones, and reaction times are not modeled. Note that our validation does not verify whether KWT itself is an appropriate model for traffic, as this has been addressed by other studies [e.g. POLSON & SOKOLOV, 2015].

4.4.1 Inclusion of source terms

The evaluation of nVT for the first network enables us to isolate the effects of source terms for the case where no downstream congestion affects the traffic flow close to the inter-corridor connection. This allows us to study the effects of different turning ratios. We assume turning ratios of $\alpha_{12} = 0.25$ and $\alpha_{21} = 0.5$. The total simulated time is $T = 500$ s, and the time-step

length which determines the size of the multi-dimensional variational graph G is $\Delta t = 0.1$ s. Recall that this specifies Δx as well.

Figure 4.6 shows the results for both corridors as contour plots. The x-axis displays time in seconds, and the y-axis space in meters. The color bar represents the density ranging from bright yellow for low densities to dark blue for high ones. These densities were obtained with the proposed nVT model. We also show the trajectories extracted from SUMO as black and white curves, on top of the contour plots. The black curves depict trajectories originating on corridor C_1, while the white ones correspond to those originating on C_2. The effects of signal phases at the intersection at $x = 250$ m are clearly identified. The dark blue areas correspond to the jam density and thus represent the evolution of queues upstream of the inter-corridor connection on both corridors. The discharge flows colored in green can also be clearly observed in the figure. More important are the evident transfer flows between both corridors. This is illustrated by two facts. First, densities $\kappa > 0$ exist downstream of I_1 also during red phases. Second, the discharge flows are split at I_1, as depicted by the different densities upstream and downstream of I_1 for each corridor even during the green phase. This is highlighted by the change in colors. The trajectories from SUMO represent a perfect match to the predicted traffic states by our nVT model. They show that our framework successfully incorporates source terms for the case when congestion does not propagate to the inter-corridor connection.

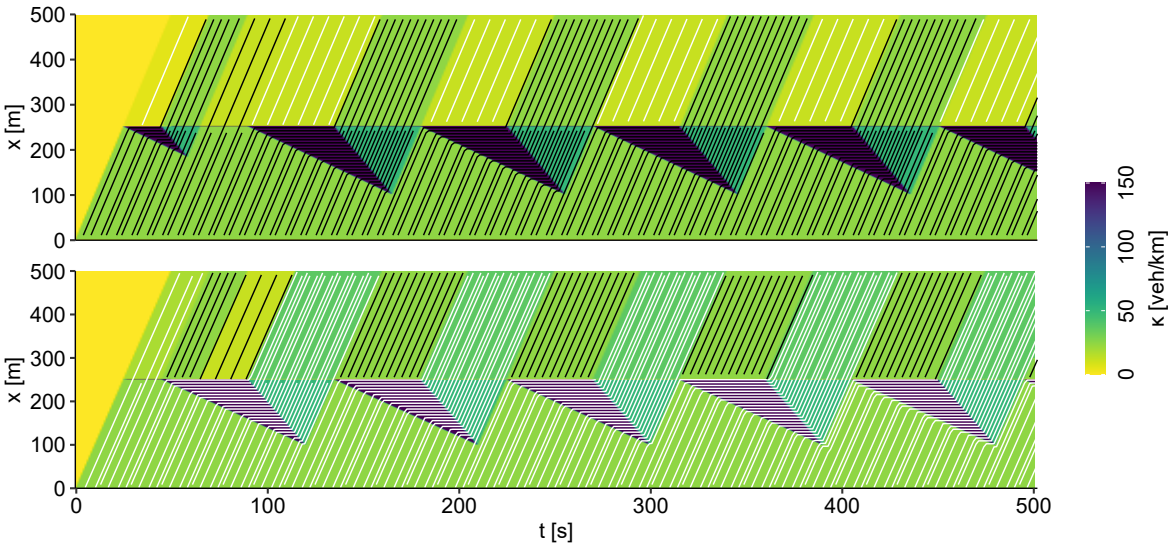


Figure 4.6: Free-flow scenario: Time-space diagrams for corridor C_1 (top) and corridor C_2 (bottom) with a traffic signal at I_1 ($x = 250$ m). The colors represent the density which was obtained with the nVT model. The curves indicate trajectories obtained with the microscopic traffic simulator, and their color indicates the origin (i.e black trajectories originate in C_1 and white ones originate in C_2).

4.4.2 Propagation of spillbacks

In the second step, we discuss a more general case where congestion appears downstream of the inter-corridor connection. Our proposed framework should propagate spillbacks along and across corridors. That is, the back of the queue has to spill over to at least one of the corridors, depending on the signal phase.

Figure 4.7 shows the simulation results. The axis and colors of the figure are the same as those of Figure 4.6. The black curves depict trajectories originating on corridor C_3, while the white ones correspond to those originating on C_4. The graph clearly shows the effects of the additional signals at $x = 400\,\text{m}$ which behave as active bottlenecks. The congestion starting on corridor C_3 propagates to I_2, where it affects both corridors upstream of I_2. This can be seen by the queues which grow on both corridors with each cycle. Also, it can be seen that there are no flow transfers to a downstream link if the other downstream link is fully congested. That again highlights the FIFO nature of the diverge model. Again, the traffic states predicted by our nVT model fit the trajectories recorded from SUMO, illustrating that our framework is able to successfully propagate congestion across inter-corridor connections.

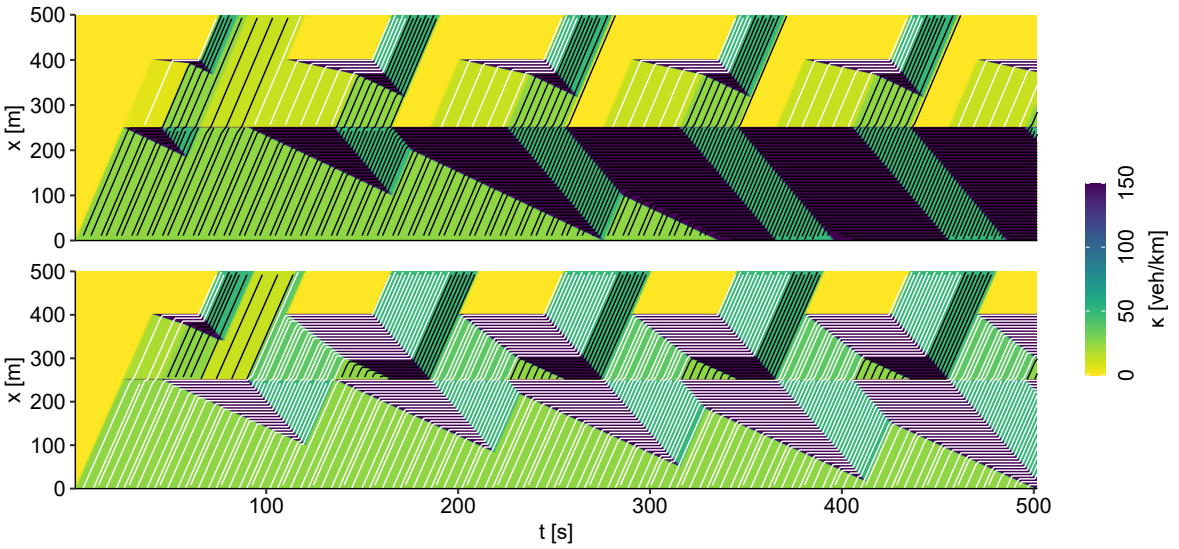


Figure 4.7: Congested scenario: Time-space diagrams for corridor C_3 (top) and corridor C_4 (bottom) with a traffic signal at I_2 ($x = 250\,\text{m}$). The colors represent the density which was obtained with the nVT model. The curves indicate trajectories obtained with the microscopic traffic simulator, and their color indicates the origin (i.e black trajectories originate in C_3 and white ones originate in C_4).

4.5 Inheritance of variational theory properties

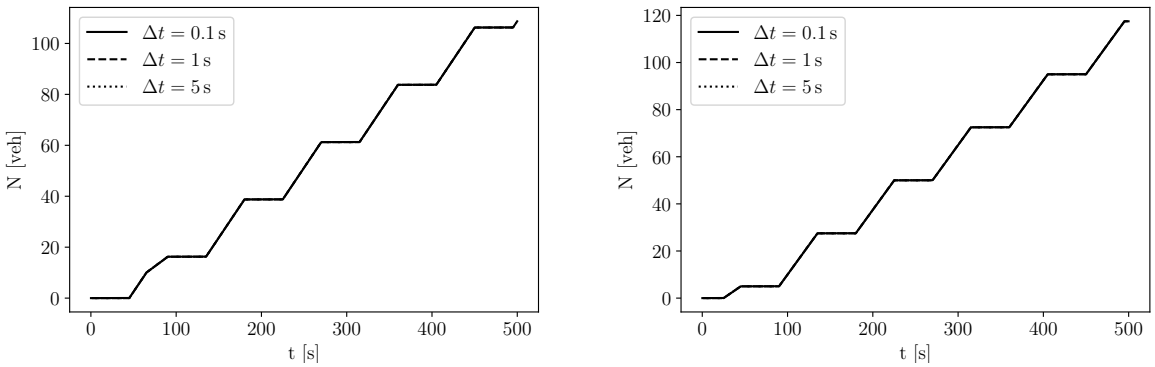
Our proposed framework extends the concept of VT to the network level by incorporating source terms. This section demonstrates that nVT inherits (i) the exact numerical calculation

of the Moskowitz function apart from potential sampling errors associated with VT, and (ii) the capability of modeling complex intra-link bottlenecks. The combination of these two properties makes nVT advantageous compared to other methods such as CTM and LTM.

4.5.1 Numerical error

In order to solve a given KWT problem with the nVT model, we have to define a multi-dimensional variational graph G. The choice of the time-step size Δt in G is crucial for the computational cost. The original VT has the advantage of being exact in determining the Moskowitz function, independently of the grid size. The only existing error is the so-called 'sampling error' according to DAGANZO & MENENDEZ [2005] that originates from an inaccurate sampling of data along the boundary.

The incorporation of discrete source terms at inter-corridor connections does not alter the fact that the framework is based on VT. It also inherits the exact numerical calculation of the Moskowitz function, as well as any potential sampling errors associated with VT. Our framework is able to consider inflows and outflows as long as the corresponding terms are mapped in the variational graph. This is not different from existing VT solutions, where bottlenecks are only considered when they are incorporated in G. Thus, nothing changes compared to the current VT techniques with regard to numerical errors. In order to illustrate this, we evaluate the congested scenario from the previous section (see Figure 4.5b) for the time-steps $\Delta t = \{0.1, 1, 5\}s$. Note that the same analysis could be conducted for any other scenario. We plot N at I_2 across t for both corridors.



(a) Moskowitz functions at I_2 on corridor C_3. **(b)** Moskowitz functions at I_2 on corridor C_4.

Figure 4.8: Impact of the time-step Δt on the results.

The y-axis displays the Moskowitz function N, the x-axis shows the time t in seconds. We plot the curves resulting from the different times-steps as solid, dashed, and dotted lines. However, comparing the numerical values of all Moskowitz functions $N_{\Delta t=0.1}$, $N_{\Delta t=1}$, and $N_{\Delta t=5}$ reveals that there is no difference between any of them. The figure illustrates this by showing the exact same shape with a perfect overlap of all curves. In other words, the Moskowitz function $N(x,t)$ is found independently of the time-step Δt, as long as all

bottlenecks and source terms are mapped in G. This confirms that our framework does not add any additional error to the original VT framework and thus is exact in determining N given a triangular FD and no sampling error at the boundary.

4.5.2 Capability of modeling complex intra-link bottlenecks

One main advantage of VT is the capability of modeling complex bottlenecks [see DAGANZO & MENENDEZ, 2005]. This section demonstrates, by means of an example, that despite the extension to networks by considering source terms at intersections, our framework is capable of modeling such complex bottlenecks.

We again apply our model to the network used in the previous section (see Figure 4.5b). Additionally, we randomly insert bottlenecks in space and time for both corridors. These could be interpreted as short blockings due to pedestrians crossing the road [KNOOP & DAGANZO, 2018]. In order to model such bottlenecks with nVT, we simply need to set the capacity $\beta = 0$ (see eq.(2.4)) where the blockings occur. The ability to represent any bottleneck by adapting the variational graph G allows for an easy and fast implementation of such.

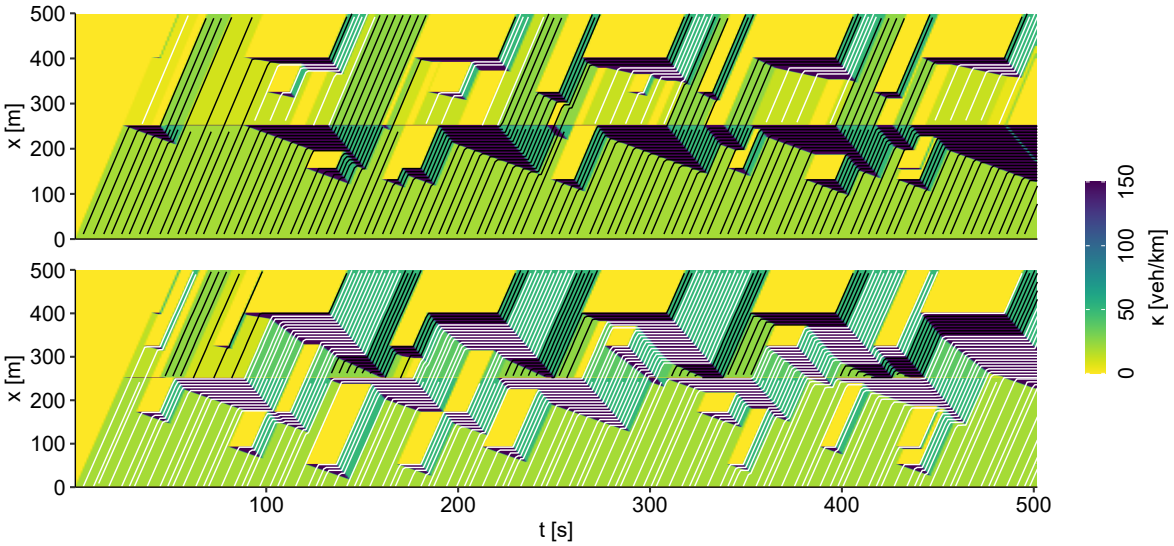


Figure 4.9: Random bottlenecks scenario: Time-space diagrams for corridor C_3 (top) and corridor C_4 (bottom) with a traffic signal at I_2 ($x = 250\,\text{m}$), and random bottlenecks. The colors represent the density which was obtained with the nVT model. The curves indicate trajectories obtained with the microscopic traffic simulator, and their color indicates the origin (i.e black trajectories originate in C_3 and white ones originate in C_4).

Figure 4.9 shows the resulting time-space diagram, where the axis and colors of the figure are the same as those of Figure 4.6. Again, the black curves depict trajectories originating

on corridor C_3, while the white ones correspond to those originating on C_4. These trajectory data are extracted from SUMO, in which we reproduced the scenario including the short bottlenecks to verify our nVT solution. Next to queues growing at intersections, the effects of short blockings are clearly visible, with many small queues appearing across the time-space region. In addition to those local effects of short queues, it can also be seen that the overall capacity of the corridor is impacted, and a larger queue in corridor C_4 downstream of intersection I_2 at $x = 250\,\text{m}$ appears, leading to a spillback. Such a spillback additionally affects the adjacent corridor C_3 and thus the entire network. This further confirms the importance of modeling such intra-link bottlenecks even on the network scale. Overall, the graph effectively demonstrates the effects of bottlenecks on traffic dynamics and clearly shows that the nVT solution fits the SUMO results very well. This illustrates that our model captures the effects of bottlenecks as well as a microscopic KWT simulation, even if such bottlenecks are inserted irregularly in time and space.

4.6 Application of the network variational theory to a realistic case

The previous section shows a very good fit of the trajectories from SUMO and the traffic states resulting from our proposed nVT model for scenarios with spillbacks and randomly occurring intra-link bottlenecks. In this section, we showcase the consistency between our nVT model and the KWT on a larger and more realistic network. For this purpose, we compare the predicted traffic states from nVT to a ground truth. Similar to the examples described in Section 4.4, we implement Newell's car-following model in SUMO to represent KWT at the network level. Note that this implementation and the associated assumptions are not related to a limitation of nVT, but are only needed to make the KWT solutions of both models comparable. This allows us to derive an appropriate ground truth for our case study and to show the consistency of nVT with KWT at the network level.

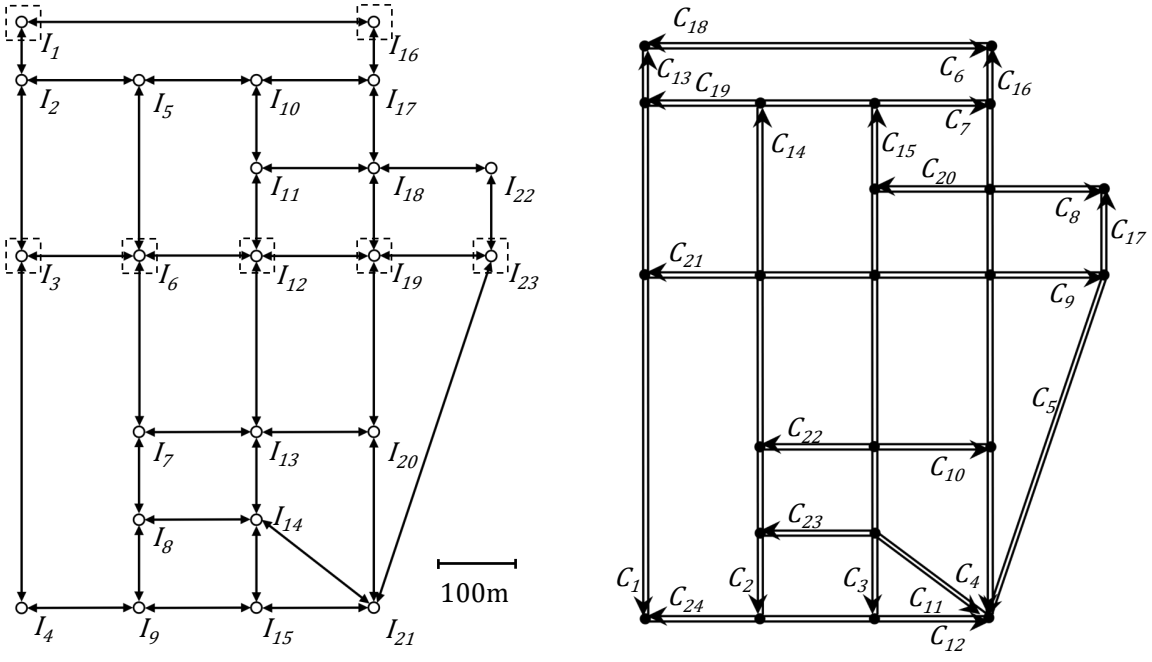
We create a network based on the well-known Sioux Falls example[1]. The solution of the KWT is the Moskowitz function $N(x, t)$. It allows us to derive all sorts of indicators, such as average flows, densities, and speeds, but also travel times and delays. Thus, we present our results as Moskowitz functions in this case study. We compare the results derived from our nVT framework to those obtained with SUMO. First, we measure the average flows per cycle according to both methods. This is a suitable aggregation for minimizing the stochastic aspects of SUMO. Second, we show the Moskowitz functions for two locations in the network for the purpose of illustration. Third, we analyze the computational effort of both methods, as the associated cost for macroscopic models is usually low and thus advantageous for numerous applications.

[1] derived from: https://github.com/bstabler/TransportationNetworks

4.6.1 Case study design

The nature of our framework and that of the microscopic simulation are fundamentally different. The former is macroscopic, and traffic can be deterministically assigned based on origin flows and turning ratios. Moreover, intersections are modeled as points in space. The latter includes a car-following model, as well as routing aspects next to origin flow and turning ratio definitions. In particular, SUMO allows specifying turning ratios at intersections which are then approximated by an internal route-swapping optimization method. This method aims to find a set of individual vehicle routes which approximates the specified turning ratios. However, deviations from the specified turning ratios still occur. Also, the geometrical layout of intersections in SUMO deviates from the representation of intersections within the nVT framework, in which intersections are modeled as points in space with no physical dimension. Having these differences between the two models in mind, we apply several simplifications, described in the following, to be able to conduct a reasonable comparison.

Figure 4.10a presents the network for which the case study is performed. For the sake of simplicity, we only consider intersections with four or fewer legs and thus reduce the original Sioux Falls network slightly. In total, the network consists of 23 intersections connected by 36 bi-directional links, as highlighted by the arrows. All intersections are signalized with a cycle time of 90 s, and a green and red phase of 45 s. All offsets are set to zero.



(a) Original network.

(b) Decomposed network and corresponding set of corridors.

Figure 4.10: Sioux Falls network for the nVT case study.

In order to apply our nVT model, origin nodes and inflows, as well as turning ratios at inter-corridor connections are required. For our case study, we define the nodes with less

than four legs as origins. At these nodes, virtual links are created from which the flows enter the network. Additionally, we define turning ratios for seven intersections, i.e. seven inter-corridor connections. Note that the number of inter-corridor connections is not limited within nVT in general. This number is chosen to further increase the interpretability of the observed phenomena in the case study as it limits the impact of vehicle routing in SUMO; thereby increasing the comparability of both methods. At the same time, these inter-corridor connections ensure the occurrence of network traffic phenomena such as circular route dependencies in the presented case study. For the sake of simplicity, turning ratios are set to $\alpha = 0.5$ for all seven inter-corridor connections. Nevertheless, any other value between 0 and 1 is feasible. The inter-corridor connections are highlighted as dashed squares in Figure 4.10. Note that by defining origin nodes and turning ratios, no destination nodes need to be set explicitly. We evaluate the case study for five different demand scenarios with volume-to-capacity ratios of $VOC \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ where the capacity again refers to that of intersections. The simulation period is one hour. We choose a time-step of $\Delta t = 0.1$ s, and an FD with the same parameters as described in the previous section.

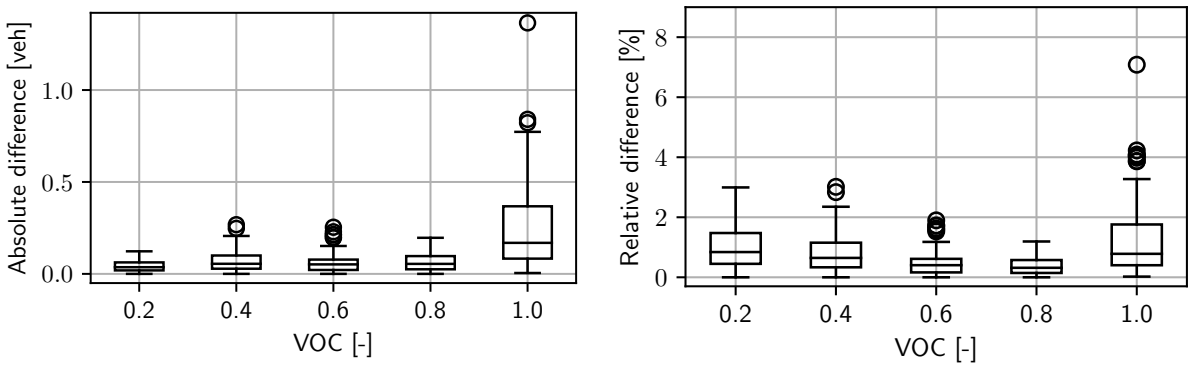
Furthermore, as described in step 1 in Section 4.3.1, we decompose the network into corridors based on the horizontal and vertical orientation of links. In total we define 24 corridors which are displayed in Figure 4.10b. Note that this network specification includes circular route dependencies for the corridors C_1 - C_9 - C_{16} - C_{18}, as well as in the opposite direction C_{13} - C_6 - C_4 - C_{21}. Recall that the method of decomposition does not affect the final result of nVT.

4.6.2 Results and discussion

Moskowitz functions

To evaluate the results of our nVT model, we extract the Moskowitz function at the upstream end of each link and compare it to the Moskowitz function derived from the SUMO output. Subsequently, we measure the average flow per cycle and calculate the differences between the results from nVT and SUMO. This results in a total of 86 values for each demand scenario (one per lane, including virtual links). Figure 4.11 shows the box-and-whisker diagrams for these differences related to each demand scenario.

On the left (Figure 4.11a) we show the absolute difference, and on the right (Figure 4.11b) the relative difference. The latter corresponds to the absolute difference divided by the average number of discharging vehicles per cycle according to SUMO. The absolute and relative errors are displayed on the y-axis, whereas the x-axis corresponds to the volume-to-capacity ratio VOC. For most cases, we observe very low differences. The relative difference for scenarios with a low VOC is marginally higher, simply because of the low flows per cycle. The exception is the last scenario that has the highest relative difference. However, these statistical outliers occur due to the high influence of stochasticity and routing complexity when the network is saturated. Therefore, this outcome is expected, as a larger number of vehicles leads to more stochasticity and a higher routing complexity in SUMO. This consequently increases the differences between the SUMO and the nVT results. Nonetheless, most of the



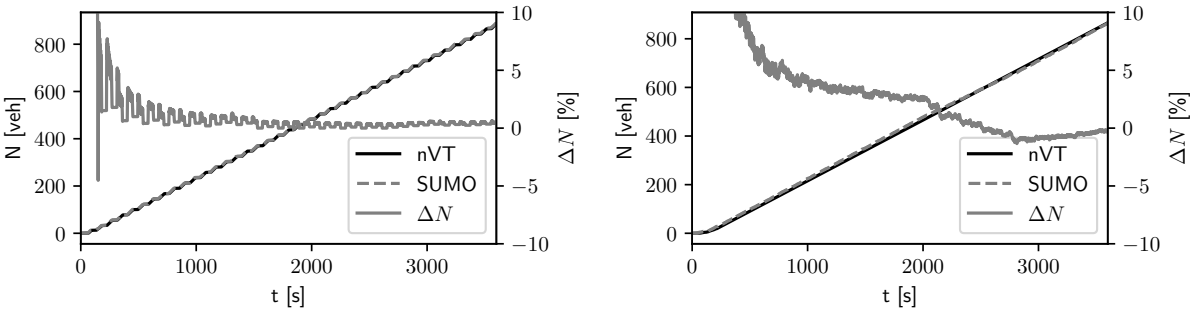
(a) Absolute differences.

(b) Relative differences.

Figure 4.11: Differences of cycle-based counts between nVT and SUMO for all demand scenarios.

observed relative differences are below 3 %. Additionally, the median relative differences are all below 1 %, which corresponds to median absolute differences below 0.2 vehicles counted per cycle for all scenarios. Given the numerical errors, the above mentioned stochasticity, the occurring circular route dependencies, and the different nature of the microscopic and macroscopic modeling approaches, we adjudge these differences as small.

In addition to these box-and-whisker diagrams, we present two Moskowitz functions in Figure 4.12 for illustration purposes. We display the Moskowitz functions N from our nVT approach as solid black lines, and SUMO as dashed grey lines for the intersection I_3 (Figure 4.12a), and the inter-corridor connection I_{18} (Figure 4.12b). Additionally, we plot the relative difference of both cumulative curves as solid grey lines.



(a) Moskowitz functions at intersection I_3.

(b) Moskowitz functions at inter-corridor connection I_{18}.

Figure 4.12: Moskowitz functions at an intersection and an inter-corridor connections from nVT and SUMO.

The left y-axis shows the Moskowitz functions N in vehicles downstream of the intersections, while the right y-axis shows the relative difference ΔN in percent. The x-axis displays the simulation time in seconds. Figure 4.12a corresponds to an intersection, where there are

no turning flows, i.e. $\alpha = 0$. One can observe the characteristic step-wise increase due to the different signal phases. The relative difference is very high in the beginning but soon converges to a value close to 0 %. The high values for low t are due to the fact that the difference is related to N from SUMO, and these values are very low at the beginning of the simulation. Moreover, stochastic influences appear to be large for short simulation durations. However, after a reasonable simulation period (i.e. warm-up period), the results from SUMO and nVT match each other very well. Figure 4.12b is an example for an inter-corridor connection. Due to the turning ratios $\alpha = 0.5$, a constant flow occurs. Again, both curves match each other very well for longer simulation periods, and the error converges to 0 % at the end of the simulation period. Note that the shape of the cumulative curves corresponding to nVT perfectly matches the expected step-wise and linear shape. Thus, the main reasons for the existing errors presumably result from the aforementioned fundamental differences between SUMO and nVT. An in-depth sensitivity analysis [e.g. GE, CIUFFO & MENENDEZ, 2015] could be conducted to further explore these differences paying special attention to the relation between the different inputs and the outputs of the SUMO simulations, but that is considered out of scope for this chapter.

Computational effort

Compared to microscopic simulations, one great advantage of our method is its computational efficiency. This results from the macroscopic nature of the model. We show a comparison of the computational effort of both methods for the presented Sioux Falls case study in Table 4.1. The methods were evaluated on a computer with an Intel(R) Xeon(R) W-2145 CPU with 3.70 GHz and 64 GB RAM.

Table 4.1: Computational cost comparison of SUMO and nVT

VOC [-]	Computation time [s]			
	SUMO	nVT ($\Delta t = 0.1$ s)	nVT ($\Delta t = 1$ s)	nVT ($\Delta t = 5$ s)
0.2	33.8	194.0	13.1	2.3
0.4	54.9			
0.6	88.5			
0.8	106.1			
1.0	188.6			

The demand scenarios are specified in the left-most column as a function of the VOC. We see that the computational cost in seconds for SUMO scales with the increasing demand. Since each vehicle is simulated separately it is logical that a growing number of vehicles leads to an increased computational effort. In contrast, this is not the case for nVT since the computational cost scales only with the multi-dimensional variational graph G. The computational cost does not scale with demand, i.e. the actual number of vehicles simulated, but only with the number of points in G. The results in the table show the effect of an increased time-step Δt, which leads to a reduced number of points in G, and therefore to a

drastic decrease of computational effort required. While $\Delta t = 0.1$ s leads to a high cost, the choice of $\Delta t = 5$ s results in a computational cost of nearly two orders of magnitude lower than the highest one of SUMO. In fact, the absolute difference of three minutes corresponds to roughly 600 % increase of computation time from a relative perspective. Such differences become important for use cases where a large number of repetitions of the traffic model are required, e.g. simulation-based optimization or model calibration [e.g OSORIO & BIERLAIRE, 2009; TILG, YANG & MENENDEZ, 2018; TILG ET AL., 2020b; AMELI, LEBACQUE & LECLERCQ, 2020; GE, CIUFFO & MENENDEZ, 2014]. At the same time, no numerical error is introduced by the larger time-step as indicated in Section 4.5.1. The appropriate choice of the time-step makes our proposed nVT model suitable for real-time applications or as part of a model-based optimization framework.

Overall, this case study clearly shows the applicability of our nVT model at the network level. The results match those of a microscopic KWT simulation for several demand scenarios very well. Moreover, the nVT framework can evaluate network-wide traffic dynamics at a low computational cost. We conclude that our nVT model represents the first step towards a fully VT-based KWT simulation applicable to large-scale networks.

4.7 Conclusion

This chapter proposes an extension of the Eulerian VT [DAGANZO, 2005b; DAGANZO, 2005a; DAGANZO & MENENDEZ, 2005] for the application at the network level. First, the network of interest is decomposed into a set of non-overlapping corridors. Second, we introduce a mathematical framework to account for non-zero inflows and outflows at intersections within VT. By considering the discontinuities in the Moskowitz function that occur due to such inflows and outflows, we account for potential violations of the flow conservation on such corridors. Our nVT model is able to propagate free-flow and congested traffic states across intersections and therefore throughout the entire network. We successfully verify our framework for a simple network with the microscopic simulation SUMO. Additionally, we show that our extension inherits the numerical exact calculation of the Moskowitz function for triangular FDs from the original VT, as well as the capability to model complex intra-link heterogeneities. Finally, we apply our framework to a case study for the Sioux Falls network. We again compare the results to the solution derived from SUMO. The traffic states predicted by our method clearly show a good fit to the ones from SUMO, with median relative differences in cycle-based flows of less than 1% in all considered demand scenarios. Moreover, our results indicate the low computational cost which can be achieved with our VT extension.

In conclusion, we successfully extended the Eulerian VT for signalized urban networks. Thus, KWT problems can be solved at the network level based on the concept of VT. Therefore, we can account for heterogeneities such as time-space dependent FD, moving bottlenecks, and randomly occurring short blockings. The brief literature review showed the wide range of applications of VT, some of which may profit from the possibility to apply VT at the network level as well. The developed model applies to use cases where a microscopic simulation

is computationally too expensive, but the existence of intra-link bottlenecks such as buses or pedestrians hinders the efficient application of other macroscopic models such as CTM and LTM. Moreover, our proposed nVT model is an important step to provide the grounds for existing VT applications at the network level, such as the MFD approximation.

From a methodological perspective, the implementation of merge schemes seems to be the next logical step, e.g. Daganzo's merge model [DAGANZO, 1995]. Moreover, our model is a first step in the direction of semi-analytical network MFD approximation. Current methodologies such as the MC apply only to single corridors, as they are based on the original VT. This topic will be further explored in the next chapter.

Chapter 5

Network method of cuts

This chapter is based on the following works:

- G. TILG, L. AMBÜHL, S. BATISTA, M. MENENDEZ, L. LECLERCQ & F. BUSCH [2021c]: Semi-analytical estimation of macroscopic fundamental diagrams: From corridors to networks. In: *Presented at the 100th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA
- G. TILG, L. AMBÜHL, S. F. BATISTA, M. MENENDEZ, L. LECLERCQ & F. BUSCH [n.d.]: From corridor to network macroscopic fundamental diagrams: A semi-analytical estimation approach. Submitted for publication.

5.1 Introduction

In Chapter 1, we highlighted the wide range of MFD-based applications. Obviously, these require the estimation of the MFD to describe the evolution of aggregated traffic states. In the absence of real data, or for hypothetical scenario analyses, one solution is to use microscopic simulations. Unfortunately, these have high computational costs, require labor-some calibration, and their results remain sensitive to the specific demand patterns. Alternatively, analytical solutions exist for estimating the MFD. For example, AMBÜHL ET AL. [2020] developed such an analytical model. However, limited or unavailable a-priori knowledge limits these simple estimations with the proposed functional form. Another possibility is to use semi-analytical approaches, such as the MC and the SA which were thoroughly analyzed in Chapter 3. So far, however, these methods only explicitly model single corridors. Hence, their application is limited to small and regular synthetic urban networks, where the difference between the MFD of an abstract corridor and that of the whole network is minor. Hitherto, the literature lacks a comprehensive framework to semi-analytically estimate MFDs for realistic, complex urban networks. Below, we highlight some major limitations of the corridor simplification.

The topological mapping of a network into a single corridor is not necessarily accurate. Existing approaches assume that control and network topology are relatively similar and regular throughout the network (i.e. similar offsets, green-to-cycle ratios, block length, etc.). Thus, the network average features can be mapped into a single corridor. As explained in

Chapter 1, the idealized MFD refers to a theoretical upper bound for traffic conditions for such a corridor [DAGANZO, 2007]. However, the topological features of the network are lost, leading to inaccuracies in the approximated MFD. For example, the capacity of a single corridor is typically constrained by the most restrictive intersection. This might not be true for general networks, since in many cases multiple routes connect the same OD pair. In addition to these deficiencies from a supply perspective, the aggregation of a network to one corridor also simplifies demand-related aspects. For example, it implies that traffic dynamics evolve in a similar manner on all corridors in the network. This is a strong assumption, as it rarely occurs in urban networks with a large number of OD pairs which lead to very complex demand patterns. Furthermore, diverging flows at intersections inevitably lead to undersaturated conditions at downstream links. Neglecting this contradicts the network-wide flow conservation. Lastly, the effects of spillbacks may propagate on adjacent corridors. Thus, the need for an efficient approximation of the network MFD considering these aspects is obvious.

We should also note that in many cases, the observed macroscopic performance of an urban network is substantially lower than the theoretically derived idealized MFD. The reason for this lies in the observed traffic heterogeneity, non-stationary traffic states, and the path-flow distribution of the demand [MAHMASSANI, SABERI & ZOCKAIE, 2013; KNOOP ET AL., 2015; LECLERCQ, PARZANI, KNOOP, AMOURETTE & HOOGENDOORN, 2015; GEROLIMINIS & SUN, 2011b; MAZLOUMIAN ET AL., 2010]. Recall that these observed aggregated traffic states are herein denoted as the realized MFD, in contrast to the idealized MFD [AMBÜHL ET AL., 2021; LODER ET AL., 2019b]. Unfortunately, the two different MFD notions are currently used interchangeably. Both MFD types are very similar for a corridor, under steady-state conditions, and homogeneous demand. However, they might differ significantly for general networks. Simply put, the idealized MFD corresponds to stationary flows resulting from a spatial demand pattern that perfectly utilizes the given supply. Thus, the ambiguity of results stemming from models that simplify the network to a corridor, as required by existing semi-analytical approaches, and neglecting the network-specific influence of demand-related aspects, becomes apparent. First approaches to tackle these issues are reported in GEROLIMINIS & BOYACI [2012] and XU ET AL. [2020], which incorporate turning flows at intersections. The former study analyzes these effects on a simulation basis and focuses on incoming turns. The second one models the effects merely stochastically. Neither of them describes network-wide spillbacks nor network conservation explicitly, which is important to identify the underlying mechanisms for congestion dynamics. Only such understanding enables the proper design of measures to mitigate congestion and increase urban network capacities. The proposed methodology addresses this by considering the effects of turning ratios on the network MFD. The realized MFD is especially important because it replicates observed traffic states based on which further applications can be designed (e.g. traffic control, transport planning, analysis of local bottleneck effects at the network level). Additionally, as we will discuss later, it could potentially be extended to also estimate the idealized MFD, although this is left for future research.

In this chapter, we propose a framework using semi-analytical methods to estimate the realized MFD at the network level. Thereby, we do not rely on extensive empirical traffic data

sets or expensive microscopic simulations. In contrast, we develop a framework based on semi-analytical methods but do not reduce the network to a corridor. We focus on a single region and explicitly consider flow conservation as well as the effects of spillbacks at the network level. Thus, we can account for the effects of different demand patterns on the realized MFD's shape. Hereafter, we refer with 'MFD' to the realized one unless explicitly stated otherwise. The contributions are threefold:

1. We develop a general framework to semi-analytically estimate the MFD at the network level. So far, existing methods were not able to account for spatial demand patterns and the related violation of the conservation of flows at intersections from the corridor perspective. Our framework proposes a way to address this challenge. We decompose a given network into a set of corridors and construct a hypernetwork in the spirit of the original MC. The decomposition maintains the connection of different corridors and, thus, turning relations are actively considered. These inter-corridor connections enable us to model the inter-dependencies of traffic dynamics between corridors in the network, e.g. the network-wide effects of spillbacks. This is a major difference to current approaches, where multiple corridors are either reduced to a single one or modeled as being largely independent of each other, which may lead to inaccurate traffic state estimations.

2. Considering the effects of turning relations, we propose three different approaches to derive the maximum flows at intersections, and two different approaches to approximate the network-wide jam density. Hereby, the challenge is to model queue propagation while accounting for traffic state interdependencies. To our best knowledge, no method exists so far to approximate the network-wide jam density. With the maximum flows at intersections and the network-wide jam density, the MFD is then estimated following the philosophy of the original MC. The proposed approaches vary in terms of modeling complexity, computational cost, and approximation accuracy. Thus, they offer high flexibility and enable the choice of a case-specific MFD approximation. Moreover, they shed light on the effects of related assumptions and thereby provide additional explanatory value.

3. We evaluate the proposed framework using a realistic network. Within this case study, we compare the proposed approaches to the state of the art for such networks and the CTM [DAGANZO, 1992; DAGANZO, 1995] acting as a ground truth. Thereby, we provide proof of our concept and test the framework for a range of different input parameters.

The remainder of this chapter is organized as follows. As the background on VT and MC was introduced in Chapter 2, we present the general methodology for our framework. Then, we explain the specific steps of this framework in more detail. Subsequently, we conduct a case study for the realistic network of Sioux Falls. Finally, we draw conclusions and outline potential avenues for future work.

5.2 General methodology for the extended method of cuts

In an attempt to overcome the mentioned drawbacks of existing methods and to address this research gap, we extend the MC in order to apply it to realistic networks with varying turning flows at each intersection. Thereby, we include network-wide traffic dynamics and do not reduce the entire network to a single corridor. Thus, we consider its original topology. This results in a general framework to approximate the realized MFD for realistic urban networks. Hereafter, we refer to this framework as 'nMC'.

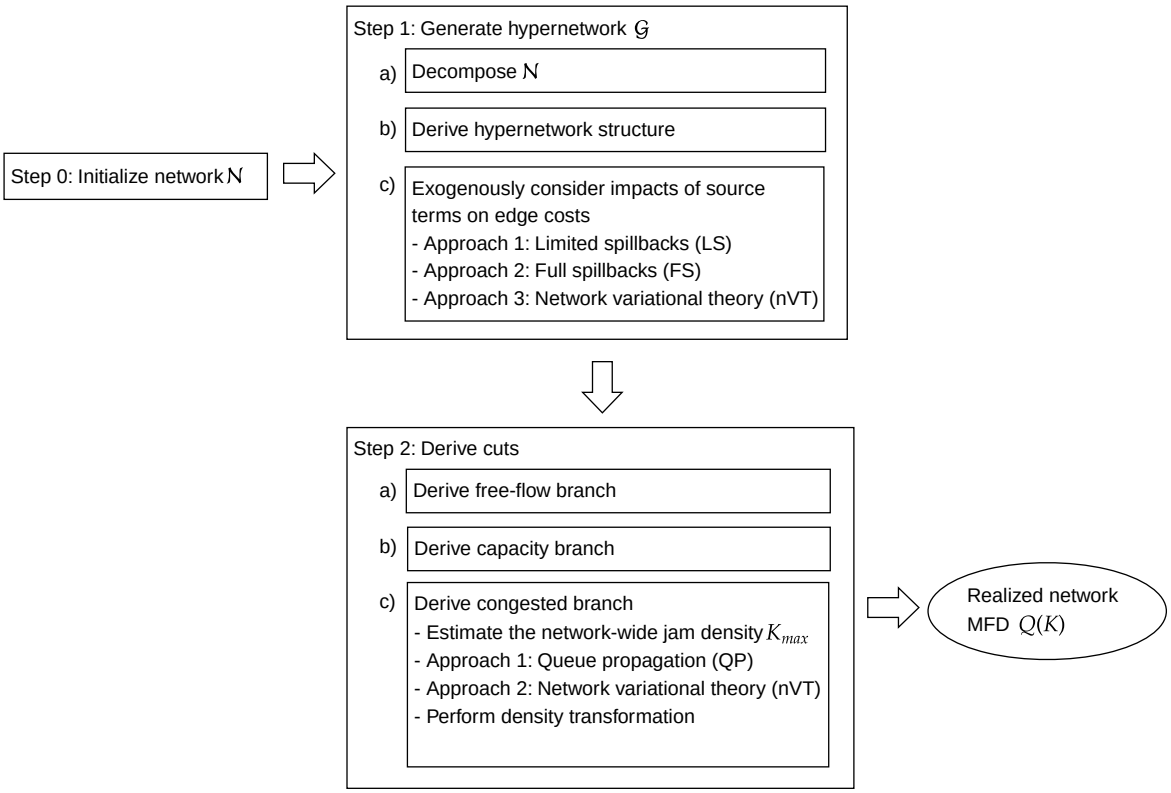


Figure 5.1: Flowchart of the proposed framework, the nMC.

Figure 5.1 describes the overall framework which consists of three steps which are explained in the following:

- *Step 0* initializes the problem by defining supply and required demand characteristics. This consists of the definition of the network topology and the signal control settings. Moreover, we define the turning ratios at each intersection. Note that this implies an indirect definition of routes, and thus reflects the observed spatial demand pattern for a specific case. These parameters describe the physical road network $\mathcal{N}$ composed by a set of directed links and intersections, as well as the demand that exists for that network.
- *Step 1* transforms $\mathcal{N}$ into a hypernetwork $\mathcal{G}$ which is the basis for deriving the MFD.

Recall that a hyperlink represents a physical corridor in the original MC (see Figure 2.5a in Chapter 2). Analogously, a hypernetwork represents the physical road network in time and space and can be seen as a multi-dimensional variational graph. To generate $\mathcal{G}$, we decompose $\mathcal{N}$ into a set of corridors, define the structure of the hypernetwork $\mathcal{G}$, and exogenously derive the costs for horizontal edges in $\mathcal{G}$. For this, we propose three different approaches with varying modeling complexity, computational cost, and approximation accuracy. These approaches are labeled as limited spillback (LS), full spillback (FS), and nVT, and are further explained in Section 5.3.

- *Step 2* derives the cuts based on which the free-flow branch, the capacity, as well as the congested branch of the MFD are estimated. In the spirit of the original MC (see Figure 2.5b in Chapter 2), we send moving observers through $\mathcal{G}$ in order to estimate stationary traffic states. This step includes the approximation of the network-wide jam density and the utilization of symmetries in the propagation of free-flow and congested traffic states across the network. Again, two different approaches are proposed, namely the queue propagation (QP) and the nVT approach, which are explained in Section 5.4.3.

The nMC enables one to estimate the realized MFD for realistic urban road networks. It accounts for turning flows and the corresponding interdependencies of traffic states across the physical road network. Compared to simulation-based approaches, this procedure describes the observed macroscopic traffic behavior in urban networks in a fast and efficient manner. We propose several approaches to (i) allow one to flexibly trade-off modeling complexity and computational burden against approximation accuracy depending on the use case at hand, and (ii) shed light on the impact of some of the modeling assumptions. Below we describe steps 1 and 2 mentioned above in more detail. Step 0 is trivial as it refers to the problem initialization and requires no further explanation.

5.3 Generation of a hypernetwork

The correct representation of $\mathcal{N}$ as a hypernetwork $\mathcal{G}$ is key to our framework. It allows the derivation of the cuts similar to the original MC (see Section 2.3) while accounting for source terms and network-wide spillback propagation. These cuts are the basis to find the realized network MFD. In the following, we list the main assumptions of our framework:

- The focus lies on unimodal networks. That is, we do not consider any modes apart from private vehicles. Nevertheless, there are multimodal extensions for the original MC which can potentially be integrated into our framework. We consider this, however, out of scope and leave it for future work.
- We only consider signalized intersections without modeling conflicting streams explicitly. However, conflicting traffic streams can be approximated by reducing the average capacity of corresponding intersection approaches accordingly, e.g. based on headway distributions [e.g. HERZ ET AL., 1976]. Related to that, other intersection types such as

4-way stops can be included if capacity estimates for the corresponding intersection approaches exist.

- We assume that vehicles follow a FIFO discipline on all links and at diverges [NEWELL, 1993]. Turning lanes can be introduced by splitting a link, and defining the turning ratios at the downstream end of each lane accordingly. For example, the turning ratio at the downstream end of a right-turning lane would be 1, while it would be 0 at the downstream end of the straight-going lane.
- For the sake of simplicity, our framework only applies to cases where vehicles at intersections can either remain on the main corridor or change to a single adjacent corridor. The possibility of describing turning flows between more than a pair of corridors is left for future work. While the general framework is able to account for multiple turning options since the basic functionality requires no substantial modification, it would at least necessitate the specification of turning ratios for each option and the coupling of more than two corridors at intersections.
- While we allow turning ratios to vary between intersections, we assume them to be constant across time.
- We assume that stationary states exist for any spatial demand pattern if the temporal change is slow. Note that this is an assumption inherited from the original MC and further applies to the general notion of the MFD. However, we do not explicitly specify an initial state in the network. Instead, we solve for boundary conditions which result in a maximum network-wide average flow. Thus, we implicitly assume the existence of an initial state which leads to the capacity state of the network.

We generate a hypernetwork $\mathcal{G}$ representing a physical road network $\mathcal{N}$ including signal settings and turning flows. First, $\mathcal{N}$ is decomposed into a set of corridors $\mathcal{C}$. Second, the structure of a hypernetwork $\mathcal{G}$ is defined. As LAVAL ET AL. [2016] showed, VT cannot endogenously account for source terms. Consequently, they are considered as an exogenous input as the overall framework is based on the concept of VT. Note that only the turning ratio is an exogenous input. The resulting costs of horizontal edges in the hypernetwork, which represent maximum flows, are endogenously derived depending on the method applied, i.e. LS, FS, or nVT. This concludes the generation of the hypernetwork $\mathcal{G}$. Note that the resulting maximum flows in $\mathcal{G}$ represent the network traffic at the capacity level.

5.3.1 Network decomposition

The decomposition of the physical road network $\mathcal{N}$ into a set of non-overlapping corridors facilitates the generation of $\mathcal{G}$. Furthermore, this set of corridors is an input to the MFD approximation (see Section 5.4.1). Chapter 4 describes such a decomposition in the context of VT. Following this methodology is advantageous for the framework here, as the proposed nVT can be utilized to derive the hypernetwork (see Section 5.3.3). We provide a brief summary of the decomposition method hereafter.

The initialization of the problem (Step 0) defines a physical road network $\mathcal{N}$ consisting of

intersections $I \in \mathcal{I}$ and links $L \in \mathcal{L}$. Moreover, it includes the specification of signal control settings, i.e. red r and green times g, cycle lengths c, offsets o, as well as turning ratios $\alpha \in A$ at each intersection I. All information from $\mathcal{N}$ including topological features, turning ratios α, and control settings for $I \in \mathcal{I}$ have to be retained when the network is decomposed (Step 1). Note that the turning ratios are an exogenous input. The two main requirements for decomposing the network $\mathcal{N}$ are:

1. The set $\mathcal{C}$ includes all links $L \in \mathcal{L}$.
2. Each link L exists only once in $\mathcal{C}$, i.e. the corridors do not have any overlapping segments.

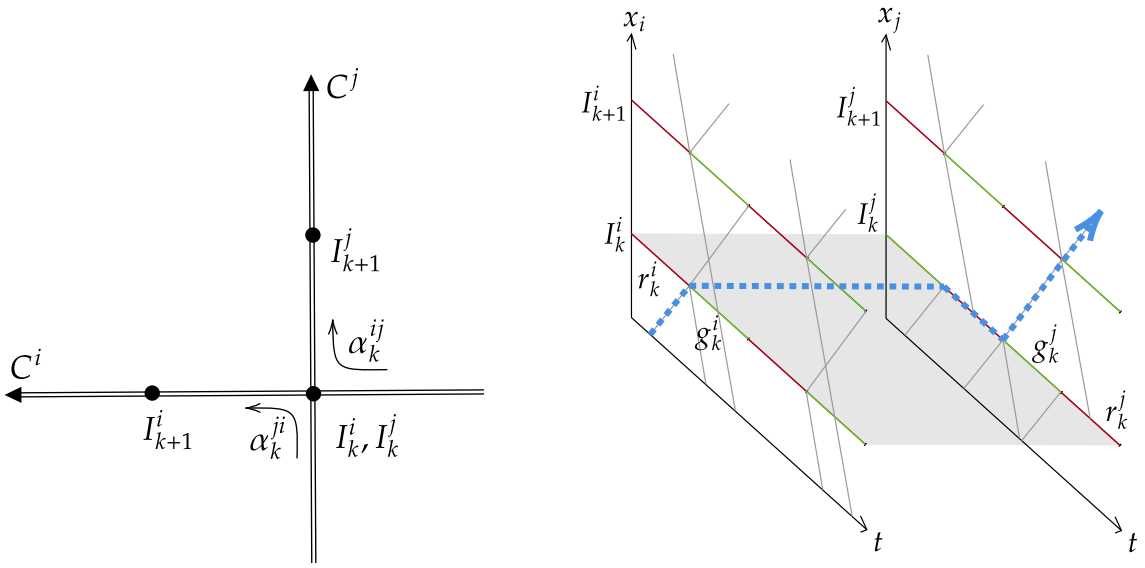
For small toy networks, we can determine the set $\mathcal{C}$ manually from $\mathcal{N}$. More general, realistic networks can be decomposed according to the actual layout of the roads. This includes arterials, avenues, and streets. Each road can be represented as a corridor C and as such be incorporated in $\mathcal{G}$. This will always satisfy both conditions mentioned above. Requirement 2 does not preclude the potential existence of circular routes in the network (see Chapter 4 for more details). Such interdependencies across links are essential to realistically model the propagation of traffic congestion. Note that the set $\mathcal{C}$ might affect the MFD to a certain extent. This aspect will be treated in Section 5.4.1 and analyzed in Section 5.5.2.

5.3.2 Structure of the hypernetwork

The set of corridors $\mathcal{C}$ allows us to generate a hypernetwork $\mathcal{G}$ representing the maximum traffic flows in the entire network $\mathcal{N}$. This includes the generation of hyperlinks for each corridor along with intersections and corresponding turning ratios α according to the physical network topology. Hence, in contrast to the original MC where only single corridors could be modeled, $\mathcal{G}$ represents networks where multiple edges downstream and upstream of an intersection might exist. This expansion of the concept of the corridor-specific hyperlink to a network-wide hypernetwork $\mathcal{G}$ represents one of the major cornerstones of our framework.

Figure 5.2a illustrates an example of a decomposed network based on which we further introduce our notation. It shows two corridors C connected at one intersection I. The indices denoted as superscripts i, j refer to a specific corridor. The index k as subscript refers to an intersection and the link directly upstream of it. The subscript is unique in combination with the corridor-related index. For example, the intersection I_k^i is the k^{th} intersection on corridor C^i. Similar to that, subscripts of maximum flows $\tilde{q}$ indicate intersections, and superscripts the corridor the flow refers to. For example, the variable $\tilde{q}_k^i$ describes the maximum cycle-based average flow on corridor C^i upstream of intersection I_k^i. By maximum flow we do not refer to the capacity, but the maximum flow that can be sustained on average during one cycle as a function of both the demand and the supply. The indices of turning ratios α denote the origin and destination corridor of the corresponding flows at a given intersection. For example, the turning ratio α_k^{ij} describes the ratio of the flow aiming to change from C^i to C^j at intersection $I_k^i = I_k^j$.

The network $\mathcal{N}$ can be translated into a hypernetwork $\mathcal{G}$, see Figure 5.2b. For a single



(a) Physical road network $\mathcal{N}$.

(b) Hypernetwork $\mathcal{G}$.

Figure 5.2: Example of a network with two corridors, and a corresponding hypernetwork excerpt.

corridor, the graph $\mathcal{G}$ consists of horizontal and slanted edges, the latter with slopes equal to u or w. The duration of green and red phases is denoted with g and r, respectively. Subscripts denote the intersection these phases refer to, while superscripts refer to the corresponding corridors. For example, g^i_k is the duration of the green phase at intersection I^i_k on corridor C^i. The horizontal edges have associated costs z equal to the maximum flow $\tilde{q}$ passing a moving observer traveling along the respective edge. In the original MC, this flow is zero during red times, i.e. $\tilde{q} = 0$, and equal to the link capacity during green times, i.e. $\tilde{q} = q_{opt}$ (see Section 2.3). However, the occurrence of inflows and outflows at intersections might affect the observed maximum flow, and consequently the related costs z. For example, the maximum flow $\tilde{q}^i_k$ during a green phase at an intersection I^i_k might be reduced when a net outflow occurs at the upstream intersection I^i_{k-1}. Also, if the inflow at an intersection I^i_k exceeds the capacity of the downstream intersection I^i_{k+1}, spillbacks occur after a while, potentially affecting $\tilde{q}^i_k$.

These examples highlight the importance of the inclusion of effects of turning flows and indicate that such impacts can be represented via $\tilde{q}$. Thus, we aim at modifying the costs z of horizontal edges to represent the effects of source terms. As explained before, these effects need to be considered exogenously. Therefore, we propose three different methods to exogenously determine the effects of source terms on the maximum flows $\tilde{q}$ at intersections, and adapt the costs of related edges in $\mathcal{G}$.

5.3.3 Exogenous consideration of source terms

To complete the generation of the hypernetwork $\mathcal{G}$, we exogenously account for the effects of source terms at intersections which represent inflows and outflows. In particular, we modify the costs of horizontal edges in $\mathcal{G}$ representing green phases. Thereby, we include both demand- and supply-related aspects, such as undersaturated intersection approaches and the occurrence of spillbacks coming from the downstream intersection. Note that while the edge costs in $\mathcal{G}$ shall represent maximum flows, similar to the original MC, the hypernetwork allows deriving an upper bound for flows during the loading and unloading phase of the network as well. For this purpose, we present three different approaches of increasing modeling complexity and computational cost in exchange for higher estimation accuracy:

1. *Approach 1 - Limited spillbacks (LS)*: We do consider spillback propagation, but only within links. The capacity reduction from downstream intersections will affect the upstream ones but will not further propagate into the network. Therefore, circular dependencies cannot be considered with this approach. The field of application of this approach is for use cases where no substantial network-wide spillbacks occur.

2. *Approach 2 - Full spillbacks (FS)*: This approach builds on the first one, but it also includes spillback propagation throughout the network and thus related circular dependencies. Therefore, it also applies to use cases with substantial network-wide spillbacks.

3. *Approach 3 - Network variational theory (nVT)*: Based on the nVT proposed in Chapter 4, this approach allows to accurately consider spillback propagation including circular dependencies at the network scale according to KWT. While being more exact and offering a larger field of applications, this approach requires significantly more modeling and computational efforts than the other two.

In the following subsections, we first introduce the mathematical framework for the approximate approaches FS and LS. Then, we describe the utilization of nVT to complete the hypernetwork, and highlight the synergies between nVT and the proposed nMC.

Approach 1 - Limited spillbacks (LS)

This approach applies to use cases where a fast and simple MFD approximation is essential, and spillbacks are limited to successive intersections. For such use cases, we propose a solution method that provides a reasonably accurate estimation of spillback effects at the link level to derive the costs z for edges representing green phases in $\mathcal{G}$. The approximation of such effects is based on analytical calculations, making it fast and efficient.

This section formulates an optimization problem with non-linear constraints considering the inter-dependencies of maximum flows $\tilde{q}$ between different intersections $I \in \mathcal{I}$ in the network $\mathcal{N}$. First, we consider demand-related aspects and the network-wide conservation of flows. Second, we account for supply reductions due to spillbacks.

Assumptions. For the sake of clarity, we list the most important assumptions for this approach below:

- The link FD is triangular.
- Traffic signals at all intersections I have a common cycle length.
- We only consider the effects of congestion at the link level and no spillback propagation across intersections.

Demand. The demand-induced reductions of maximum flows $\tilde{q}_{k,d}$ due to source terms need to be represented in $\mathcal{G}$. This is of particular importance for the approximation of the capacity branch of the MFD. The index d indicates that this maximum flow is constrained by the demand. Note that such effects were implicitly included in the original MC, e.g. undersaturated intersections due to upstream bottlenecks. However, source terms eliminate the possibility for such an implicit consideration. The following example lets us derive the key formulation to represent demand-related effects in our model.

Consider a case where net outflows at the intersection I^i_{k-1} lead to flows at the downstream link L^i_k and intersection I^i_k that are lower than the link capacity q_{opt}. This is illustrated in an excerpt of the time-space diagram corresponding to C^i in Figure 5.3.

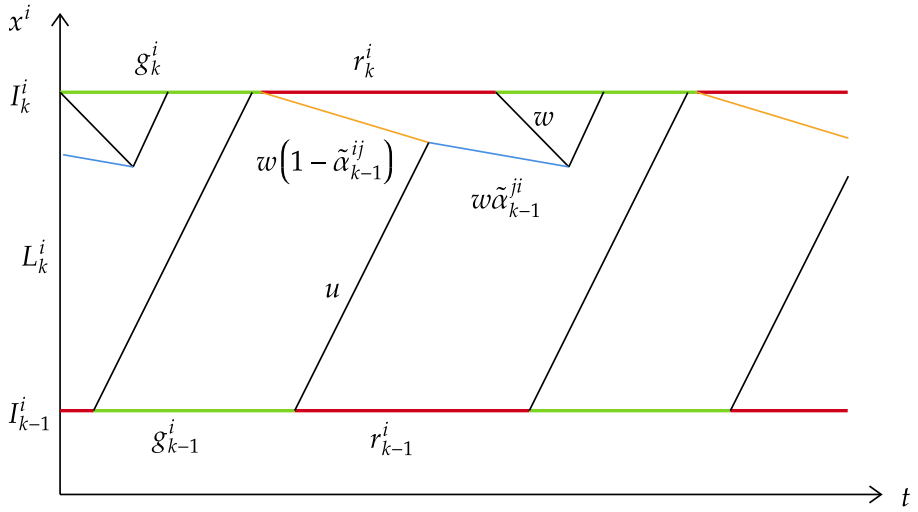


Figure 5.3: Influence of undersaturated links.

The figure shows a link L^i_k between two intersections, I^i_{k-1} and I^i_k, on corridor C^i. The red and green lines represent the signal phases at both intersections. Shock waves with the slopes u and w are shown as black lines. We assume turning ratios such that the total flow during g^i_k is undersaturated, i.e. $\tilde{q}^i_k < q_{opt} \cdot (g^i_k/c^i_k)$ where c^i_k denotes the cycle length at I^i_k. The shock waves illustrating the propagation of the queue from I_k are shown as orange and blue lines. The orange shock wave results from flows staying on C^i, i.e. retaining flows, which join the queue. Thus, the queue grows with speed $w \cdot (1 - \tilde{\alpha}^{ij}_{k-1})$. The blue shock

wave results from turning flows from C^j which join the queue. Thus, the queue grows with $w\tilde{\alpha}^{ji}_{k-1}$. Assuming a triangular FD, the parameter $\tilde{\alpha}$ can be derived as shown in appendix A and written as:

$$\tilde{\alpha} = \alpha \cdot \frac{u}{u + w \cdot (1 - \alpha)}. \tag{5.1}$$

This example shows that the maximum flow $\tilde{q}^i_k$ can be undersaturated since the flow which discharges from the queue does not last for the entire green phase g^i_k. Accounting for intersection-specific signal settings, we can describe the demand $\tilde{q}^i_{k,d}$ for an entire cycle as:

$$\tilde{q}^i_{k,d} = \tilde{q}^i_{k-1} \cdot (1 - \alpha^{ij}_{k-1}) \cdot \frac{g^i_{k-1}}{g^i_k} + \tilde{q}^j_{k-1} \cdot \alpha^{ji}_{k-1} \cdot \frac{r^i_{k-1}}{g^i_k}. \tag{5.2}$$

This equation describes the maximum flow $\tilde{q}^i_{k,d}$ which can occur considering demand only. Note that if a net inflow exists, i.e. $\tilde{q}^i_{k,d} > q_{opt}$, the demand exceeds the link capacity q_{opt}. To address such cases, the required supply-related constraints are introduced in the following.

Supply. Next, we model the effects of spillbacks on the maximum flows, i.e. $\tilde{q}_{k,s}$, denoted by the additional subscript s. Note that such effects are implicitly considered in the original MC as flow conservation is kept. However, the occurrence of inflows and outflows at intersections might lead to a violation of flow conservation at the corridor level even though it is still satisfied at the network level. Therefore, in order to consider the effects of spillbacks, we propose to find the most constraining spillback duration σ_k exogenously as explained in the following. Note that the assumption of a FIFO diverge implies that a spillback occurring on a single link downstream of I_k affects all other outgoing links of I_k, too, if there is turning demand. Below, we derive the spillback duration and incorporate that into $\tilde{q}_{k,s}$ by means of an example.

Recall that flow conservation applies at the link level despite the existence of source terms as so far we have assumed those source terms are located at intersections. Thus, we can derive the duration σ_k during which a spillback in a link affects the upstream intersection I_k by applying traditional concepts of traffic flow theory.

Figure 5.4 displays an example of such a spillback occurrence for the stationary case. It shows an excerpt of a time-space diagram for a link L^i_{k+1} between two intersections, I^i_k and I^i_{k+1}. The red and green lines represent the signal phases at both intersections. Shock waves with the slopes u and w are shown as black lines. For this case, we assume turning ratios such that the total flow during g^i_{k+1} is oversaturated, i.e. the demand exceeds the intersection capacity $\tilde{q}^i_{k+1,d} > q_{opt} \cdot (g^i_{k+1}/c^i_{k+1})$. The queue grows with speed w while saturation flows reach the queue. Once vehicles from upstream of I^i_k join the queue, its propagation speed changes. This is highlighted by the blue and orange curves in the figure. The blue curve represents a queue growth due to inflows from the adjacent corridor (i.e. turning vehicles coming from corridor C^j to C^i), similar to Figure 5.3. The orange one represents a queue growth due to the straight flows (i.e. vehicles staying on C^i). Once the orange shock wave reaches I^i_k, the flow of the corresponding upstream link is blocked and the queue spills over to this link. This marks the beginning of the spillback duration, $t^i_{k,b}$. The spillback ends once the queue at I^i_k

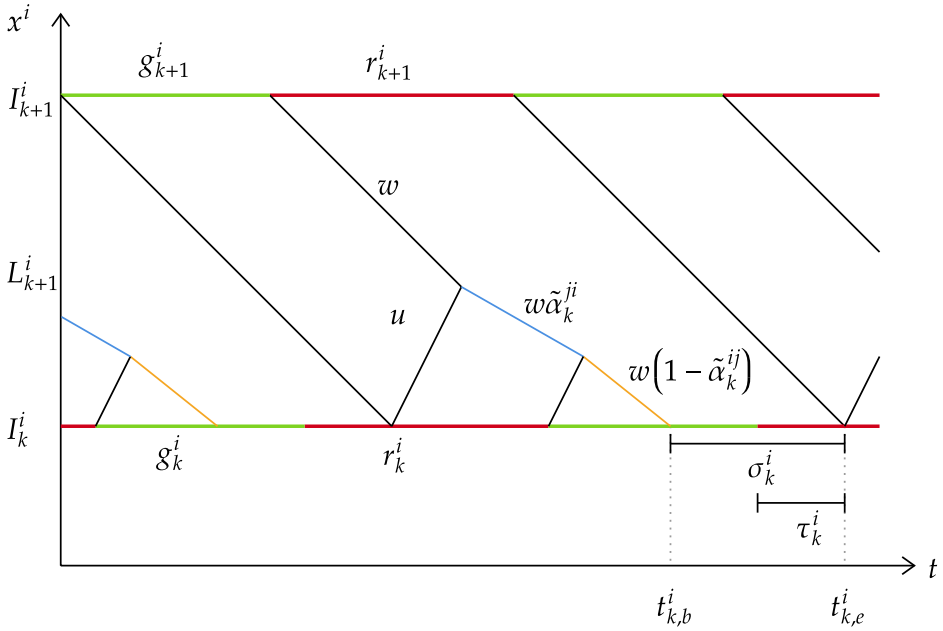


Figure 5.4: Spillback occurrence in a time-space diagram.

dissolves again at the time instant $t^i_{k,e}$. The total duration of the spillback impact is denoted as σ^i_k. This blockage affects the capacity of upstream links and the demand for downstream links on all adjacent corridors. Note that $t^i_{k,e}$ occurs during a red phase in the shown example, and the following considerations refer to that case. Nevertheless, the concept is valid for $t^i_{k,e}$ occurring during a green phase as well. The equations can be derived analogously.

First, we calculate the spillback impact ending time $t^i_{k,e}$ which can easily be found from the time-space diagram:

$$t^i_{k,e} = r^i_{k+1} + \frac{l^i_k}{w}. \tag{5.3}$$

Then, we derive the parameter τ^i_k, which denotes the time between the beginning of the active phase and $t^i_{k,e}$. Considering the offset o^i_k, we derive the parameter as follows:

$$\tau^i_k = mod(t^i_{k,e}, c^i_k) - o^i_k. \tag{5.4}$$

where mod is the modulo operator that calculates the remainder of the division of $t^i_{k,e}$ and c^i_k.

To facilitate the estimation of σ^i_k, we shift our perspective from the time-space diagram to a cumulative plot, as illustrated in Figure 5.5. We then resort to the illustrative and well-established technique of comparing cumulative counts N at different locations introduced as the 3-detector problem by NEWELL [1993]. As described in the previous chapter, a discontinuity in the Moskowitz function N occurs at intersections with source terms. Thus, it is important to note that the following explanations always refer to a link segment where the flow conservation applies and no discontinuity exists. Correspondingly, we derive the cumulative count N^i_{k+1} at the downstream intersection I^i_{k+1}, and $N^i_{k+\delta x}$ at an infinitesimally

small distance downstream of I^i_k, denoted as $I^i_{k+\delta x}$. Furthermore, we assume the spillback duration at I^i_k and $I^i_{k+\delta x}$ to be the same, $\sigma^i_k = \sigma^i_{k+\delta x}$. The count N^i_{k+1} represents the cumulative curve corresponding to the capacity, and $N^i_{k+\delta x}$ the demand. Since we are interested in deriving the spillback impact duration at $I^i_{k+\delta x}$, we shift N^i_{k+1} by l^i_k/w in the temporal dimension, and $l^i_k \cdot \kappa_{max}$ in the N-dimension, where κ_{max} is the link jam density. Figure 5.5 illustrates the shifted cumulative curves, where N^i_{k+1} and $N^i_{k+\delta x}$ are depicted as

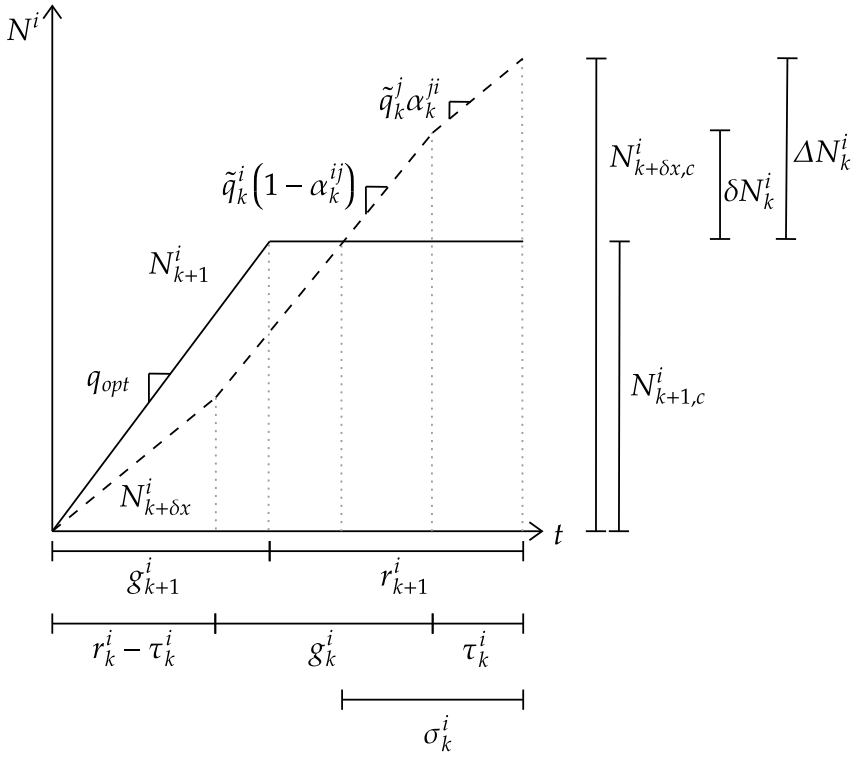


Figure 5.5: Spillback occurrence in a cumulative plot.

a solid and as a dashed curve, respectively. Both N-curves refer to a single cycle length c. This is sufficient, as we assume a common cycle length for all intersections, and analyze stationary states which occur periodically with such cycle length. The maximum cumulative count during the respective cycle is denoted with the subscript c. Based on both cumulative curves, we can derive the spillback impact duration σ^i_k as follows.

The cumulative count $N^i_{k+1,c}$ is equal to the product of link capacity q_{opt} and the green phase at the downstream intersection I^i_{k+1}:

$$N^i_{k+1,c} = g^i_{k+1} \cdot q_{opt}. \tag{5.5}$$

Note that this represents the essential assumption of the LS approach, i.e. queues blocking I^i_{k+1} are not considered in deriving σ^i_k. In other words, spillbacks propagating from links downstream are not considered here. However, this assumption is relaxed in the FS approach, see Section 5.3.3, which is then capable of accounting for such spillback propagation which includes circular dependencies within the network

The cumulative count related to the demand $N^i_{k+\delta x,c}$ equals the sum of the inflows for one cycle. Building upon eq.(5.2), we can write:

$$N^i_{k+\delta x,c} = \tilde{q}^i_k \cdot (1 - \alpha^{ij}_k) \cdot g^i_k + \tilde{q}^j_k \cdot \alpha^{ji}_k \cdot r^i_k. \tag{5.6}$$

The comparison of both N-curves lets us determine the spillback duration σ^i_k at intersection I^i_k. More specifically, we can determine the actual cumulative plot at I^i_k as the lower envelope of both N-curves. The spillback begins once the flow is constrained by the cumulative curve at the downstream intersection, I^i_{k+1}. During that period, the flow is zero. We determine the difference between the demand and the capacity accounting for the spillback:

$$\Delta N^i_k = N^i_{k+\delta x,c} - N^i_{k+1,c} = \tilde{q}^i \cdot (1 - \alpha^{ij}) \cdot g^i_k + \tilde{q}^j \cdot \alpha^{ji} \cdot r^i_k - g^i_{k+1} \cdot q_{opt}. \tag{5.7}$$

Next, we determine the part of ΔN^i_k which remains after subtracting the blocked vehicles during τ^i_k, denoted as δN^i_k. This simplifies the notation to derive σ^i_k.

$$\delta N^i_k = \Delta N^i_k - \tau^i_k \cdot \tilde{q}^j_k \cdot \alpha^{ji}_k. \tag{5.8}$$

Note that a positive δN^i_k corresponds to a case where $\sigma^i_k > \tau^i_k$, i.e. the spillback spans both phases. This case is illustrated in Figure 5.5. On the other hand, a negative δN^i_k corresponds to the case where $\sigma^i_k < \tau^i_k$, i.e. where the spillback only occurs during the red phase.

To derive the total spillback impact duration σ^i_k, we distinguish between the two cases referring to the sign of δN^i_k:

$$\sigma^i_k = \begin{cases} \tau^i_k + \frac{\delta N^i_k}{\tilde{q}^i_k \cdot (1-\alpha^{ij})} & \text{, if } \delta N^i_k > 0 \\ \tau^i_k + \frac{\delta N^i_k}{\tilde{q}^j_k \cdot \alpha^{ji}} & \text{, otherwise.} \end{cases} \tag{5.9}$$

With this equation, one can calculate the spillback duration due to a queue on C^i, which can be relevant for both corridors C^i and C^j at the intersection I^i_k. To approximately account for a FIFO diverge behavior, we propose to calculate ΔN_k for all outgoing links, and subsequently consider the maximum ΔN_k to compute σ_k. During a stationary state, the inflow and outflow in a link are balanced by the spillback duration. Thus, the maximum ΔN_k is required to ensure all flows are stationary. Note that this is an approximation, as we neglect the fact that spillbacks can also affect inflows on adjacent links. However, this approach does not aim at an extremely precise estimation of σ_k, but at estimating the spillback duration sufficiently precise for an accurate MFD approximation at the network level.

The total spillback duration σ^i_k is further divided into the corridor-specific blockage times, which are denoted as $\sigma^{i\to i}_k$ and $\sigma^{i\to j}_k$. These depend on the length of σ^i_k, i.e. whether the spillback lasts only during one signal phase, or spans two phases. Note that σ^i_k can occur during two phases at a maximum, as we assume a common cycle length for all intersections and therefore the theoretically maximum blockage duration equals to the red phase at the downstream intersection which cannot exceed a cycle length. As we assume that $t^i_{k,e}$ occurs during a red phase, $\sigma^{i\to j}_k$ is either τ^i_k or σ^i_k, if $\sigma^i_k < \tau^i_k$. Similar to that, $\sigma^{i\to i}_k$ is the difference

of the total spillback duration and τ_k^i, or equal to zero in $\sigma_k^i < \tau_k^i$. To cover both cases, we write:

$$\begin{aligned} \sigma_k^{i \to i} &= \max\left\{0, \sigma_k^i - \tau_k^i\right\}, \\ \sigma_k^{i \to j} &= \min\left\{\tau_k^i, \sigma_k^i\right\}. \end{aligned} \tag{5.10}$$

It is left to the interested reader to verify that the framework can be derived analogously for the other case, where $t_{k,e}^i$ occurs during an active green phase.

Eventually, the corridor-specific spillback impacts allow us to derive the respective capacity constraints. Therefore, we can formulate the reduction of maximum flows on corridor C^i and corridor C^j due to spillbacks on corridor C^i:

$$\begin{aligned} \tilde{q}_{k,s}^{i \to i} &= q_{opt} \cdot (1 - \frac{\sigma_k^{i \to i}}{g_k^i}), \\ \tilde{q}_{k,s}^{i \to j} &= q_{opt} \cdot (1 - \frac{\sigma_k^{i \to j}}{r_k^i}). \end{aligned} \tag{5.11}$$

Maximum flows. The maximum flow on each link is then the minimum of the demand (eq.(5.2)) and the supply (eq.(5.11)):

$$\tilde{q}_k^i = \min\left\{\tilde{q}_{k,d}^i, \tilde{q}_{k,s}^{i \to i}, \tilde{q}_{k,s}^{j \to i}\right\}. \tag{5.12}$$

This equation is formulated for each link and results in a non-linear system of equations. Its structure assembles widely applied concepts such as the CTM or VT. For network entry and exit links, the demand and supply parts have to be treated separately. For example, there are no demand-related restrictions for links on which flows enter the network. Similarly, there are no capacity constraints for links on which flows exit the network. These links represent origin and destination nodes within the network.

The maximum network-wide average flow does not necessarily result from maximum inflows at origin links. This is because internal bottlenecks might become active at lower inflows and the corresponding congestion can limit the network-wide maximum flows. The maximum average flow is reached when the inflows are as high as possible, but the effects of activated internal bottlenecks remain low. To treat this problem adequately, we formulate an optimization problem based on eq.(5.12). The capacity of a network is defined by the maximum stationary flows possible. We define the inflows at origin links as decision variables. Therefore, the equivalent of eq.(5.12) for origin links, i.e. the link index $k = 1$, can be written as follows:

$$\tilde{q}_1^i = \min\left\{\lambda^i \cdot q_{opt}, \tilde{q}_{1,s}^{i \to i}, \tilde{q}_{1,s}^{j \to i}\right\}. \tag{5.13}$$

where the parameter λ is a factor to determine the demand at origin links as the share of q_{opt}. This factor is found within an optimization problem and corresponds to the highest demand level possible where the effects of internal bottlenecks are low such that the network-wide

average flows are maximized. The overall optimization problem for all links in the network can be formulated as:

$$\max_{\lambda^i} \quad \sum_{i=1}^{n} \tilde{q}_k^i \tag{5.14a}$$

$$\text{s.t.} \quad 0 \leq \lambda^i \leq 1 \tag{5.14b}$$

since all constraints are effectively included in eq.(5.12).

In summary, this subsection proposes an approximate approach to derive the maximum flows $\tilde{q}_k^i$ at the intersection level. These are necessary to define the costs of horizontal edges in $\mathcal{G}$ representing green phases. The approach assumes a triangular FD and common cycle lengths. The effects of spillbacks are only considered at the intersection level and not across entire corridors. Thus, we label it 'Limited spillbacks'. For use cases, where a more exact determination of the maximum flows $\tilde{q}_k^i$ is desired, we propose an alternative in the following subsection.

Approach 2 - Full spillbacks (FS)

The second approach is in large parts similar to the first one. The difference lies in the consideration of spillbacks at downstream intersections. For this purpose, we only need to adapt eq.(5.5) as explained in the following.

We modify the way to estimate the cumulative count $N_{k+1,c}$ representing the maximum number of vehicles passing the downstream intersection I_{k+1} during a cycle. In this approach, $N_{k+1,c}$ is equal to the product of the maximum flow $\tilde{q}_{k+1,s}^i$ at the downstream intersection I_{k+1} and the corresponding green phase:

$$N_{k+1,c}^i = g_{k+1}^i \cdot \tilde{q}_{k+1,s}^i. \tag{5.15}$$

Recall that the index s denotes that this flow can itself be impacted by spillbacks at the respective intersection. This dependency shows the recursive nature of this approach, and it allows us to model the propagation of spillbacks downstream of I_k^i. Since $\tilde{q}_{k+1,s}^i$ can be reduced by spillback from *all* links downstream of I_k^i, this equation allows to account for circular dependencies within the network. Nevertheless, the effects of any constraint are uniformly spread across the green time. This constitutes the main assumption of the FS approach.

The modification of the method to derive $N_{k+1,c}$ further affects the computation of ΔN_k^i. Based on eq.(5.7) and eq.(5.15) above, we write:

$$\Delta N_k^i = N_{k+\delta x,c}^i - N_{k+1,c}^i = \tilde{q}^i \cdot (1 - \alpha^{ij}) \cdot g_k^i + \tilde{q}^j \cdot \alpha^{ji} \cdot r_k^i - g_{k+1}^i \cdot \tilde{q}_{k+1,s}^i. \tag{5.16}$$

The difference in the demand and the capacity ΔN_k^i can then be put into eq.(5.8), and thus enables one to follow the procedure described for the previous approach to eventually calculate the corridor-specific spillback impacts. The FS approach enables one to consider

congestion propagation throughout the network. Thus, we label this approach 'Full spillbacks'. However, this ability comes at the price of higher modeling complexity. More specifically, $\tilde{q}^i_{k+1,s}$ in eq.(5.15) is a variable in contrast to q_{opt} in eq.(5.5) being a constant. Thus, the overall system of equations becomes more complex. However, the traffic dynamics are still simplified. To further increase the accuracy of modeling such dynamics, an appropriate KWT solution scheme shall be used. The application of such a scheme within the context of the proposed framework is presented in the following section.

Approach 3 - Network variational theory (nVT)

The nVT as proposed in the previous chapter enables one to describe the evolution of traffic states in a signalized network with precision and efficiency. Its advantage compared to other macroscopic traffic models is the ability to model complex intra-link bottlenecks. Its advantage compared to microscopic simulation models is the numerical efficiency. Moreover, nVT requires the same inputs as nMC and builds upon a numerical grid similar to a hypernetwork. This leads to further synergies, as the hypernetwork $\mathcal{G}$ can be utilized for such a numerical grid. Nevertheless, compared to the other two approaches it requires the highest modeling efforts to derive the costs for horizontal edges in $\mathcal{G}$.

Cumulative count. The numerical grid is defined by the time-step length Δt, the spatial step length $\Delta x = u \cdot \Delta t$, the length of each corridor $C \in \mathcal{C}$, and the length of the analysis period (see Section 2.2). The latter has to be chosen such that stationary traffic states are reached on each link $L \in \mathcal{L}$, since the MFD is based on such stationary traffic states. In other words, the total computation time of the numerical grid needs to be sufficiently long such that stationary states can be reached within this period. Lastly, turning ratios $\alpha \in A$ are input parameters. Note that this grid has some overlap with the hypernetwork $\mathcal{G}$ (e.g. at intersections) but is generally more fine-meshed as grid nodes also cover space within links.

The nVT finds the cumulative count N at each node $P(x,t)$ in the numerical grid. For a given node $P(x,t)$ within links, the classical VT formulation applies (see Section 2.2). Furthermore, we proposed to exogenously account for the effects of source terms at the points P which are exactly at, and right downstream of an intersection. These equations ensure the correct propagation of free-flow and congested traffic states across intersections despite the existing discontinuities in the Moskowitz surface at such locations. The result is the precise cumulative count N at each point on the numerical grid. For further details on the methodology, please refer to Chapter 4.

Edge costs. The cumulative count $N(x,t)$ found by nVT lets us now derive the costs for horizontal edges z^i_k in the hypernetwork $\mathcal{G}$. Recall that these costs correspond to the flow which would pass a moving observer traveling along such edges. Such moving observers always travel until the termination of a red phase in the original MC, which includes the entire green phase. Hence, it is sufficient to derive the cycle-specific average flows from the nVT solution to specify the costs of horizontal edges in $\mathcal{G}$. We measure the difference in N which

occurs during the beginning and the end of green phases at all intersections to calculate the costs z. These costs consider both, the fact that intersections can be undersaturated due to net outflow at the upstream intersection, and spillbacks that block certain parts of the green phase. In both cases, the costs of the corresponding edge are effectively reduced. More formally, we can determine the costs z_k^i related to a green phase at intersection I_k^i as:

$$z_k^i = \frac{N_{k,s}^i - N_{k,e}^i}{g_k^i}. \tag{5.17}$$

where $N_{k,s}^i$ and $N_{k,e}^i$ denote the cumulative count at intersection I_k^i at the start and end of a green phase, respectively. Note that the numerator is always positive, as N is a monotonic increasing function in t.

Moreover, it is necessary to derive the spillback duration. Again, such durations can be extracted from the nVT solution by analyzing the Moskowitz surface.

Optimization. Similar to the approximate approaches, we find the capacity state of the network with eq.(5.14). However, the maximum flows $\tilde{q}$ are obtained based on eq.(5.17) since the cost of horizontal edges in $\mathcal{G}$ corresponds to these flows.

5.4 Derivation of cuts

The previous section describes three approaches with different levels of modeling complexity to define the costs of horizontal edges in $\mathcal{G}$ representing green phases, which include the effects of source terms on network-wide traffic state evolution for the capacity state. Building upon $\mathcal{G}$, in this section, we further develop our method to estimate the realized network MFD. Thereby, we apply methods from the original MC and utilize recent findings in MFD-related research. First, we estimate the free-flow branch by deriving cuts from $\mathcal{G}$. Similar to LECLERCQ & GEROLIMINIS [2013], moving observers are free to choose any path in $\mathcal{G}$ and are not restricted to one direction as in DAGANZO & GEROLIMINIS [2008]. Second, we approximate the capacity branch also based on $\mathcal{G}$. Finally, we estimate the network-wide jam density and exploit the symmetry of the link FD indicated in LAVAL & CASTRILLÓN [2015] and [DAGANZO & KNOOP, 2016] to derive the congested branch.

5.4.1 Free-flow branch

Similar to the original MC, we rely on the concept of moving observers that travel through the network. We aim at estimating the free-flow branch of each route MFD to be able to aggregate them to the free-flow branch of the network MFD. However, flow conservation within a route might be violated when source terms exist, and thus MC cannot be directly applied (see Section 2.3). To address this, we propose an approximate approach. It is

based on the common notion of path flows, which we will refer to as *route flows* to avoid any confusion with the term *path* as used in Section 2.2.

The route flows consist of vehicles sharing a common route from an origin to a destination and flow conservation applies. Thus, we can apply the original MC by deriving hyperlinks for these routes from the hypernetwork $\mathcal{G}$. Spillback effects are accounted for as the costs are derived from $\mathcal{G}$, and demand-related effects are not considered as flow conservation applies. The general cost of an edge representing a green phase relates to the share of the route flow to the total flow at the specific intersection approach. By deriving all route MFDs and subsequently taking the route length-weighted average, one can approximate the network MFD. Note that this corresponds to the aggregation of traffic states according to Edie's definitions [EDIE, 1963].

Unfortunately, the number of routes connecting each OD pair can become intractable for realistic networks. Hence, to reduce the number of evaluated routes, we conjecture that the network MFD is estimated sufficiently well when the network variability is covered by the evaluated hyperlinks. The corresponding set of routes is non-overlapping and includes each intersection approach. Note that such a set is given by the set of corridors $\mathcal{C}$ defined by the network decomposition in Section 5.3.1. We derive the cuts corresponding to forward-moving observers as described in the original MC for each hyperlink. Eventually, we average the route-specific free-flow branches to derive the one corresponding to the network MFD. We conduct a sensitivity study in Section 5.5.2 to support this procedure and the involved assumptions.

5.4.2 Capacity branch

To derive the capacity branch of the network MFD, we fully utilize the hypernetwork $\mathcal{G}$. It already includes estimates for the maximum flows on each link, $\tilde{q}_k$. The computation of these flows considered both decreased demand due to a net outflow, and spillback induced capacity reductions. We take the link-weighted average to derive the network-wide capacity, as proposed in GEROLIMINIS & DAGANZO [2008]. Note that phenomena such as the short-blocks problem [e.g. DAGANZO & LEHE, 2016] are implicitly taken into account in our methods to define the hypernetwork as we consider spillback effects explicitly.

5.4.3 Congested branch

To conclude the approximation of the MFD we derive its congested branch as follows. First, we estimate the network jam density K_{max}. Then, we exploit a symmetry in the evolvement of free-flow and congested traffic states considering the estimation for K_{max}. This symmetry was shown by LAVAL & CASTRILLÓN [2015] and DAGANZO & KNOOP [2016] concerning the FD which exists when a canonical density transformation is applied. It reflects the symmetric propagation of free-flow and congested traffic states throughout a link. They utilized this symmetry also for the MFD. Our main assumption is that the symmetry still holds when the network jam density is reduced. We discuss this assumption below.

Network jam density

To our best knowledge, no explicit attempts to estimate the network-wide jam density in the context of MFD approximation exist. The network jam density differs from the link jam density because the network is not necessarily fully jammed in case of gridlock [DAGANZO ET AL., 2011; MAHMASSANI ET AL., 2013; MAZLOUMIAN ET AL., 2010]. In the case study presented in Section 5.5 the significant improvement of the MFD approximation considering the network jam density becomes apparent.

First, we let the traffic in the physical network reach the capacity state first. Then, we define virtual links at the destination nodes in the network, hereafter referred to as 'destination links', and simultaneously set their capacity to zero. This results in queues originating on those links which propagate through the network until a gridlock state is reached. By approximating the queue growth throughout the network, we can derive the network jam density K_{max}. Note that gridlocks in reality might occur due to an interplay of many different aspects such as outflow reductions due to spillbacks, driver heterogeneity, restrictive internal bottlenecks such as traffic lights, and additional demand generation inside or at the fringe of the given network. Nevertheless, we assume that reducing the outflow capacities at destination links leads to the highest number of jammed links in the network. Therefore, it results in the highest value for the network jam density which is preferable as we look for the MFD as an upper bound.

When we restrict capacities at destination links, the density K is influenced by the remaining network outflow and inflow. The strongest outflow restriction is represented by a simultaneous closing of all destination links. If some links were not closed, vehicles could still leave the network. However, the queue growth throughout the network that results from the outflow restriction highly depends on the turning ratios at each intersection. A blocked intersection leads to zero flows at all downstream links independently of the respective queues. Thus, the link densities can be smaller than κ_{max}. The network inflow is determined by the maximum of the demand and the capacity of origin links. While the demand is independent of the outflow restriction, the capacity of origin links depends on whether queues have already reached these links. Due to traffic interdependencies in the network, queues can grow even though some destination links might not be restricted. In other words, a queue growing from one restricted destination link can finally block multiple origin links. Thus, for the case of a non-simultaneous outflow restriction, we assume that net outflows out of the system are larger than in the case of a simultaneous outflow restriction. Therefore, we conjecture that the maximum number of vehicles in the network, and thus K_{max} considering spatial demand patterns, can be achieved in the latter case. In the following, we propose an approximate approach and one based on nVT for the estimation of K_{max}.

Approach 1 - Queue propagation (QP). We assume the capacity of outgoing links at the boundary of $\mathcal{N}$ to be zero and trace the propagation of queues throughout the network. Thus, we label this approach as 'queue propagation' (QP). Once a stationary gridlock state is reached, we derive the number of vehicles per link, and subsequently, take the link-weighted average to derive K_{max}.

For each link L_k^i we derive the time instant $t_{k,dq}^i$ when the queue starts to grow at the link's downstream end, and approximate the time when it reaches the link's upstream end, $t_{k,uq}^i$. Note that the latter time instant equals to $t_{k-1,dq}^i$, i.e. the time when the queue reaches the upstream link L_{k-1}^i, if no earlier blockage from an adjacent link occurs. To derive $t_{k,uq}^i$, we distinguish between two cases.

In the first case, we focus on links where no spillback occurred when the network was at capacity, i.e. $\tilde{q}_k^i < q_{k,s}^i$. Such a case is displayed in Figure 5.6. It shows a time-space diagram for the link L_k^i between two intersections I_{k-1}^i and I_k^i.

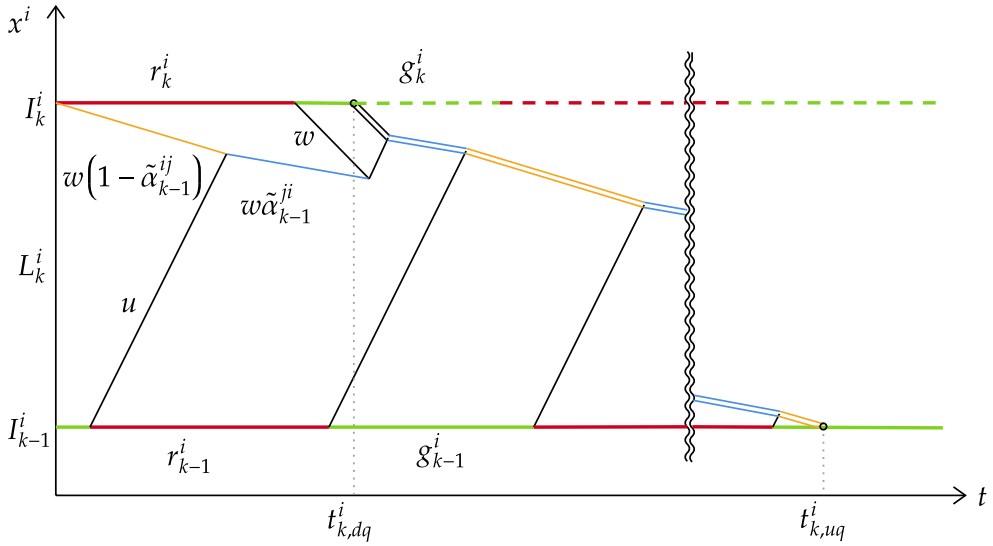


Figure 5.6: Queue growth for undersaturated links.

The left part of the figure reflects the traffic conditions for the network capacity state (see Section 5.3.3). At time $t_{k,dq}^i$ a queue reaches the intersection I_k^i from downstream. Note that $t_{k,dq}^i$ can result from downstream links from both corridors, C^i and C^j. However, the earlier queue is decisive for the further propagation, as we assume FIFO diverging behavior. If the earliest queue arrives from C^j, it occurs during a red phase on C^i. The queue on link L_k^i starts to grow with the start of the respective red phase, as vehicles stopping due to a red phase will not be able to continue their travel once the green phase starts due to the existing queue downstream. Formally, the time when the queue starts to grow at the downstream end of link L_k^i can be written as:

$$t_{k,dq}^i = \min \left\{ \lfloor \frac{t_{k+1,uq}^j}{c_k^i} \rfloor \cdot c_k^i, t_{k+1,uq}^i \right\}, \tag{5.18}$$

where the $\lfloor . \rfloor$ is the *floor* function, which ensures $t_{k,dq}^i$ starts with a cycle and therefore with the red phase if the decisive queue comes from C^j.

From this time onward, the capacity at I_k^i is reduced to zero and the queue starts to grow on link L_k^i. During the occurrence of discharging flows, the queue grows with speed w as

shown in the figure. In case that flows from the upstream intersection I^i_{k-1} reach the queue, it grows with speeds $w \cdot (1 - \tilde{\alpha}^{ij}_{k-1})$ and $w \cdot \tilde{\alpha}^{ji}_{k-1}$. The growth of the queue is displayed as a double line in the figure. For the sake of simplicity, we neglect the impact of discharging flows which are supposed to be small in the considered case, and approximate the growth as the cycle-based average of the blue and orange shock waves:

$$t^i_{k,uq} = t^i_{k,dq} + \frac{l^i_k}{w} \cdot \frac{c^i_k}{r^i_k \cdot \tilde{\alpha}^{ji}_{k-1} + g^i_k \cdot (1 - \tilde{\alpha}^{ij}_{k-1})}. \tag{5.19}$$

In the second case, we focus on an intersection I^i_{k+1} that is oversaturated when the network is at capacity state, and therefore a spillback occurs. Thus, the queue grows faster as in the first case since incoming flows are typically close to saturation.

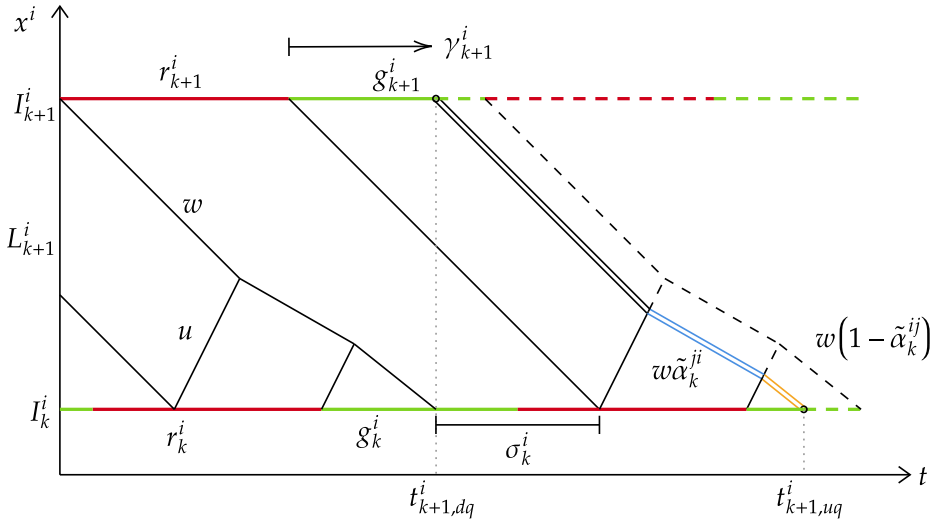


Figure 5.7: Queue growth for oversaturated links.

This is shown in Figure 5.7 as a time-space diagram for link L^i_{k+1} between the intersections I^i_k and I^i_{k+1}. Note that a spillback occurs at I^i_k denoted as σ^i_k. In the figure, a queue reaches I^i_{k+1} at $t^i_{k+1,dq}$ from downstream. The queue grows with speed w as long as saturation flow reaches it. Closer to I^i_k, the queue grows depending on the inflows with speed $w \cdot (1 - \tilde{\alpha}^{ij}_{k-1})$ and $w \cdot \tilde{\alpha}^{ji}_{k-1}$, and reaches I^i_k at $t^i_{k+1,uq}$. If $t^i_{k+1,dq}$ equals the beginning of the red phase r^i_{k+1}, we can calculate $t^i_{k+1,uq}$ based on eq.(5.9) in Section 5.3.3, i.e. by deriving the spillback time σ^i_k. Contrarily, if $t^i_{k+1,dq}$ is close to the beginning of the green phase g^i_{k+1}, the queue grows with speed w nearly all the time. However, if $t^i_{k+1,dq}$ is exactly equal to the beginning of the green time, the queue starts growing at the beginning of the previous red phase. To describe this formally, we introduce the parameter $\gamma = [0, 1]$, as shown in the figure, which is 1 if $t^i_{k,dq} = r^i_k$, and close to 0 if $t^i_{k,dq}$ is close to g^i_k. To approximate all cases in between, we linearly interpolate between $\gamma = (0, 1]$. The case of $\gamma = 0$ is treated separately.

$$t^i_{k+1,uq} = \begin{cases} t^i_{k+1,dq} + \frac{l^i_{k+1}}{w} + \gamma^i_{k+1} \cdot (r^i_{k+1} - \sigma^i_k) & \text{, if } \gamma^i_{k+1} > 0 \\ t^i_{k+1,dq} + \frac{l^i_{k+1}}{w} - \sigma^i_k & \text{, otherwise.} \end{cases} \tag{5.20}$$

With equations (5.18)-(5.20) we approximate congestion propagation throughout the network. However, a blocked intersection due to a queue on C^i also blocks flows to C^j due to the FIFO diverging behavior. This leads to reduced inflows to the respective link. In the extremal case, the entire link on C^j might become empty due to such a blockage. This is essentially the reason for K_{max} being smaller than the link jam density κ_{max}. To account for such cases, we trace the last vehicle which enters and exits a link. A corresponding example is illustrated in Figure 5.8. In this example, a blockage occurs at intersection I^i_{k-1} at $t^i_{k-1,dq}$. This is due to a queue reaching the intersection from the adjacent corridor C^j, which is not shown in the figure. The remaining vehicles on link L^i_k join the existing queue at I^i_k, and leave the link at $t^i_{k,dq}$ during the respective green phase. In this example, no queue from downstream of I^i_k occurs at $t^i_{k,dq}$ and therefore the link L^i_k becomes empty. In contrast to the cases described above, the flows at I^i_k are reduced to zero due to a lack of demand and not due to a queue-induced blockage.

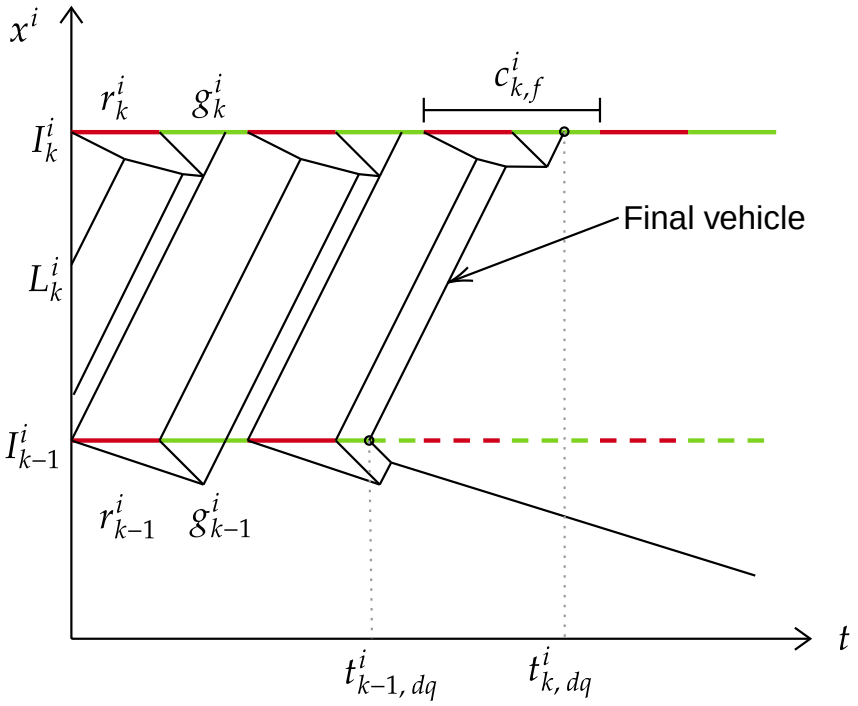


Figure 5.8: Downstream effects of spillbacks.

To describe this formally, let $c^i_{k,f}$ denote the beginning of the cycle in which the final vehicle reaches the existing queue at the downstream end of the link.

$$c^i_{k,f} = c^i_k \cdot \lfloor \frac{t^i_{k-1,dq} + \frac{l^i_k}{u}}{c^i_k} \rfloor, \tag{5.21}$$

The remaining flow relevant to $c^i_{k,f}$ is denoted as $q^i_{k,f}$. It can be derived based on inflows at the upstream intersection. The ratio of the remaining flow to the saturation flow q_{opt}

corresponds to the part of the green phase where flows larger than zero occur at I_k^i. Thus, the time $t_{k,dq}^i$ when $q = 0$ is equal to the red phase r_k^i of the cycle $c_{k,f}^i$, and the part of the green time until the intersection is cleared.

$$t_{k,f}^i = c_{k,f}^i + r_k^i + g_k^i \cdot \frac{q_{k,f}^i}{q_{opt}}, \tag{5.22}$$

To account for this, we rewrite eq.(5.18):

$$t_{k,dq}^i = \min \left\{ \lfloor \frac{t_{k+1,uq}^j}{c_k^i} \rfloor \cdot c_k^i, t_{k+1,uq}^i, t_{k,f}^i \right\}. \tag{5.23}$$

This completes the set of equations needed to track congestion propagation throughout the network for the gridlock case. The equations (5.19), (5.20), and (5.23) are applied for each link L_k^i on a corridor C^i of the set $\mathcal{C}$. However, the queues from an adjacent corridor might reach an intersection earlier than the current one being evaluated with the equations. Thus, an iterative approach needs to be employed until the resulting times $t_{k,uq}^i$ and $t_{k,dq}^i$ converge.

Subsequently, we estimate the maximum density per link K_k^i during the gridlock state. Thereby, we consider that a link might not fully load due to a blockage from an adjacent corridor at the upstream intersection. The time span needed to fully load the link is $t_{k,uq}^i - (t_{k,dq}^i - l_k^i/u)$, which is based on the equations above. However, the actual loading duration for a link is $\max\{t_{k-1,dq}^i - t_{k,dq}^i + l_k^i/u, 0\}$ since $t_{k-1,dq}^i$ also accounts for blockages from adjacent corridors. The maximum operator is applied to exclude infeasible values. We approximate the actual maximum density per link as follows:

$$K_k^i = \kappa_{max} \cdot \frac{\max\left\{t_{k-1,dq}^i - t_{k,dq}^i + l_k^i/u, 0\right\}}{t_{k,uq}^i - t_{l,dq}^i + l_k^i/u}, \tag{5.24}$$

Finally, the network-wide jam density is calculated as:

$$K_{max} = \frac{\sum_{i,k} K_k^i}{\sum_{i,k} l_k^i}, \tag{5.25}$$

Approach 2 - Network variational theory (nVT). Similar to estimating the cost of horizontal edges in $\mathcal{G}$, we propose a second approach based on nVT to estimate the network-wide jam density K_{max}. Compared to the QP approach, it has the advantage of being more exact as fewer assumptions are involved. On the other hand, this approach has a higher modeling complexity and computational cost. Alternatively, any other numerical solution method to solve network KWT problems applies. However, there exist synergies between nVT and the nMC framework proposed in this chapter which we intend to utilize (see Section 5.3.3).

To estimate K_{max}, we let the network reach the capacity state as in the estimation of costs for $\mathcal{G}$, and then block the outflow links. This leads to queues being propagated throughout

the entire network. Once the system reaches a stationary state, the link density can be derived by evaluating the difference of the Moskowitz function in the spatial dimension for each link, and subsequently dividing it by the link length:

$$K_k^i = \frac{N_{k,x=0}^i - N_{k,x=l_k^i}^i}{l_k^i}. \tag{5.26}$$

Based on K_k^i for each link, we eventually apply eq.(5.25) to derive K_{max}.

Density transformation

In order to finish the approximation of the congested branch of the network MFD, we utilize the symmetry between the free-flow and congested branch, as suggested by LAVAL & CASTRILLÓN [2015] and DAGANZO & KNOOP [2016].

More specifically, we transform densities of the free-flow and capacity MFD branch:

$$\kappa' = \kappa - \frac{1}{2} \cdot \left(1 - \left(\frac{1}{-w} - \frac{1}{u}\right) \cdot q\right). \tag{5.27}$$

where κ' denotes the transformed density. Note that this equation is based on a normalized FD with $q_{opt} = 1$ and $\kappa_{max} = 1$. Based on the approximation of the free-flow and the capacity branch in Sections 5.4.1 and 5.4.2, we know the corresponding (q, κ) values. As noted in LAVAL & CASTRILLÓN [2015], the respective transformed densities range between $[-0.5, 0]$.

To approximate the congested branch, we mirror those traffic states by changing the sign of κ' for each traffic state (q, κ'). Re-transforming the original as well as the newly generated traffic states based on eq.(5.27), and considering K_{max} instead of the link jam density κ_{max} leads to a fully defined MFD including the congested branch. Hereby, the main assumption is that the symmetry still holds even though we utilize the reduced network jam density instead of the link jam density. This assumption is equivalent to assuming an increase of the average speed with which congestion spreads at the corridor level. Thus, also the average backward wave speed can be decreased (i.e. the value becomes more negative). This seems to be reasonable given the following example. Consider a corridor with three links in a network. Let us assume the most downstream link becomes fully congested, but then a queue from an adjacent corridor blocks the upstream intersection such that the second link remains empty, and the queue continues to grow on the most upstream link. Eventually, only the most downstream and most upstream links are fully congested, while the middle link remains empty. Congestion spreads with the corresponding shock wave speed on the most downstream link, skips the middle link, and then spreads on the most upstream link. The total time to reach a stationary state divided by the corridor length can be interpreted as the speed with which congestion spreads for this specific corridor. This speed is higher than for the case without any network effects. This corresponds to the implicit assumption that the average speed with which congestion spreads is increased, when we apply the shear transformation using the network jam density K_{max}. Thus, we conjecture that assuming the symmetry to hold for K_{max} is reasonable.

5.4.4 Summary

The sections above described the nMC framework we propose to estimate the realized MFD for realistic networks. This includes the effects of source terms, such as undersaturated links and spillbacks. For clarity, we hereafter provide a brief summary.

The main required input is the network $\mathcal{N}$. This includes the network topology such as the number of intersections, respective control settings, number of links, the link FDs, as well as the turning ratios at intersections.

The next step is the generation of the hypernetwork (Section 5.3). This requires the decomposition of the network $\mathcal{N}$ into a set of corridors $\mathcal{C}$ (Section 5.3.1). Based on these corridors, the hypernetwork is created to reflect supply conditions induced by signal control and network topology (Section 5.3.2). Subsequently, we propose to exogenously consider the effects of source terms. We do so with the two approximate approaches LS and FS (Section 5.3.3 and 5.3.3), as well as the application of nVT (Section 5.3.3) as substantial methodological synergies exist with the herein proposed nMC.

With the hypernetwork $\mathcal{G}$ we can then estimate the free-flow branch (Section 5.4.1) and the capacity branch (Section 5.4.2). To estimate the congested branch, we propose a method including an approximate and an nVT approach which estimates the network-wide jam density and utilizes symmetries in the FD (Section 5.4.3).

5.5 Case study

5.5.1 Case study design

Network initialization

The case study is conducted based on the bi-directional Sioux Falls network as shown in Figure 5.9. For the sake of simplicity, we alter the original Sioux Falls network slightly by only considering intersections with four or fewer legs, and we disallow left-turns at intersections. In total, the network consists of 23 intersections, shown as filled and empty circles, connected by 36 bi-directional links, as highlighted by the arrows. All intersections are controlled by a fixed signal control scheme with a cycle time of 90 s, and a green and red phase each of 45 s. All offsets are set to zero. We assume an identical FD for all links, characterized by $q_{opt} = 1800$ veh/h, $u = 10$ m/s, and $w = -5$ m/s.

We define origin and destination nodes as those with less than four legs. They are shown as filled circles in the figure. At these nodes, we add virtual links which serve as origin and destination links and thereby allow to consider demand generation within the network. Additionally, we randomly select reasonable but different turning ratios for each intersection between $0.25 \leq \alpha \leq 0.75$. These limits are chosen to exclude extremal turning ratios which lead to special and rather unrealistic spillback occurrences in the network, as numerical investigations have shown. More importantly, the chosen turning ratios lead to an interdependency of traffic dynamics across the entire network. This includes the propagation

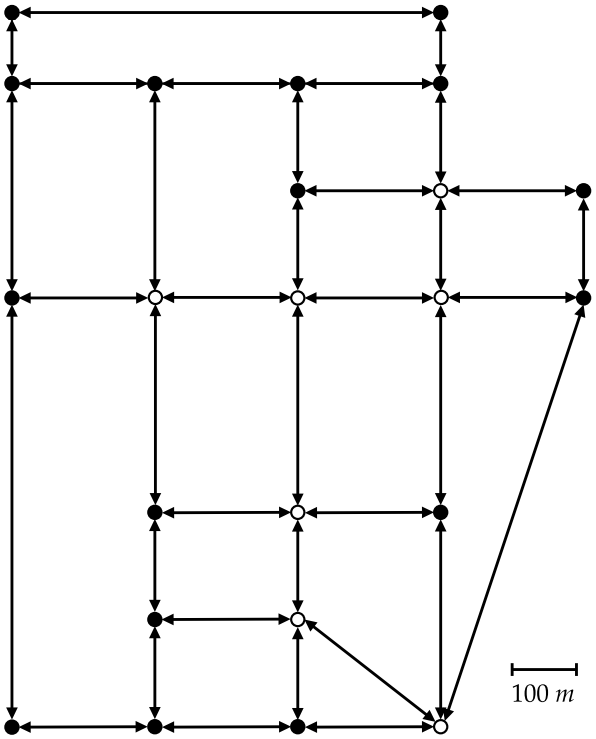


Figure 5.9: Sioux Falls network for the nMC case study.

of spillbacks within and across corridors, i.e. circular dependencies exist in our experiment. This becomes clear when considering that turning ratios at all intersections are non-zero and that the network is bidirectional. Thus, circular flows can occur between any set of at least four intersections forming a rectangle within the network.

Ground truth

In order to further evaluate our MFD approximation results, we compare it to a ground truth derived from the CTM which is a macroscopic, time- and space-discrete traffic model. We implement the same network topology and the total simulation time to 2 h. We set the simulation time step to $\Delta t = 0.25$ s, and the spatial step size to $\Delta x = u\Delta t$ to satisfy the Courant-Friedrichs-Lewy condition [COURANT, FRIEDRICHS & LEWY, 1928]. Setting a lower time-step was not feasible due to the high computational cost.

To derive the full range of the MFD with theCTM, we define several scenarios which are characterised by a constant inflow level and an exogenous reduction of the capacity at destination links. The former aspect is required in order to fully load the network and to reach capacity. The latter measure is applied to create significant congestion spreading throughout the network and affecting the stationary network-wide traffic state accordingly. While congestion can generally occur in networks also without the explicit reduction of outflows, e.g. due to very restrictive intersection control, this procedure facilitates the replicability of the experiments and the reproducibility of the corresponding results. To

generate multiple stationary traffic states in the free-flow and congested regime of the MFD, we consider multiple inflow and exogenous capacity reduction levels for the CTM simulation. More specifically, we specify ten different inflow levels between 0 and 900 veh/h per origin link, and 20 different capacity reduction factors between 0 and 1 at destination links. In each scenario, the network is initially empty and the same level of inflow and capacity reduction applies to all origin and destination links, respectively. Subsequently, the network is loaded with the specified inflow, and traffic dynamics are simulated until a stationary state is reached. If a congested traffic state is to be obtained in a scenario, the respective exogenous capacity reduction at destination links applies once the network is stationary at the capacity level. This ensures the transition from a stationary state. To plot the related MFD, we aggregate flow and density values for one cycle length once stationary states are reached at the end of the simulation. We verify the existence of a network-wide stationary traffic state by ensuring that no differences exist between the average densities of the last two cycles in each simulation run.

Note that the CTM includes a certain error in the calculated density and flow values due to the numerical diffusion. This error converges to zero with a decreasing time step Δt. The chosen time-step represents a reasonable trade-off between accuracy and computational cost. Nevertheless, the ground truth results should be treated accordingly when being interpreted.

Settings for the network variational theory approach

The nVT is applied to derive the hypernetwork, and to estimate the network-wide jam density. For both applications, a numerical grid is defined based on the decomposed network (see Section 5.3.1). The total computation period is set to 2 h to reach stationary traffic states, and the time-step is set to $\Delta t = 0.1$ s. This small time-step highlights the lower computational cost and consequently the increased accuracy of nVT compared to the CTM's one.

In order to estimate K_{max}, the capacity of destination links is set to zero after 1.5 h. This ensures that the stationary traffic states related to the network capacity are reached before the gridlock state occurs.

State of the art

We choose the MC by DAGANZO & GEROLIMINIS [2008], by LECLERCQ & GEROLIMINIS [2013], and the SA by LAVAL & CASTRILLÓN [2015] as state of the art for this case study. Hereafter, we abbreviate them with 'Dag', 'Lec', and 'Lav'. The first method is chosen as it represents the original version of the MC applied to estimate the MFD. The authors find a representative corridor of the network, consider average supply parameters such as block lengths and signal settings, and estimate the MFD. The second method was the first to explicitly account for irregularities in corridors, i.e. differing block lengths and signal settings. The method can be applied to each corridor, and the resulting MFDs be averaged to approximate the network MFD. The last method accounts for stochasticity in the network.

In the evaluated case, the level of stochasticity remains small and only refers to varying block lengths. Nevertheless, we include the method for comprehensiveness. Note that none of these methods explicitly account for the effects of source terms, i.e. the exchange of flows across corridors. Furthermore, these methods implicitly focus on the idealized MFD for corridors. Still, we consider them as state of the art for estimating the realized MFD as no better alternatives exist to our best knowledge. More specifically, methods that estimate the MFD while accounting for turning flows in realistic networks do not exist, as the reported studies all apply to simplistic networks such as regular grids or two-ring networks.

5.5.2 Results and discussion

In the following, we present several MFDs. The x-axis always displays the network-wide average density K in veh/km and the y-axis the network-wide average flow Q in veh/h.

Impact of hypernetwork generation method

We compare the MFDs estimated by our proposed framework with the one derived by the 'Dag', 'Lec', and 'Lav' methods, as well as with the CTM ground truth. The results are displayed in Figure 5.10. The CTM results are shown as grey diamonds. The MFDs from the proposed framework are based on the nVT approach, as well as on the LS and FS approaches to derive $\mathcal{G}$ combined with the QP approximation for the network jam density. They are displayed as a solid black curve, a dashed black curve, and a dotted black curve, respectively. The MFD resulting from the state-of-the-art methods are shown as solid, dashed, and dotted grey curves.

The ground truth MFD reaches a capacity of ca. 670 veh/h, and a network-wide jam density of ca. 94 veh/km.

The nVT approach to derive the hypernetwork and estimate the network-wide jam density results in the most accurate MFD approximation as indicated by the figure. The capacity is ca. 25 veh/h lower than the one of the CTM, while the network-wide jam density is the same. The reason for this is that nVT itself solves the KWT problem at the network level accurately, and its integration into the hypernetwork generation, as well as into the estimation of K_{max}, let the overall framework profit from such accuracy. As expected, the estimated MFDs resulting from the hypernetwork built based on the approximate approaches are less accurate. The first approach, abbreviated as 'LS-QP', which does not consider network-wide spillback propagation overestimates the capacity by ca. 75 veh/h and the jam density by ca. 10 veh/km. The second one, abbreviated as 'FS-QP', which includes network-wide spillback propagation, estimates Q_{max} as accurately as the nVT approach, with an error of 25 veh/h. However, here the difference results from an overestimation rather than an underestimation. The error in the jam density estimation is ca. 8 veh/h and thus similar to the LS-QP approach. In reality, an accurate capacity branch approximation is more important as the congested branch is rarely observed [e.g. LODER ET AL., 2019b]. Thus, we adjudge the increased modeling complexity of the FS approach as valuable for the presented case.

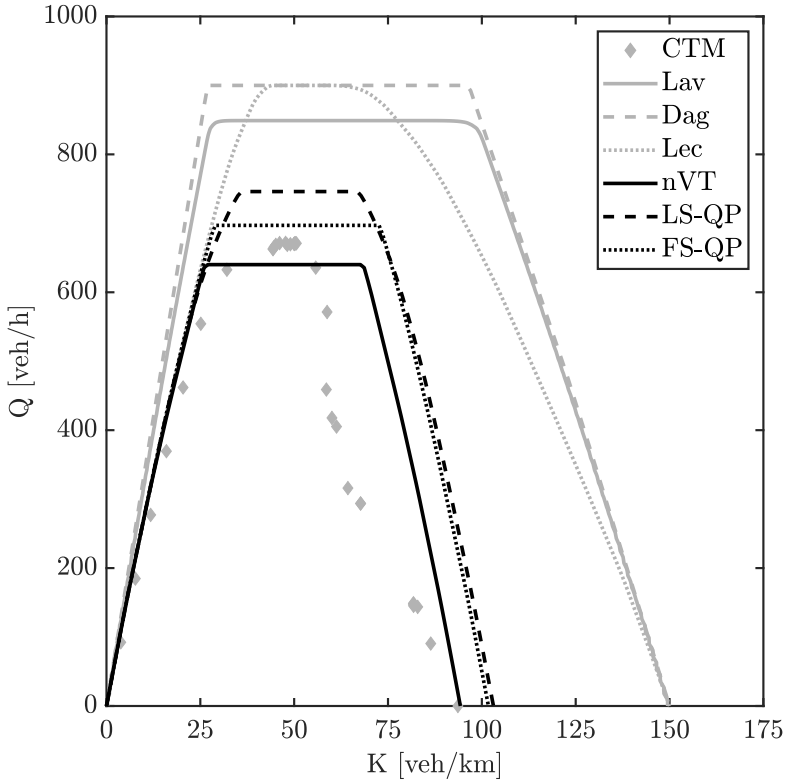


Figure 5.10: Resulting MFDs from the proposed framework based on the nVT, the LS-QP, and the FS-QP approach, the state-of-the-art methods represented by the original method of cuts by DAGANZO & GEROLIMINIS [2008] ('Dag') and by LECLERCQ & GEROLIMINIS [2013] ('Lec'), the stochastic approximation by LAVAL & CASTRILLÓN [2015] ('Lav'), and the CTM ground truth.

The differences in the network-wide capacity and the jam density estimation by the approximate approaches and the nVT one are due to the assumptions made within the approximations. In particular, averaging the downstream spillback impact uniformly on the downstream capacity neglects the temporal component of the spillback occurrence. However, such temporal aspects impact the further propagation of the spillback, i.e. to what share it is propagated to corridors C^i and C^j. Similar to that, we assume the inflow into a link as the equally weighted mean of both retaining and turning flows to derive K_{max}. The shorter a link is the more likely these assumptions are violated.

The results from the state-of-the-art methods substantially differ from the ground truth. Both the capacity and the jam density are significantly overestimated. However, this is expected, as these methods cannot account for the effects of source terms on the network MFD. In other words, they cannot account for the effects of spatial demand patterns, i.e. turning ratios, which is important to derive an estimate for the realized MFD as our results clearly show.

Given the error included in the ground truth and the magnitude of improvement compared to the results of the state-of-the-art methods, the differences in Q_{max} and K_{max} appear to be small for all MFDs resulting from our proposed framework. Moreover, these first results indicate that indeed the overall approximation accuracy is the highest for the nVT approach, and the consideration of spillback propagation throughout the network in the FS approach is valuable.

Computational costs

The modeling complexity of the nVT approach is clearly the highest compared to the other two semi-analytical approaches. This is not surprising, as it numerically solves the underlying KWT problem at the network level. For the case study settings, the nVT takes 4 h 50 min to derive the MFD. The FS-QP approach consists of more decision variables in the optimization problem than the LS-QP one and thus has the highest complexity of the two. Both approaches take 1.2 min for the case study settings and are thus substantially cheaper than the nVT from a computational perspective. The computational cost of the nVT strongly scales with the chosen time-step (see Chapter 4). For example, if the time-step is set to $\Delta t = 1$ s, the computation time is only 14.3 minutes for the case study. The other two methods scale primarily with the number of links in the network.

The computation times for these three approaches are of the magnitude of some minutes, while the CTM requires several days to derive the MFD for this case study.[1] This high computational time for the ground truth results from the small time-step (even though it is still 2.5 times larger than that used in the nVT approach), the resulting large number of cells, the necessity to reach stationary states, and the need to replicate the simulation for multiple demand levels. Similar or even larger computational requirements could be equally expected from other simulations. Furthermore, note that the optimization procedure integrated in the nMC framework ensures that highest network-wide flows are found, while the CTM-related MFD only represents a point-wise evaluation of traffic states which can miss the true capacity.

Impact of route selection

As described in Section 5.4.1, an assumption made in order to reduce modeling complexity and computational burden while approximating the network MFD refers to the set of routes based on which the free-flow branch is estimated. In Section 5.4.1, we propose to estimate the free-flow branch for each corridor $C \in \mathcal{C}$ being aware that this is a simplification. In order to study the implication of this simplification, we estimate the MFD based on the nVT approach for two different sets of routes in the network. First, we evaluate the MFD based on $\mathcal{C}$ (as proposed in Section 5.4.1), which results in 24 different routes. Second, we extract the shortest paths for all OD pairs in the Sioux Falls network and define each path as a route for which the MFD is estimated. In this case, the number of routes increases to a total of 576.

[1] The methods were evaluated on a computer with an Intel(R) Xeon(R) W-2145 CPU with 3.70 GHz and 64 GB RAM.

Note that the capacity branch is not affected by this assumption and the congested branch equals the transformed free-flow one. Therefore, we only evaluate the free-flow branch of the MFD for the two sets of routes and compare them below.

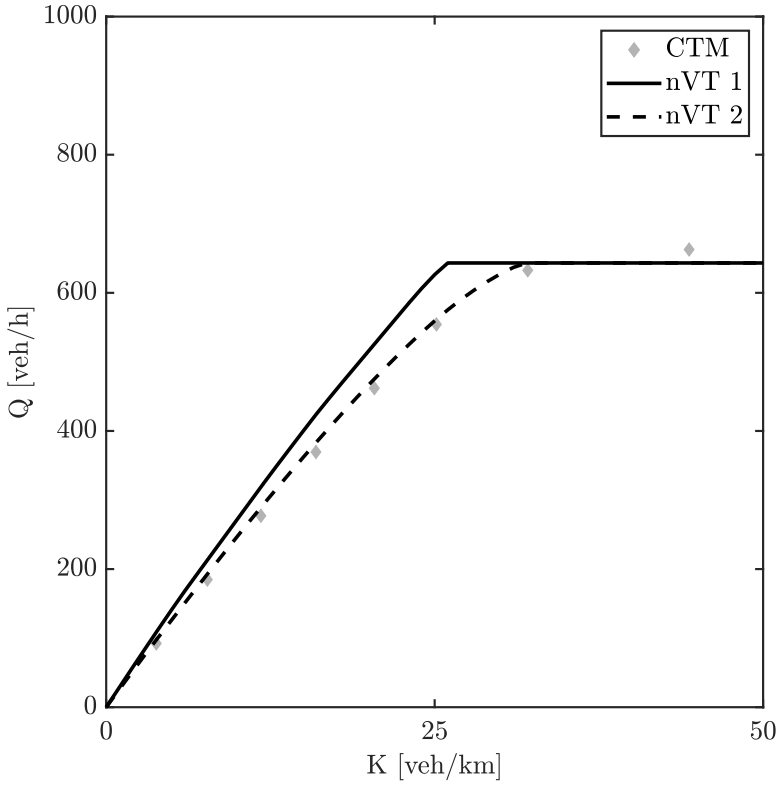


Figure 5.11: Analysis of the impact of route selection.

Figure 5.11 shows the free-flow branch of both MFDs. The solid black curve labeled as 'nVT 1' represents the one based on $\mathcal{C}$. The dashed black curve 'nVT 2' illustrates the derived free-flow branch which results from considering the second set of routes. The figure indicates that indeed the approximation's accuracy is further increased by the latter approach as the curve is closer to the ground truth MFD. However, the number of paths to be evaluated is 24 times higher for our case study. This leads to a significant increase in computational cost. In the end, the selection of routes for the MFD approximation is use-case dependent and the choice represents a trade-off between computational burden and accuracy.

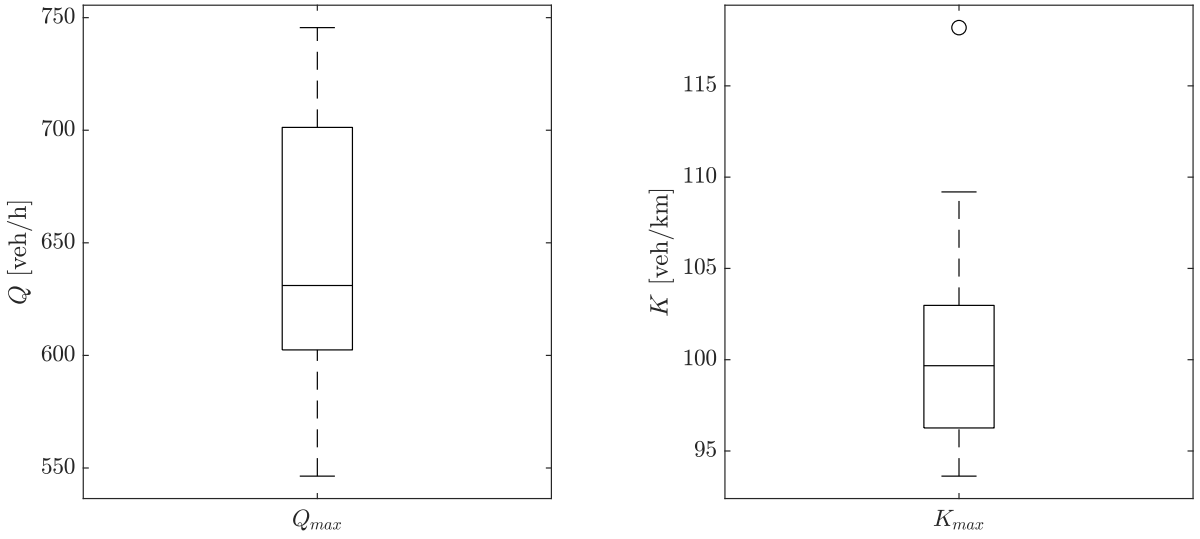
Impact of turning ratios

Next, we aim at evaluating our framework for the presented network with different turning ratio sets $\mathcal{A}$. More specifically, we first analyze the variability of the resulting MFDs from the ground truth to validate that spatial demand patterns indeed have an impact. Second, we compare the results from our proposed framework to the state of the art and the ground truth to assess the quality of our approach. Last, we pick a specific set $\mathcal{A}$, introduce small

deviations, and investigate the effects on the resulting MFDs based on our framework. This indicates the robustness of our MFD approximation method against different spatial demand patterns.

For the first analysis, we randomly select ten different reasonable sets $\mathcal{A}$ with turning ratios ranging again between $\alpha = [0.25, 0.75]$ with a mean at 0.5. We then evaluate the CTM ground truth, the LS, the FS, and the nVT approaches to derive the hypernetwork, and the QP and nVT approaches to approximate the network jam density K_{max}. Moreover, the state-of-the-art methods are applied. We estimate the free-flow branch only for the set $\mathcal{C}$, as we have shown in the previous section that the impact of this assumption is minor.

As the proposed methodology primarily focuses on improving the estimation of the maximum flows $\tilde{q}$ and thus the network-wide capacity Q_{max}, as well as the jam density K_{max}, we merely compare the corresponding results. This allows us to drastically reduce the scenarios to be run with the CTM. The results from the CTM simulation lead to varying Q_{max} and K_{max} for each scenario as illustrated in Figure 5.12. A range of about 200 veh/h for Q_{max} can be observed. The values of K_{max} cover a range of nearly 25 veh/km. The significant variability in these results shows that the MFD indeed depends on the spatial demand pattern in the network. This further indicates that a framework to estimate the MFD for specific spatial demand patterns is beneficial. Despite the assumption regarding the existence of stationary states, which might be violated in reality, the approximated MFDs can be useful.



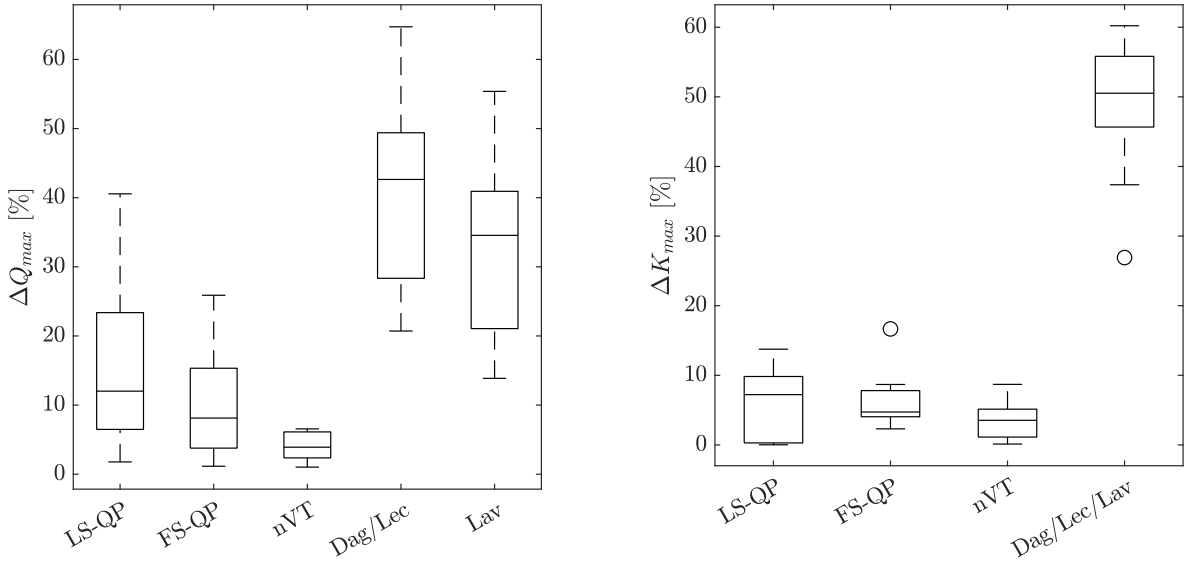
(a) Q_{max} from CTM simulations with varying turning ratios.

(b) K_{max} from CTM simulations with varying turning ratios.

Figure 5.12: Box-and-whisker diagrams showing the variability of Q_{max} and K_{max} across the evaluated scenarios.

To further assess the nMC, we calculate the absolute relative differences ΔK_{max} and ΔQ_{max} for the MFD approximations from both the proposed framework and the state of the

art, compared to the ground truth. For example, the absolute relative difference ΔQ_{max} between the nVT approach and the CTM can be calculated as $\Delta Q_{max} = |\frac{Q_{max,nVT} - Q_{max,CTM}}{Q_{max,CTM}}|$. Analogous calculations are performed to derive the other absolute relative differences. Note that Q_{max} and K_{max} for all state-of-the-art methods remain invariant across the scenarios since they do not account for turning ratios.



(a) Absolute value of the relative difference ΔQ_{max}.

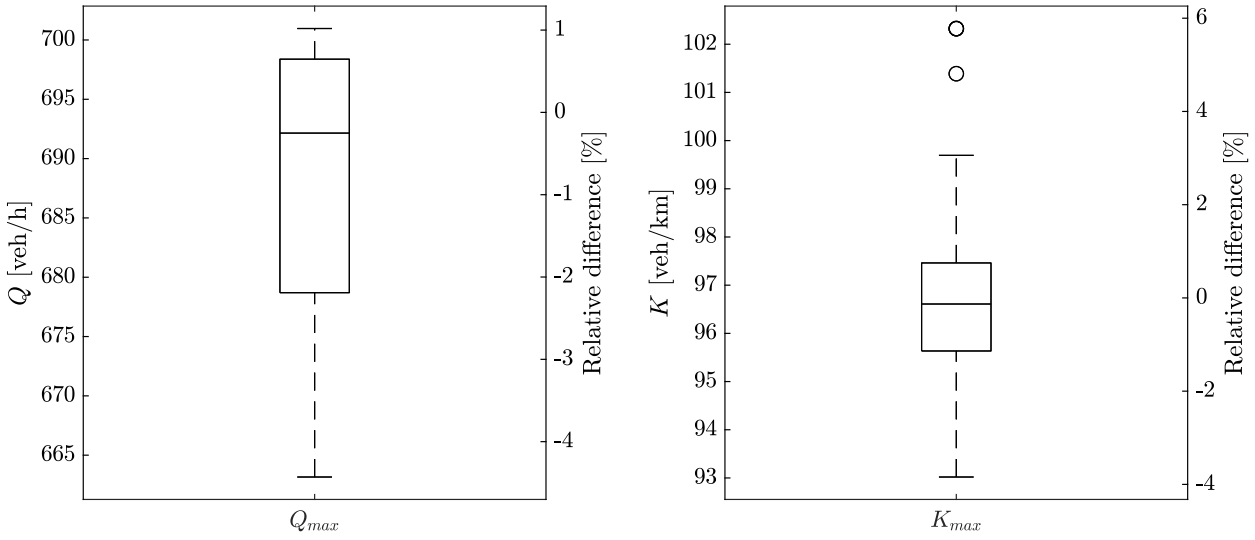
(b) Absolute value of the relative difference ΔK_{max}.

Figure 5.13: Analysis of the estimated network-wide capacity Q_{max} and jam density K_{max}. The evaluated approaches are the proposed framework, i.e. LS-QP, FS-QP, and nVT, as well as the state of the art, i.e. Dag, Lec, and Lav.

Figure 5.13 summarizes the results. Figure 5.13a shows the difference in the capacity approximation from the proposed LS-QP, the FS-QP, the nVT approach, and the state of the art compared to the ground truth values. The differences of the methods 'Dag' and 'Lec' do not vary since their estimated capacity is equal. The figure clearly shows the substantial improvement in the approximation accuracy regarding the network-wide capacity of all three proposed methods. Moreover, it is apparent that the FS-QP approach, which models network-wide spillback propagation in more detail, is more exact than the LS-QP approach, which does not account for such effects. Furthermore, the nVT approach is even more precise in estimating the network-wide capacity. The average capacity approximation of the proposed approaches is more than five times as accurate as the estimate by the state of the art. Additionally, we investigate the estimation of the network-wide maximum density K_{max} (see Figure 5.13b). While the average estimation of K_{max} by the nVT approach seems to be more accurate than the one by both approximate approaches, the differences are less obvious. Again, the proposed methods substantially improve the estimation of K_{max} compared to the state of the art, which assumes the jam density equal to the one of the link FD, i.e. $K_{max} = \kappa_{max}$. On average, the estimated K_{max} is more than five times closer to the

ground truth value than the estimate from the state of the art.

The analysis above showed that our proposed nMC framework has reasonable accuracy. It enables analyses that are not possible with other methods, such as the CTM, due to their high computational cost. To showcase this, we further examine the robustness of the estimated MFD with regard to small deviations of the turning ratios. To analyze the effect of small deviations of turning ratios, we first define a reference set $\mathcal{A}$ and then introduce random variations with a maximum of $\pm 5\,\%$ relative turning flows. We sample 100 sets with different variations and evaluate the MFD for each of them based on the FS-QP approach. The results are displayed in Figure 5.14 as box-and-whisker diagrams for Q_{max} and K_{max}, including the relative differences to values referring to the reference set $\mathcal{A}$. The values of



(a) Q_{max} from FS-QP-based MFD approximation with varying turning ratios.

(b) K_{max} from FS-QP-based MFD approximation with varying turning ratios.

Figure 5.14: Variability of Q_{max} and K_{max} for the evaluated scenarios with small deviations of turning ratios.

the capacity Q_{max} lie in a range of about 40 veh/h. Compared to the results displayed in Figure 5.12 which correspond to a strong variation of turning ratios, this is clearly a small range of variation. This is further highlighted by the relative differences which are below 4 %. However, it is apparent that the capacity of the estimated MFD is sensitive to turning ratios. Fortunately, due to the low computational cost of the proposed approach, one can increase the robustness of the results by simply introducing small deviations in the turning ratios and then taking the average capacity Q_{max}. Regarding the network-wide jam density, the absolute variation lies in a range of 5 veh/km. Considering that the relative differences are also below 4 %, except for two outliers, the sensitivity of the network jam density is similar to the capacity. Nevertheless, the robustness can also be increased by taking the average from an array of estimated values. These results can indicate how often the MFD has to be estimated during the day. More specifically, evaluating our proposed framework for

specific turning ratios enables one to quantify how small deviations in the turning ratios affect the MFD. Thereby, one can identify the maximum variation of turning ratios for a specific intersection, or multiple ones, that corresponds to a given admissible variation of the MFD. This, in turn, allows one to identify the times of the day when the MFD has to be re-estimated to reflect variations in the turning ratios. This example further confirms that spatial demand patterns indeed impact the MFD and highlights the value the proposed framework provides for the analysis of urban traffic.

5.6 Conclusion

In this chapter, we propose a framework, the nMC, to approximate the network-wide realized MFD without relying on extensive empirical data sets or microscopic simulations. Its flexibility allows to trade-off modeling complexity and computational cost against approximation accuracy. We base our framework on the well-established MC [DAGANZO & GEROLIMINIS, 2008; LECLERCQ & GEROLIMINIS, 2013] and follow its philosophy while exogenously accounting for the effects of source terms at intersections. The framework consists of two main steps after the problem initialization. First, we build a hypernetwork that reflects the network flows in the capacity state. For this purpose, we propose three different methods including two approximate and a precise one based on Chapter 4. The hypernetwork is the basis to derive the network MFD. Similar to the original MC, we derive cuts to approximate the free-flow branch of the MFD. Moreover, the hypernetwork allows us to efficiently estimate the network capacity. Last, we estimate the congested branch of the network MFD by approximating the maximum density in the network during gridlock and utilizing symmetries of the underlying link FD. We thoroughly investigate the proposed framework in a case study. This includes the comparison to a ground truth provided by the CTM, as well as to the original MC [DAGANZO & GEROLIMINIS, 2008; LECLERCQ & GEROLIMINIS, 2013] and the SA [LAVAL & CASTRILLÓN, 2015] as state of the art for realistic networks. Furthermore, we analyze our framework's robustness to different spatial demand patterns. The results demonstrate that the proposed framework enables one to estimate the realized network MFD sufficiently well for a realistic network. Moreover, they evidently show the improvement of the approximation accuracy of the network-wide capacity and jam density by our proposed framework compared to existing methods.

Our methodology allows the approximation of the realized MFD for realistic urban networks without the information loss induced by the reduction of networks to a single corridor as implicitly done by existing methods. Therefore, our framework is able to account for different spatial demand patterns and determine the upper bound of the network-wide average flow. The advantages of our model regard the flexible deployment, the low computational cost, and the explanatory value regarding the effects of important assumptions. Following the empirical study of AMBÜHL ET AL. [2021] and the theoretical work by LECLERCQ & PAIPURI [2019] which establish the demand-dependency of the MFD from a temporal perspective, our framework is a methodological proof that the effects of the spatial aspects of demand on the realized MFD cannot be neglected as well.

Future work includes the derivation of the idealized MFD based on our proposed framework since it corresponds to the system optimum which is relevant to a wide array of applications. Moreover, multi-modal aspects can be incorporated by modeling the effects of buses onto traffic flow and integrating a passenger model, leading to an estimation of the three-dimensional MFD [GEROLIMINIS ET AL., 2014]. These aspects are thoroughly discussed in the following chapters.

Chapter 6

Discussion

This chapter discusses the overall results from Chapters 3 to 5 apart from the numerical case-study specific outcomes which were examined in the corresponding sections. The structure of this section follows the 'life-cycle' of a model, starting from its design, then covering its verification and validation, and eventually illustrating its applications. Thereby, we discuss the proposed framework's limitations and interpret our results in light of existing literature and the greater context to further highlight the value that our developed methodologies add to the current state of the art.

6.1 Model design

6.1.1 Network variational theory

The current state of the art regarding VT-based traffic modeling is not able to describe traffic dynamics in urban networks. Since only a single spatial dimension can be considered, source terms would be required to be integrated within the solution space to model inflows and outflows at intersections. However, they violate the conservation of flow in the general case. To our best knowledge, only LAVAL ET AL. [2016] studied the integration of source terms into VT. Yet, they limited the integration to continuous source terms that are larger than zero. Therefore, they did not cover the case of urban intersections where inflows and outflows depend on signal phases and are indeed zero during red times. The proposed nVT, described in Chapter 4, fills this research gap by integrating source terms that represent inflows and outflows at signalized intersections. The core of the designed framework is its capability to model network-wide traffic dynamics by decomposing the network into a set of corridors and accounting for inflows and outflows at intersections of each corridor. Hence, our framework extends the original concept of VT from the corridor to the network level and models the propagation of free-flow and congested traffic states throughout the entire network.

Nevertheless, a number of assumptions are included in the overall design which leads to some limitations of the applicability of the model. Most importantly, VT and thus nVT is based on KWT. The underlying first-order PDE was originally developed to replicate traffic on highways. However, the corresponding solution methods such as the CTM and LTM are

also widely applied to urban areas [e.g. HAN ET AL., 2015; LONG ET AL., 2011]. Therefore, we assume that no major restrictions arise due to KWT being the theoretical foundation of our proposed models. Related to that, the FD is required to be concave and piecewise differentiable. If the FD has a triangular shape, this further leads to a convenient variational graph that facilitates the solution process of nVT. Also, bottlenecks and source terms can only be considered if they are fully mapped in the variational graph. This is because such aspects need to be included explicitly in the solution space. By adapting the time-step length, the accuracy in mapping these aspects can be controlled. Traffic-dependent bottlenecks, e.g. buses or slow-moving trucks, are more difficult to include but theoretically possible. Potential ways to consider them can be designed as a hybrid modeling approach, where the trajectories of the moving bottlenecks are derived from prevalent traffic conditions of the previous time-step [e.g. LAVAL & DAGANZO, 2006; TILG ET AL., 2018].

Another assumption is that vehicles follow a FIFO queueing discipline on links and at diverges. The implementation of other queueing disciplines might not be trivial, but of interest for complex intersection typologies. For such cases, the division of the intersection typology into a system of diverges can be a potential way to account for such disciplines. For example, a link with a turning lane can be modeled by splitting the representation of the link in VT at the beginning of the turning lane. Thereby, an additional node is generated that can be modeled as a diverge. Thus, more complicated queueing orders can readily be implemented.

Furthermore, limitations of the nVT framework result from the considered network elements. More specifically, we do not account for unprotected turns that occur, for example, at intersections where left-turners do not obtain a dedicated signal phase. From a theoretical perspective, such flows can be considered based on merge models [DAGANZO, 1995] and headway-based considerations [HERZ ET AL., 1976]. The merge model is necessary to describe how the left-turners and the right-turners from the opposing travel direction affect each other on the downstream link. The implementation of a merge model into our proposed framework could potentially be based on the capacity-related term of the nVT equations. Headway-based considerations are required because left-turners need to cross the conflicting traffic stream and the corresponding capacities need to be estimated. These estimates could again be integrated into the capacity-related term of the nVT equations. We conjecture that unprotected turns have low impacts if the absolute number of such movements is low. Furthermore, we assume that the turns would be protected and turning lanes exist if the flows are high. Hence, we adjudged the integration of unprotected turns as out of scope for this dissertation. Nevertheless, challenging cases occur when turning flows are significantly high such that the effects are observable, but no protected signal phases exist. Therefore, future work should indeed focus on integrating a merge model and capacity values which reflect conflicting streams as described above. This would further enable the modeling of fully unsignalized intersections and roundabouts.

Also specific to modeling network traffic in nVT, we assume turning ratios are constant during the entire computation period, for the sake of simplicity. Such turning ratios could be estimated from a higher-level traffic assignment model. This could result from a coupling framework of nVT and such model. Also, turning ratios can be estimated based on empirical

data, e.g. FCD, which is becoming increasingly available [e.g. BARMPOUNAKIS & GEROLIMINIS, 2020]. Varying ratios can potentially be included in the modeling framework. More specifically, the mathematical consideration of turning ratios (see eq.(4.7) in Section 4.3.2) assumes the turning ratio to be merely constant for a couple of time-steps. Therefore, a promising approach can be based on the notion of moving averages for modeling such varying turning ratios. Since the existing formulations would still apply, this would only require an additional computational layer to calculate the moving average in each time step.

6.1.2 Network method of cuts

In Chapter 5, we introduced the nMC to semi-analytically estimate the realized network MFD. The current state of the art reduces the network's complexity drastically, often only evaluating a representative corridor [e.g. DAGANZO & GEROLIMINIS, 2008; LECLERCQ & GEROLIMINIS, 2013; LAVAL & CASTRILLÓN, 2015]. Moreover, spatial demand patterns in networks are neglected by the majority of methods. XU ET AL. [2020] considered a network-wide average turning ratio in their extension of the SA. GEROLIMINIS & BOYACI [2012] focused only on specific corridors and did not fully integrate the effects of turning ratios. To fill this research gap, we proposed several different approaches to account for the impacts of network-wide traffic dynamics on capacities and the network jam density. Our framework does not reduce entire networks to single corridors and considers the effects of spatial demand patterns. The case study results demonstrated the substantial improvement of the accuracy of the MFD approximation of our method compared to the current state of the art.

The assumptions of the framework are partly similar to the ones of nVT. This regards the triangular FD, that bottlenecks need to be included in the variational graph, and the FIFO discipline on links and at diverges. Additionally, when the nVT approach is utilized within the nMC framework, the corresponding assumptions apply. The analytical approaches related to the derivation of the network-wide capacity and jam density make additional assumptions. More specifically, the LS and FS approach simplify the FIFO diverging behavior. Moreover, the LS approach does not propagate spillbacks across intersections, while the FS approach does but in a simplified manner. The QP approach simplifies the inflow into a link which affects the growth rate of the queue. While all these assumptions may impact the final MFDs, they are an explicit aspect of the overall framework. In other words, by applying the different approaches to approximate the network-wide capacity and jam density, one can directly compare the impacts of the related assumptions. Therefore, the framework offers the possibility to assess these assumptions and thus shed light on the corresponding impact for a specific problem setting. Next to the explanatory value this provides, the accompanying flexibility allows the modeler to trade-off estimation accuracy, modeling complexity, and computational effort for the problem at hand.

Important to note is that the concept of the MFD itself refers to the existence and uniqueness of stationary traffic states, similar to the link FD. The same assumption is made by the original MC. Some of the developed approaches of the nMC framework inherit this assumption. Moreover, it is worth noting that the assumption of stationary states, which correspond to an equilibrium state, is made beyond the field of MFD research. For example, in transport

planning one commonly assumes that equilibrium states can be reached [e.g. LEBLANC, 1975]. We assume the existence of stationary states for any spatial demand pattern when the temporal demand profile is constant. Recall that the MFD refers to a network with finite length and traffic states as well as shock waves travel with finite speeds. Thus, shock waves reach the boundaries of the network within finite time. Since shock waves represent changing traffic states, equilibrium at the network level is reached once all shock waves arrive at the boundary. The state of equilibrium is only prevented when temporal demand changes occur faster than shock waves reach the network boundary [e.g. LECLERCQ & PAIPURI, 2019; DAGANZO, 2007]. While the existence of stationary states can reasonably be assumed, the uniqueness of such is less clear. In particular, JIN ET AL. [2013] found that congested stationary traffic states might not be unique in simple networks (e.g. two-ring networks) as they depend on the initial distribution of vehicles. In the proposed nMC framework, we merely consider two different initial states. More specifically, to derive the network capacity, we assume an initially empty network. To derive the network-wide jam density, we assume the network is in the capacity state. Moreover, the employed node models are deterministic. Therefore, we conjecture the approximated MFD by the nMC to be unique. Nonetheless, fast-varying demand leads to non-stationary traffic states and hence unique traffic states cannot reasonably be assumed. In summary, the application of the nMC or the original MC, and the utilization of the MFD for region-based traffic modeling demands to keep in mind that the results are only valid when stationary traffic states exist. This might not be necessarily the case in reality as the observed traffic states might deviate from the semi-analytical estimated ones due to, among others, fast-varying demand profiles. The corresponding phenomenon is called hysteresis and emerges as *loops* in the (Q, K) plane. More information on its characteristics is provided in Chapter 1. Unfortunately, the nMC as well as other existing MC-based approximation methods cannot directly model hysteresis effects. Recall that such an ability would be beneficial, as hysteresis phenomena in urban areas are able to explain scatter in the realized MFD. However, semi-analytical methods based on the MC inherently seek an upper limit of traffic flow and can therefore not estimate hysteresis loops. Yet, the semi-analytical MFD provides valuable information. In particular, it enables to specify the maximum outflow of the given network which is a key element for models which characterize hysteresis phenomena, i.e. the system's loading and unloading during peak hours [e.g. MARIOTTE ET AL., 2017].

6.2 Model verification and validation

Model verification and validation are key to this dissertation since its main contributions in chapters 3 to 5 directly relate to these processes. RAKHA, HELLINGA, VAN AERDE & PEREZ [1996] stated that the effective use of a traffic model requires its validity for the application, as well as its ability to provide credible and reliable results. To ensure this, a model needs to be verified, validated, and calibrated. The authors assigned the former two processes to the model developer, i.e. the model verification and validation. Thus, these are the ones interesting for this dissertation. The calibration is highly important for the final application of a traffic model in the context of a real-world scenario. Since this is beyond the scope of this

dissertation, this step is therefore not addressed in this section.

RAKHA ET AL. [1996] defined the model verification as '*the process of determining if the logic that describes the underlying mechanics of the model [...] is faithfully captured by the computer code*'. In this dissertation, several applied models were verified. For the original MC and the SA applied in the comparative study in Chapter 3, the codes were obtained from the authors of the original works, and their technical correctness was confirmed based on the examples provided in the corresponding papers [LECLERCQ & GEROLIMINIS, 2013; LAVAL & CASTRILLÓN, 2015]. The development of nVT included the verification examples concerning the propagation of free-flow and congested traffic states for a simple intersection (see Chapter 4), where the effective inheritance of properties of the original VT was confirmed. The nMC in Chapter 5 builds partly on the verified nVT. For the additionally included models, the verification is performed along with the validation in the case study. Thus, we claim all models applied and developed in this dissertation as successfully verified.

Furthermore, RAKHA ET AL. [1996] defined the model validation as '*the process of determining to what extent the model's underlying fundamental rules and relationships are able to adequately capture the targeted emergent behavior, as specified within the relevant theory and as demonstrated by field data*'. Note that this definition of 'validation' refers to validating the suitability of the proposed model for representing the real behavior and traffic dynamics, and not to the validation after the calibration process. Again, substantial efforts were conducted to perform this validation process within the dissertation. The study presented in Chapter 3 where we compare the MC and SA to empirical data is per se a validation study. Thus, we inherently perform the validation and a subsequent evaluation of the corresponding results. The two novel methodologies introduced in this dissertation, the nVT and nMC, were both validated in a Sioux Falls case study. While the former was compared to a microscopic simulation, and the latter to the CTM, both studies aimed at validating the models' consistency with KWT at the network level. In both cases, these validation processes were performed for several demand scenarios to increase the quality of the validation. However, we did not validate the models against empirical data yet and consider this out of scope. RAKHA ET AL. [1996] confirmed that validation against a traffic simulation or analytical model where the underlying mechanics are known and can be controlled is the most important and additionally most informative part of the overall validation process. Nevertheless, future work should aim at including an empirical validation of the nVT and the nMC. This would allow evaluating potential mismatches of the predicted MFD and the empirical one, and thus would enable considering the corresponding errors during the practical application of the proposed models.

The validation results highly depend on the chosen ground truth, which underlines the importance of such for the general interpretability of the results. While conducting model validation for the studies presented in Chapters 3 to 5, we observed the crucial role of the ground truth for the overall results' interpretation. Empirical data is seemingly the first choice, as a model's ultimate goal should be the representation of reality. However, the growing body of literature on biases in empirical data and their impacts on the MFD estimation demonstrates that also an empirical ground truth needs to be treated with care (see Chapter 1). We followed these considerations explicitly for the comparison of existing

semi-analytical methods for MFD approximation as reported in Chapter 3. A ground truth can also be obtained from simulation. Yet, it might not be the unadulterated truth either. The validation of the nVT revealed that simple differences arising from the microscopic simulation being of a different nature than the macroscopic model, such as VT, can lead to errors, even though both are representations of KWT. For example, discrete vehicles are not divisible, while a macroscopic flow can be a floating decimal. While this issue does not arise when validating with a macroscopic model such as the CTM, as performed in Chapter 5, one should keep in mind that simulations often include errors themselves. In the case of the CTM, this is the numerical viscosity (see Chapter 2). After all, these observations show that while the validation against a ground truth is highly important and rightly best-practice, the final results still require cautious interpretation. It is crucial to bear in mind that the ground truth itself can be subject to uncertainty.

6.3 Model applications

6.3.1 Network variational theory

As mentioned above, the main application of nVT is to solve network KWT problems. For specific application requirements, other solution methods than the nVT might be more suitable, e.g. CTM, LTM, and Lagrangian VT. This is not surprising since models in general focus on certain aspects of the reality they aim to reproduce, and thus different models for different use cases exist. An example where nVT might not be the right choice is the modeling of large freeway networks. For these cases, the LTM might be more beneficial due to its extremely low computational cost, and the fact that merges are not fully supported within nVT yet.

Nevertheless, the properties and capabilities of nVT make it advantageous for an array of applications. In Chapter 5, we demonstrated the fact that the proposed nVT framework inherits important properties of the classical VT concept, such as the numerical precision and the ability to model complex intra-link heterogeneities. The former aspect makes the nVT superior to the CTM when both accuracy and computational cost are of importance. This was illustrated in a case study where the nVT based approach to derive the MFD was compared to a CTM ground truth. The latter inherited aspect is beneficial for use cases, where multi-modal interactions shall be examined at the network level. This relates to frequently but randomly occurring pedestrian crossings on certain network segments [e.g. KNOOP & DAGANZO, 2018] or public transport operation [e.g. GULER ET AL., 2016]. The corresponding stationary and moving bottlenecks can easily be implemented in the variational graph and thereby considered in the traffic state estimation. This is advantageous in comparison to other solution methods where the implementation of such bottlenecks is more cumbersome, e.g. the LTM or the Lagrangian VT formulation. Also, the computational effort of the proposed extension scales similar to the original concept with space and time, and the density of the numerical grid. In Section 4.6, we were able to demonstrate the beneficial use of nVT compared to a microscopic simulation. More specifically, there are magnitudes of efficiency gains from a computational cost perspective since microscopic

models usually scale with the number of simulated vehicles. This advantage becomes even more convincing having the previously described aspects in mind, which separates nVT from other similarly efficient macroscopic models.

As the last point, we would like to emphasize the compatibility of our extension with other works on VT, some of which are mentioned in Section 4.2. Apart from the MFD approximation thoroughly studied in this dissertation, nVT's fields of application range from data fusion [MEHRAN ET AL., 2012; SUN & BAN, 2013] to multi-modal urban traffic analyses [GULER ET AL., 2016; WU ET AL., 2017] as these can also be formulated as a KWT problem. Therefore, the nVT does not only extend the current state of the art but also the field of applications of methods that utilize the classical VT formulation. For example, VT-based data fusion algorithms can now also be extended to the network level.

6.3.2 Network method of cuts

The main field of application for the nMC is clear: the approximation of the network realized MFD. While this might sound overly specific at the first glimpse, it can contribute quite substantially to the field of traffic flow theory. Recall that the semi-analytical MFD currently estimated by the original MC is an input to many applications. For example, trip-based models need a speed-density relation as input that is currently often provided by the MC [MARIOTTE ET AL., 2017]. However, as such relation corresponds to a reservoir that represents a network and not a single corridor, the application of nMC to estimate such relation leads to an important increase in logical consistency of this modeling approach. As indicated above, the trip-based models allow accounting for hysteresis, and other aspects of urban traffic flow, such as fast-varying demands, which would violate one of the main assumptions of the accumulation-based notion of the MFD (see Section 1.1.4). Furthermore, the functional form for the MFD as proposed by AMBÜHL ET AL. [2020] takes a semi-analytical approximated MFD as input. Using the nMC to provide such again leads to a higher logical consistency, as the functional form is usually estimated at the network level. Similar to that, other fields (e.g. traffic management and control, transportation planning) where the original MC provides input but the application refers to networks can profit from higher logical consistency by MFD approximations provided by the nMC. Moreover, the explanatory value provided by the nMC framework can shed light on the network-wide impact of local bottlenecks with respect to a specific spatial demand pattern and thereby increase our understanding of the MFD itself.

Next to increasing our understanding of network traffic flow, the nMC can be of use for the design of traffic management strategies or transportation systems. Its comparably low computational cost allows its integration into real-time applications, or simulation-based optimization to design urban traffic systems. First studies already explore the integration of the MFD for such purposes, for example, the design of signal coordination, or the evaluation of route guidance schemes (see Section 1.1.4). Furthermore, our proposed methodology can be integrated into an optimization framework to find the idealized MFD considering spatial demand patterns at the network level. By defining the turning ratios as decision variables and the capacity as objective value, one could find the spatial demand pattern that matches the supply to maximize average flows. Applying the nMC within the optimization

further highlights the benefit of low computational costs. Note that the maximum average flows for a given density lead to the maximum average speeds, and therefore to the shortest average travel times in the considered network. Therefore, the idealized MFD can be interpreted as the system optimum which is interesting to many applications in the overall field of transportation [WARDROP, 1952].

6.3.3 Coupling of both proposed frameworks

After emphasizing the range of applications for both, the nVT and nMC, we want to highlight the advantages which arise from integrating the nVT into the nMC framework. In particular, this integration allows the MFD approximation to profit from potential nVT extensions. For example, nVT could account for stochastic bus arrivals as proposed by DAKIC ET AL. [2020] for classical VT. Then, nVT would allow estimating the effects of multi-modal interactions on network-wide traffic flows. Furthermore, the integration of continuous exogenous source terms as proposed by LAVAL ET AL. [2016] would be interesting to represent trip generation and completion within the network. Since the nMC framework can utilize nVT, the resulting framework would allow accounting for multi-modal interactions and trip generation and completion when estimating the realized network MFD. Hence, more and more assumptions of current semi-analytical approximation methods could be relaxed. Moreover, a potential passenger routing model could be developed by defining OD pairs in a given network and deriving travel times based on the approximated MFD, if a public transport system including bus routes and a time-table is given. The combination of the semi-analytical multi-modal MFD approximation with a passenger routing model could eventually result in a larger framework to account for mode choice and to eventually derive the three-dimensional passenger MFD [GEROLIMINIS ET AL., 2014]. This is evidently interesting for future work as the overall perspective in traffic engineering and planning is being shifted from vehicles towards persons and multi-modality.

6.3.4 Physics-based vs. data-driven models

Last, we want to draw attention to the application of the proposed frameworks in the context of traffic state estimation and prediction. Since the MFD is a concept to estimate and predict aggregated traffic states at the network level, it can be applied for such purposes. The semi-analytical approximation of the realized network MFD as proposed in this dissertation, as well as the nVT as a solution method for network KWT problems, can be categorized as physics-based models. This is due to the fact that their key elements represent the laws of nature, for example, the occurrence of shock waves and the existence of a fundamental diagram. While a certain amount of data is required for calibration and validation of these models, they cannot exploit huge traffic data sets.

In case that such data are available, methods based on statistical and machine learning concepts might be more suitable. These methods are often categorized as ‘data-driven’ to illustrate the extensive reliance on the underlying data set. For the case that a large amount

of traffic data is available, the nMC and nVT might become obsolete in the context of traffic state estimation, as potential machine learning-based approaches might be more accurate, or a simple empirical MFD estimation could be sufficient.

In the literature, it is stated that one of the advantages of data-driven techniques is the low number of involved assumptions. As an example in the context of traffic state estimation, they do not assume an existing FD. For example, ANTONIOU, KOUTSOPOULOS & YANNIS [2013] utilized a data-driven approach to predict traffic states, and proposed to replace traditional FD in a traditional traffic simulation with their estimated speed-density functions. Also, MA, ANTONIOU & TOLEDO [2020] proposed an approach based on machine learning and statistical methods to estimate network-wide traffic states. Nevertheless, more general critics often refer to the strong dependency on historical data of these methods [SEO ET AL., 2017]. For example, if special or unexpected events occur which are not represented in the training data, these methods often fail to predict or estimate with precision. Furthermore, they are sometimes considered as a black box as their explanatory value is low [SEO ET AL., 2017]. VLAHOGIANNI, KARLAFTIS & GOLIAS [2014] identified numerous challenges for data-driven forecast techniques. Amongst them is the development of algorithms that are responsive to unexpected events, such as bad weather conditions or accidents. Moreover, they noted that the research on traffic state prediction for urban settings including arterials and corridors is still at an early stage. Also, they found the effects of control schemes, such as traffic signals, to impede the use of statistical models. The authors recommended a combination of statistical methods with classical traffic flow models to overcome such issues.

Recently, first study results on such a successful coupling of both model worlds have been published. In the context of the MFD, an example has been reported in SAFFARI, YILDIRIMOGLU & HICKMAN [2020]. The authors utilized principal component analysis to reduce the necessary data for the MFD estimation. Moreover, the combination of both model families is presented in the very recent research trend of so-called physics-informed deep learning models. For example, SHI, MO & DI [2021] presented an approach to couple KWT-based traffic models with deep learning to estimate traffic states for a simple highway. Also, REMPE, LODER & BOGENBERGER [2021] applied a physics-informed model for fusing LDD and FCD to estimate traffic states. These first studies indicate that coupling deep learning methods and physics-based models is technically feasible. Therefore, combining the developed nVT and nMC with a deep neural network could result in a framework that can potentially account for unexpected events as well as for the impact of traffic signal control, but at the same time exploit large traffic data sets. Hence, we see this topic as an interesting line of future research.

Chapter 7

Conclusion

Urban capacities are being pushed to their limits. This inevitably leads to congestion and a related increase of other external costs of traffic. Therefore, efficient and parsimonious traffic models are in need since they can help to design effective traffic management strategies to mitigate congestion. It is increasingly acknowledged by the research community that the macroscopic fundamental diagram (MFD) represents such a model. In this dissertation, we contribute to its usability by developing a framework for its semi-analytical approximation. Thereby, we distinguish between the idealized and realized MFD. The idealized MFD describes the maximum average network-wide flow at a given density independently of the demand. The realized one refers to maximum flows for a given density and a specific demand pattern. However, the current MFD estimation techniques do not suffice at the network level in case no extensive database or simulation model exists. For such cases, semi-analytical methods can be of advantage due to their low input data requirements and computational efficiency. Furthermore, they often add explanatory value and thus can help to facilitate the understanding of urban traffic phenomena. Yet, the available semi-analytical methods to approximate the MFD are limited to the corridor level and do not apply to urban networks. Also, they only refer to the idealized MFD and can thus not account for a specific spatial demand pattern. To address this research gap, this dissertation formulates the following overarching research question:

Can the realized MFD be semi-analytically approximated for urban networks?

Indeed, we can affirm this based on the successful application of the proposed network method of cuts (nMC) framework, which semi-analytically approximates the network MFD, in a case study for the Sioux Falls network. Thus, this dissertation strengthens the relevance of the semi-analytical MFD approximation for modeling, monitoring, and control of traffic in urban networks. Additionally, the thesis provides multiple contributions to the field that lie apart from the pure approximation of the MFD and concern more general macroscopic traffic flow modeling.

Section 7.1 summarizes the thesis, and relates the findings to the research questions **RQ1**-**RQ6** as specified in Section 1.2.1. Afterward, we highlight the implications of the work in Section 7.2. Then, Section 7.3 presents the limitations. Finally, Section 7.4 gives an outlook for future work based on the obtained results and the remaining open questions.

7.1 Summary

Motivated by the wide range of applications of the MFD, we conduct an in-depth analysis of corresponding semi-analytical approximation methods in Chapter 3. More specifically, we analyze the method of cuts (MC) [LECLERCQ & GEROLIMINIS, 2013] and the stochastic approximation (SA) [LAVAL & CASTRILLÓN, 2015] to see whether a deterministic or stochastic approach is more suitable for a realistic urban setting. Both were derived from the original method of practical cuts introduced by DAGANZO & GEROLIMINIS [2008] and each extends certain aspects of the original one. The MC is deterministic and applies to irregular corridors, i.e. with varying block lengths and signal settings at intersections. It is based on a sufficient variational graph that is conjectured to lead to a tight MFD estimation. The SA allows the explicit consideration of stochastic effects and, thus, is expected to perform better for corresponding problem settings. Despite both methods conceptually merely referring to the corridor level, researchers applied them to simple networks in the past [e.g. LECLERCQ & GEROLIMINIS, 2013; GIRAULT ET AL., 2016; LAVAL & CASTRILLÓN, 2015]. Other extensions of the method of practical cuts focus on improving the mathematical formulation and do not substantially modify the underlying concept [e.g. DAGANZO & LEHE, 2016]. Therefore, we identify the MC and the SA as the state of the art in the semi-analytical MFD approximation.

The in-depth analysis concerns a segment of an urban corridor in Munich, namely the Leopoldstraße. The corresponding empirical data sets are traffic counts from loop detectors and traffic signal phase time stamps. These data sets enable the estimation of the empirical MFD as ground truth, the derivation of the input parameters for the MC and the SA, and the analysis of the sensitivity of the MFD approximation results to input parameters related to signal control and public transport operation. The empirical estimation is conducted based on the state of the art and includes the examination of the involved loop detector placement bias. The sensitivity study is performed as a Monte Carlo Simulation.

The MC matches the ground truth more closely than the SA from the overall perspective. This result indicates the affirmation that a deterministic approach to semi-analytically approximate the MFD is closer to reality for the studied case than a stochastic one (**RQ1**). Especially the capacity branch is estimated more accurately, which is key to many MFD-based applications. The identified reason is the occurrence of substantial differences in the capacity of single intersections within the corridor. The effect of single constraining intersections cannot be captured adequately by the SA due to underlying assumptions related to the probability distribution of green phase durations and cycle lengths. In contrast, the MC assumes that the corridor capacity is determined by the most constraining intersection, or even lower due to short blocks. Since the studied urban corridor indeed contains one constraining intersection, the MC estimates the MFD more precisely. However, the SA estimates the free-flow branch of the MFD more accurately which relates to traffic states during the loading phase of the corridor. Furthermore, the analysis of the placement bias in the empirical data reveals that the free-flow speed is indeed underestimated in the empirical MFD. Thus, the SA is even closer to reality during the loading of the corridor than the MC.

We conduct a systematic Monte Carlo Simulation study to analyze the sensitivity of these approaches to input parameters reflecting the complexity of urban settings (**RQ2**). The

sensitivity analysis shows a significant dependency of the MFD on the signal parameters. This is an expected result as it is widely acknowledged that urban capacities are driven by intersection control. Neither the MC- nor the SA-based MFD approximations are substantially affected by public transport operation-related parameters. This further confirms the importance of knowledge about signal control parameters for an accurate MFD approximation. The fact that the methods can be analyzed based on a Monte Carlo Simulation illustrates the low computational costs involved. Such a study could not efficiently be conducted based on microscopic simulation due to the involved computational burden.

Overall, the study shows the MC's superiority in estimating the MFD for the given case. Moreover, the deterministic nature of the MC adds explanatory value to the model by enabling the explicit consideration of characteristics of single intersections that dominate traffic dynamics in the entire corridor. In contrast, the stochastic nature of the SA is based on distributions of the input parameters and their first and second moments and, hence, not able to account for such single dominating intersections, as the analysis showed. Due to these reasons, the MC is chosen for being extended to the network level.

In Chapter 4, we develop a mathematical framework that constitutes an extension of variational theory (VT) from the corridor to the network level (**RQ3**), the network variational theory (nVT). The motivation for this development lies in the fact that the MC is largely based on the tenets of VT. This extension to networks facilitates the affirmation of the overarching research question of this dissertation. The classical VT as introduced by DAGANZO [2005b], DAGANZO [2005a] and DAGANZO & MENENDEZ [2005] only applies to the corridor level and not to entire networks. Therefore, this thesis follows the idea of decomposing a network into a set of corridors and connecting these at intersections by accounting for turning vehicles. However, one of the main assumptions of VT is the conservation of flows within the solution space. This implies the nonexistence of inflows and outflows, or, at least that the net transfer flows equal zero, and that their effects on traffic dynamics are negligible. This fact was confirmed only recently by LAVAL ET AL. [2016] who investigated the possibility of integrating such transfer flows into the concept of VT, which are mathematically represented as so-called source terms. The authors concluded that the endogenous consideration of source terms is neither possible in the Eulerian nor the Lagrangian VT formulation. Fortunately, they further found the Eulerian formulation to allow an exogenous consideration of source terms, and therefore the possibility to account for transfer flows. Nevertheless, the study did not investigate source terms representing the discrete change of transfer flows due to traffic signals which are of high importance in an urban setting. Addressing this research gap, we introduce a set of equations enabling the consideration of source terms representing transfer flows at signalized intersections. To describe network-wide traffic dynamics, we first decompose the network into a set of corridors and then model the evolution of corridor-specific traffic states while accounting for transfer flows with the developed mathematical framework.

We apply the proposed framework for a simple intersection to demonstrate its ability to model upstream and downstream traffic state propagation that is an important building block of network traffic modeling. To further assess the value of our framework, we compare it to other existing kinematic wave theory (KWT) solution methods as well as to a microscopic simulation (**RQ4**). This evaluation consists of presenting demonstrative examples, a case

study, as well as a thorough qualitative discussion. While we successfully extend VT to the network level and thereby increase the model's complexity, we can maintain important properties of the original formulation. This includes its numerical precision as well as its capability to model complex intra-link bottlenecks with ease. The former aspect makes VT superior to the widely known cell transmission model (CTM) which involves the numerical diffusion error, i.e. a smoothing of traffic densities near shock waves. The latter aspect makes it competitive to the link transmission model (LTM), which does not allow simple modeling of intra-link bottlenecks. Both aspects are demonstrated in Chapter 4 for a simple intersection. The numerical precision of VT is demonstrated by the fact that a reduction of the time-step length in the solution method does not affect the accuracy of the result. The ability to model intra-link bottlenecks is highlighted in an example by introducing many short bottlenecks randomly in space and time, and verifying the corresponding results in the time-space diagram. Furthermore, the nVT is applied for the Sioux Falls network with turning ratios at several intersections. The chosen ground truth is Newell's car-following model implemented in the microscopic simulation software SUMO. We demonstrate that the traffic states modeled by nVT match the ground truth at all intersections and for five different loading scenarios. All errors are below 5 % which is a satisfying result considering that a certain error is expected due to the different resolution scales of both methods. Moreover, the nVT outperforms the microscopic simulation from the perspective of computational cost. In summary, the nVT is a method that allows to solve network KWT problems and is superior to alternatives for specific use cases.

The development of the nVT is a cornerstone for our proposed extension of the MC to the network level in Chapter 5, the nMC (**RQ5**). We explicitly focus on the realized MFD instead of the idealized one and thereby are able to include the effects of specific spatial demand patterns. The proposed methodology follows the corresponding network decomposition approach to derive a set of corridors since it has proven to be useful for the nVT in Chapter 4. We then develop a framework that finds the MFD based on the network topology, the control settings, and spatial demand patterns described by turning ratios at intersections. For this purpose, we propose three different approaches to generate a so-called hypernetwork in analogy to the one-dimensional hyperlink from the original MC. These three approaches differ in modeling complexity, computational cost, and estimation accuracy. The most complex, costly, but accurate approach utilizes the nVT. The hypernetwork allows the consideration of the effects of turning flows and therefore the effects of network-wide spillbacks and flow conservation in the MFD approximation. Based on the hypernetwork, we derive the free-flow and capacity branch of the MFD. The capacity branch is approximated based on an estimation of the network-wide jam density and a density transformation as reported in LAVAL & CASTRILLÓN [2015] and DAGANZO & KNOOP [2016]. For the jam density approximation, we again propose a novel analytical approach and one based on nVT. Next to offering high flexibility for the model application, providing multiple options for the hypernetwork generation and jam density estimation can shed light on the related assumptions and thus adds explanatory value (**RQ6**).

We evaluate the nMC in a case study for the Sioux Falls network to provide a proof of concept by comparing the results to a CTM ground truth. Additionally, the proposed methodology is

compared to the state of the art as described by DAGANZO & GEROLIMINIS [2008], LECLERCQ & GEROLIMINIS [2013], and LAVAL & CASTRILLÓN [2015]. The methods are applied for a variety of different turning ratios to increase the robustness of the framework's validation. On average, the proposed nMC is five times more accurate concerning the network-wide capacity and maximum density estimation. These results fundamentally support the relevance of the nMC for the realized network MFD approximation. Furthermore, we apply the nMC to evaluate the effects of small deviations of the turning ratios on the MFD. The results clearly show that the realized MFD indeed depends on spatial demand patterns.

7.2 Implications

This dissertation provides a deeper insight into the semi-analytical approximation of the network MFD and related models. This concerns the quality of existing approaches to approximate the MFD at the corridor level, and the possibility to apply VT and an MC-based framework at the network level as outlined in Chapter 6.

The analysis of MC and SA presented in Chapter 3 extends our knowledge about the semi-analytical MFD approximation for realistic and highly complex urban corridors. While the SA results in a substantial overestimation of the corridor capacity, the MC is able to estimate the MFD significantly more accurately. Considering the assumptions of the underlying KWT as well the complex operation of traffic, i.e. multi-modal interactions and actuated traffic signal control, the derived results are quite satisfactory. They imply that the MC is certainly suitable to model and monitor traffic at an aggregated level. While finer changes in the signal timings might not be assessable, the effects of traffic management measures such as a change in the green-to-cycle ratio of specific intersections can indeed be evaluated based on the MC. The SA seems to have fewer benefits for the presented case. However, this is rather expected given the inability to account for a single dominating intersection as occurring in the analyzed case. While findings from a single case can hardly be generalized, the good fit of the MC to empirical data is confirmed by previous studies [e.g. DAGANZO & GEROLIMINIS, 2008]. Moreover, the semi-analytical nature of the model facilitates the revealing of the underlying reasons for the good fit, i.e. the consideration of a single dominant intersection in the corridor. We conjecture that the longer the corridor is and the more intersections are included, the more exact is the SA. The reason lies in the fact that in such a case single intersections play a minor role as turning flows at downstream intersections might outweigh capacity-induced flow reductions. The longer the corridor is, the higher is the probability that such flows exist. Thus, taking the average of intersection capacities, as the SA proposes, becomes closer to the real corridor capacity. For such cases, the SA is expected to excel due to its parsimony and effectiveness.

The nVT, explained in Chapter 4, represents an effective extension of the concept of VT for signalized urban networks. Given the usual supply- and demand-related input parameters, the evolution of traffic states can be modeled throughout the network and in accordance with KWT. In contrast to other well-established models, nVT allows one to consider heterogeneous complex problem settings including a time-space dependent fundamental diagram (FD),

moving bottlenecks, and randomly occurring short blockings. The review of VT in Section 4.3 illustrates the wide range of applications from data fusion to traffic state estimation and prediction. Apart from the network MFD approximation as proposed in this dissertation, this includes additional applications which may profit from the extension of VT to networks. For instance, nVT enables one to evaluate multi-modal traffic management schemes at the network level where the interaction between modes is of significance, e.g. the design and operation of intermittent bus lanes. Thanks to its modeling accuracy and computational efficiency, other use cases such as the incorporation in real-time simulation or the integration into a model-based optimization framework become interesting. In summary, the proposed extension of nVT applies to scenarios where a microscopic simulation is too expensive from a computational perspective, but the existence of intra-link bottlenecks such as buses or pedestrians hinders the efficient application of other macroscopic models such as CTM and LTM.

Finally, Chapter 5 integrates the knowledge obtained in the analysis of the MC and the SA, as well as the developed nVT, into a larger methodological framework to approximate the realized network MFD. While the network MFD was already approximated based on currently existing methods in the past, these methods reduce the network to a corridor and thereby neglect corresponding complexities. Our developed framework enables one to estimate the network MFD without the information loss induced by the reduction of the network at hand to a single 'representative' corridor. Therefore, our methodology accounts for important aspects of the network topology, as well as for spatial demand patterns represented by turning ratios at intersections. We incorporate several approaches of different complexity to approximate the MFD. Thus, the potential of the nMC lies in its flexibility. For example, the different approaches shed light on the impact of related assumptions on the network-wide traffic state propagation on the MFD. Hence, our developed framework can reveal new insights in the field of network traffic flow theory. Also, compared to simulation-based MFD estimation methods, our framework benefits from low computational costs that can be beneficial for a couple of applications. Furthermore, as it is among the first semi-analytical approaches to approximate the realized MFD and thus to account for the effects of heterogeneous spatial demand patterns on the network MFD, it can be seen as methodological proof that such effects cannot be neglected. Therefore, it becomes clear that the nMC is of practical use, as traffic management schemes and transport policies can be evaluated based on the realized MFDs, which refers to observations made in reality or simulation. Furthermore, the consideration of spatial demand aspects complements studies on temporal ones with similar conclusions [e.g. LECLERCQ & PAIPURI, 2019; AMBÜHL ET AL., 2021]. Next to this explanatory value on aggregated traffic state dynamics, our framework delivers important inputs for further applications such as aggregated traffic modeling with the trip-based approach [MARIOTTE ET AL., 2017; MARIOTTE & LECLERCQ, 2019; BATISTA, LECLERCQ & GEROLIMINIS, 2019]. In the end, the nMC contributes to hypothetical scenario analyses related to the design of transportation systems where the MFD is considered either a modeling tool or a performance indicator.

7.3 Limitations

Naturally, there exist limitations to the informative and explanatory value of this dissertation. These occur due to assumptions and simplifications associated with the utilized models and underlying theories. Hereafter, we list these limitations, where some of which can be addressed by future research efforts (see Section 7.4):

- **Limited empirical data:** The comparative analysis in Chapter 3 is based on empirical data. While traffic and signal data concerning several days are compared and adjudged as similar, the total data set does merely cover working days in the month of October. Seasonal and weekday effects can therefore not be investigated. However, the study of AMBÜHL ET AL. [2021] illustrated that such effects can be seen in empirical MFDs. From a spatial perspective, only one single corridor was investigated. Thus, future validation studies should focus on corresponding data sets to examine the performance of the MC and SA for such cases.

- **Traffic signal control:** The original MC refers only to fixed-time traffic signal control from a conceptual perspective. However, the investigated corridor in Chapter 3 was operated based on actuated signal control. This aspect is partly addressed by conducting the Monte Carlo Simulation, however, an adaptation of the MC towards more flexible control algorithms could shed more light on these aspects. For example, approaches based on stochastic variational graphs [e.g. DAKIC ET AL., 2020] could be employed to investigate related effects on the accuracy of the estimated MFD.

- **KWT:** The concepts of VT and the MC are based on KWT and correspondingly come with a set of assumptions. While some of them such as a triangular FD facilitate the overall modeling process, others are strictly necessary. These assumptions are inherited by our developed models, i.e. the nVT and the nMC. To further analyze the role they play for estimation accuracy, empirical validation studies can be the first step.

- **Hysteresis phenomenon:** The hysteresis phenomenon which is observed in empirical MFDs cannot be directly modeled with the proposed nMC. This is due to the modeling philosophy inherited from the original MC, i.e. the assumption that stationary traffic states can be reached. For use cases where this phenomenon is of interest, e.g. during times of fast-changing demand, the nMC can be coupled with other models (e.g. the trip-based MFD) to provide valuable information on urban traffic dynamics.

- **Model validation:** We validate the nVT and nMC with specific simulations in the corresponding case studies. The validation results might depend on the choice of simulation model. However, as the selected models represent KWT at the network level, we do not expect a substantial validity loss in case alternative ground truths are used.

7.4 Outlook

Despite these limitations, the semi-analytical approximation of the realized network MFD is certainly a step in the right direction to efficiently model, monitor, and control urban traffic. Moreover, it opens the door to interesting topics for future research, some of which we list in the following:

- **Validation with empirical data:** The proposed methodological extensions, the nVT and the nMC are both validated with KWT simulation models for the Sioux Falls network. The corresponding results confirm their consistency with the theory. Nevertheless, a validation with empirical data will shed light on the benefits of applying these frameworks in practice. The assumptions inherent to KWT are well known and are discussed in this dissertation in the context of the developed frameworks. Still, especially for the case of the nMC, additional assumptions regarding network loading and unloading apply. The related effects on the quality of the MFD approximation can best be evaluated by comparing it to empirical data since this would indicate its applicability for practical purposes. Additionally, such an analysis would allow the quantification of the impact of these assumptions in line with the study conducted in Chapter 3 for the currently available approximation methods.

- **Idealized MFD:** The proposed nMC approximates the realized MFD at the network level. However, the idealized MFD can potentially be derived as well and can be interesting as it corresponds to the system optimum [WARDROP, 1952]. More specifically, the nMC can be integrated into a simulation-based optimization framework where the turning ratios at intersections are defined as decision variables and the maximum capacity as objective value. This would lead to the spatial demand pattern matching the supply conditions to maximize average flows. This corresponds to maximum average speeds, and thus to shortest average travel times, which is relevant to the system optimum.

- **Public transport and the passenger perspective:** This dissertation focused on vehicular flows. However, as described above, the strength of VT is the consideration of complex bottlenecks. This ability is inherited by the nVT and the nMC. Being able to model the vehicular interactions between different modes allows investigating the related effects on traffic flows. Similar extensions of the MC already exist for the corridor level [DAKIC ET AL., 2020; CHIABAUT, 2015]. From the derived traffic states, travel times between certain origin-destination (OD) pairs can be estimated. This could offer the basic inputs to a passenger model which, of course, would require further modeling efforts. Nevertheless, a larger framework to include mode choice, and thus reasonably take the perspective of passenger flows could eventually be created. Hence, one could estimate the three-dimensional passenger MFD as introduced by GEROLIMINIS ET AL. [2014]. This is an interesting line of future research since the successful application of the passenger MFD can be beneficial for the design of multi-modal transport systems [e.g. TILG ET AL., 2020b].

- **Coupling of physics-based and data-driven models:** The proposed frameworks in this dissertation are based on physics-based models. The emergent increasing availability of traffic data led to a focus on data-driven models in the research community to estimate and predict traffic states. However, both modeling families have their strengths and weaknesses. Hence, future research should be undertaken to investigate the coupling of the MFD with statistical and machine learning methods, and the integration of the concept into physics-informed deep learning methods to connect the advantages from both worlds.

As a final statement, we want to highlight that while the MFD has a wide range of applications and offers novel ways to model, monitor, and control traffic, at the end of the day it relates mainly to capacity when used as a performance indicator. It shall therefore not remain the single indicator in traffic analyses, since other impacts especially those related to transport emissions should become more important in the light of climate change. Fortunately, first works are exploring the potential of MFD-based modeling with respect to such aspects [e.g. BATISTA, TILG & MENENDEZ, 2021a; BARMPOUNAKIS, MONTESINOS-FERRER, GONZALES & GEROLIMINIS, 2021]. Thus, in the long run, this dissertation desirably contributes to the emergence of more sustainable cities.

Bibliography

ABOUDOLAS, K. & GEROLIMINIS, N. [2013]: Perimeter and boundary flow control in multi-reservoir heterogeneous networks. In: *Transportation Research Part B: Methodological* 55, pp. 265–281. ISSN: 01912615. DOI: 10.1016/j.trb.2013.07.003.

AFRIN, T. & YODO, N. [2020]: A survey of road traffic congestion measures towards a sustainable and resilient transportation system. In: *Sustainability* 12.11, p. 4660.

AGHAMOHAMMADI, R. & LAVAL, J. [2021]: Macroscopic fundamental diagram parameter estimation: A maximum likelihood approach. In: *Presented at the 100th Transportation Research Board Annual Meeting*. Washington, D.C., USA.

AMBÜHL, L. & MENENDEZ, M. [2016]: Data fusion algorithm for macroscopic fundamental diagram estimation. In: *Transportation Research Part C: Emerging Technologies* 71, pp. 184–197. DOI: 10.1016/j.trc.2016.07.013.

AMBÜHL, L.; LODER, A.; MENENDEZ, M. & AXHAUSEN, K. W. [2017]: Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data. In: *Presented at the 96th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

AMBÜHL, L.; LODER, A.; BLIEMER, M. C.; MENENDEZ, M. & AXHAUSEN, K. W. [2018]: Introducing a resampling methodology for the estimation of empirical macroscopic fundamental diagrams. In: *Transportation Research Record: Journal of the Transportation Research Board* 2676.20, pp. 239–248.

AMBÜHL, L.; LODER, A.; ZHENG, N.; AXHAUSEN, K. W. & MENENDEZ, M. [2019]: Approximative network partitioning for MFDs from stationary sensor data. In: *Transportation Research Record: Journal of the Transportation Research Board* 2673.6, pp. 94–103. DOI: 10.1177/0361198119843264.

AMBÜHL, L.; LODER, A.; BLIEMER, M. C.; MENENDEZ, M. & AXHAUSEN, K. W. [2020]: A functional form with a physical meaning for the macroscopic fundamental diagram. In: *Transportation Research Part B: Methodological* 137, pp. 119–132. ISSN: 01912615. DOI: 10.1016/j.trb.2018.10.013.

AMBÜHL, L.; LODER, A.; LECLERCQ, L. & MENENDEZ, M. [2021]: Disentangling the city traffic rhythms: A longitudinal analysis of MFD patterns over a year. In: *Transportation Research Part C: Emerging Technologies* 126, p. 103065. DOI: 10.1016/j.trc.2021.103065.

AMELI, M.; LEBACQUE, J.-P. & LECLERCQ, L. [2020]: Simulation-based dynamic traffic assignment: Meta-heuristic solution methods with parallel computing. In: *Computer-Aided Civil and Infrastructure Engineering* 35.10, pp. 1047–1062. DOI: https://doi.org/10.1111/mice.12577.

AMINI, S. & TILG, G. [2018]: Estimating the effects of temporary bottlenecks on the capacity of urban arterial: An MFD-approach. In: *Presented at the 7th Symposium of the European Association for Research in Transportation*.

AMINI, S.; TILG, G. & BUSCH, F. [2018]: Evaluating the impact of real-time traffic control measures on the resilience of urban road networks. In: *2018 21th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 519–524. DOI: 10.1109/itsc.2018.8569678.

AMINI, S.; TILG, G.; CAGRI TEKIN, A.; TIDDI, D. & BUSCH, F. [2019a]: Automatic calibration of the link fundamental diagram for macroscopic traffic simulation models. In: *mobil.TUM 2019*. Munich, Germany.

AMINI, S.; TILG, G. & BUSCH, F. [2019b]: Calibration of mesoscopic simulation models for urban corridors based on the macroscopic fundamental diagram. In: *Presented at the 8th Symposium of the European Association for Research in Transportation*.

AMINI, S.; AMBÜHL, L.; TILG, G.; BOGENBERGER, K. & MENENDEZ, M. [2020a]: Generating and calibrating large-scale, mesoscopic SUMO networks. In: *SUMO User Conference 2020*.

AMINI, S.; TILG, G. & BUSCH, F. [2020b]: Macroscopic traffic dynamics in urban networks during incidents. In: *Presented at the 99th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

ANGEL, E. & BELLMAN, R. [1972]: *Dynamic programming and partial differential equations*. Academic press, New York.

ANTONIOU, C.; KOUTSOPOULOS, H. N. & YANNIS, G. [2013]: Dynamic data-driven local traffic state estimation and prediction. In: *Transportation Research Part C: Emerging Technologies* 34, pp. 89–107.

ARNOTT, R. [2013]: A bathtub model of downtown traffic congestion. In: *Journal of Urban Economics* 76, pp. 110–121. DOI: 10.1016/j.jue.2013.01.001.

BARMPOUNAKIS, E. & GEROLIMINIS, N. [2020]: On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. In: *Transportation Research Part C: Emerging Technologies* 111, pp. 50–71.

BARMPOUNAKIS, E.; MONTESINOS-FERRER, M.; GONZALES, E. J. & GEROLIMINIS, N. [2021]: Empirical investigation of the emission-macroscopic fundamental diagram. In: *Transportation Research Part D: Transport and Environment* 101, p. 103090. ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2021.103090.

BATISTA, S.; TILG, G. & MENENDEZ, M. [2021a]: Exploring the potential of aggregated traffic models for estimating network-wide emissions. In: *Transportation Research Part D: Transport and Environment*. Submitted for publication.

BATISTA, S. F. A.; LECLERCQ, L. & GEROLIMINIS, N. [2019]: Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. In: *Transportation Research Part B: Methodological* 122, pp. 192–217. ISSN: 0191-2615. DOI: doi.org/10.1016/j.trb.2019.02.009.

BATISTA, S. F. A.; INGOLE, D.; LECLERCQ, L. & MENÉNDEZ, M. [2021b]: The role of trip lengths calibration in model-based perimeter control strategies. In: *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11. DOI: 10.1109/tits.2021.3049679.

BOYACI, B. & GEROLIMINIS, N. [2011]: Estimation of the network capacity for multimodal urban systems. In: *Procedia - Social and Behavioral Sciences* 16, pp. 803–813. ISSN: 18770428. DOI: 10.1016/j.sbspro.2011.04.499.

BUISSON, C. & LADIER, C. [2009]: Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. In: *Transportation Research Record* 2124.1, pp. 127–136.

BUITELAAR, E.; VAN DER HEIJDEN, R. & ARGIOLU, R. [2007]: Managing traffic by privatization of road capacity: A property rights approach. In: *Transport Reviews* 27.6, pp. 699–713. DOI: 10.1080/01441640701262949.

CASTRILLON, F. [2015]: Theoretical analysis of the effects of bus operations on urban corridors and networks. PhD thesis. Georgia Institute of Technology, p. 92.

CASTRILLON, F. & LAVAL, J. [2017]: Impact of buses on the macroscopic fundamental diagram of homogeneous arterial corridors. In: *Transportmetrica B: Transport Dynamics* 6.4, pp. 286–301. ISSN: 21680582. DOI: 10.1080/21680566.2017.1314203.

CHEN, P.; WEI, L.; MENG, F. & ZHENG, N. [2020]: Vehicle trajectory reconstruction for signalized intersections: A hybrid approach integrating Kalman Filtering and variational theory. In: *Transportmetrica B: Transport Dynamics* 0.0, pp. 1–20. DOI: 10.1080/21680566.2020.1781707.

CHIABAUT, N. [2015]: Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram. In: *Transportation Research Part B: Methodological* 81, pp. 410–420. ISSN: 01912615. DOI: 10.1016/j.trb.2015.02.005.

CHOW, A. H.; LI, S.; SZETO, W. & WANG, D. Z. [2015]: Modelling urban traffic dynamics based upon the variational formulation of kinematic waves. In: *Transportmetrica B: Transport Dynamics* 3.3, pp. 169–191.

CLAUDEL, C. G. & BAYEN, A. M. [2010a]: Lax–Hopf based incorporation of internal boundary conditions into Hamilton–Jacobi equation. Part I: Theory. In: *IEEE Transactions on Automatic Control* 55.5, pp. 1142–1157.

CLAUDEL, C. G. & BAYEN, A. M. [2010b]: Lax–Hopf based incorporation of internal boundary conditions into Hamilton-Jacobi equation. Part II: Computational methods. In: *IEEE Transactions on Automatic Control* 55.5, pp. 1158–1174.

CLEVELAND, W. S. [1979]: Robust locally weighted regression and smoothing scatterplots. In: *Journal of the American Statistical Association* 74.368, pp. 829–836. DOI: 10.1080/01621459.1979.10481038.

COSTESEQUE, G. & LEBACQUE, J.-P. [2014]: A variational formulation for higher order macroscopic traffic flow models: Numerical investigation. In: *Transportation Research Part B: Methodological* 70, pp. 112–133. ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2014.08.012.

COURANT, R.; FRIEDRICHS, K. & LEWY, H. [1928]: Über die partiellen Differenzengleichungen der mathematischen Physik. In: *Mathematische annalen* 100.1, pp. 32–74.

COURBON, T. & LECLERCQ, L. [2011]: Cross-comparison of macroscopic fundamental diagram estimation methods. In: *Procedia-Social and Behavioral Sciences* 20, pp. 417–426. DOI: 10.1016/j.sbspro.2011.08.048.

DAGANZO, C. F. [1997]: *Fundamentals of transportation and traffic operations*. Pergamon Oxford.

DAGANZO, C. [1992]: *The cell transmission model. Part I: A simple dynamic representation of highway traffic*. Tech. rep. University of California, Berkeley.

DAGANZO, C. F. [1995]: The cell transmission model. Part II: Network traffic. In: *Transportation Research Part B: Methodological* 29.2, pp. 79–93. DOI: 10.1016/0191-2615(94)00022-r.

DAGANZO, C. F. & MENENDEZ, M. [2005]: A variational formulation of kinematic waves: Bottleneck properties and examples. In: *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*. Maryland: Elsevier, pp. 345–364. ISBN: 0080446809.

DAGANZO, C. F. [2005a]: A variational formulation of kinematic waves: Solution methods. In: *Transportation Research Part B: Methodological* 39.10, pp. 934–950. ISSN: 0191-2615. DOI: 10.1016/j.trb.2004.05.003.

DAGANZO, C. F. [2007]: Urban gridlock: Macroscopic modeling and mitigation approaches. In: *Transportation Research Part B: Methodological* 41.1, pp. 49–62. DOI: 10.1016/j.trb.2006.03.001.

DAGANZO, C. F. & KNOOP, V. L. [2016]: Traffic flow on pedestrianized streets. In: *Transportation Research Part B: Methodological* 86, pp. 211–222. ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2015.12.017.

DAGANZO, C. F. [2005b]: A variational formulation of kinematic waves: Basic theory and complex boundary conditions. In: *Transportation Research Part B: Methodological* 39.2, pp. 187–196. ISSN: 0191-2615. DOI: 10.1016/j.trb.2004.04.003.

DAGANZO, C. F. & GEROLIMINIS, N. [2008]: An analytical approximation for the macroscopic fundamental diagram of urban traffic. In: *Transportation Research Part B: Methodological* 42, pp. 771–781. ISSN: 01912615. DOI: 10.1016/j.trb.2008.06.008.

DAGANZO, C. F.; GAYAH, V. V. & GONZALES, E. J. [2011]: Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. In: *Transportation Research Part B: Methodological* 45.1, pp. 278–288. ISSN: 01912615. DOI: 10.1016/j.trb.2010.06.006.

DAGANZO, C. F. & LEHE, L. J. [2016]: Traffic flow on signalized streets. In: *Transportation Research Part B: Methodological* 90, pp. 56–69. ISSN: 01912615. DOI: 10.1016/j.trb.2016.03.010.

DAGANZO, C. F.; LEHE, L. J. & ARGOTE-CABANERO, J. [2018]: Adaptive offsets for signalized streets. In: *Transportation Research Part B: Methodological* 117, pp. 926–934. ISSN: 01912615. DOI: 10.1016/j.trb.2017.08.011.

DAKIC, I. & MENENDEZ, M. [2018]: On the use of Lagrangian observations from public transport and probe vehicles to estimate car space-mean speeds in bi-modal urban networks. In: *Transportation Research Part C: Emerging Technologies* 91, pp. 317–334. DOI: 10.1016/j.trc.2018.04.004.

DAKIC, I.; AMBÜHL, L.; SCHÜMPERLIN, O. & MENENDEZ, M. [2020]: On the modeling of passenger mobility for stochastic bi-modal urban corridors. In: *Transportation Research Part C: Emerging Technologies* 113, pp. 146–163. ISSN: 0968090x. DOI: 10.1016/j.trc.2019.05.018.

DANDL, F.; TILG, G.; ROSTAMI-SHAHRBABAKI, M. & BOGENBERGER, K. [2020]: Network fundamental diagram based routing of vehicle fleets in dynamic traffic simulations. In: *2020 23th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8. DOI: 10.1109/itsc45102.2020.9294204.

DANDL, F.; ENGELHARDT, R.; HYLAND, M.; TILG, G.; BOGENBERGER, K. & MAHMASSANI, H. S. [2021]: Regulating mobility-on-demand services: Tri-level model and Bayesian optimization solution approach. In: *Transportation Research Part C: Emerging Technologies* 125, p. 103075.

DEPARTMENT FOR TRANSPORT [Oct. 2015]: *Free flow vehicle speeds statistics: Great Britain 2014*. Tech. rep. Department for Transport, UK.

DRAKE, J.; SCHOFER, J. & MAY, A. [1966]: A statistical analysis of speed-density hypotheses. In: *Highway Research Record* 154, pp. 53–87.

DU, J.; RAKHA, H. & GAYAH, V. V. [2016]: Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. In: *Transportation Research Part C: Emerging Technologies* 66, pp. 136–149. DOI: 10.1016/j.trc.2015.08.015.

DURET, A. & YUAN, Y. [2017]: Traffic state estimation based on Eulerian and Lagrangian observations in a mesoscopic modeling framework. In: *Transportation Research Part B: Methodological* 101, pp. 51–71.

EDIE, L. [1963]: Discussion of traffic stream measurements and definitions. In: *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*. Ed. by J. ALMOND. Paris, France: Oecd, pp. 139–154.

EUROSTAT [2021]: *Passenger cars per 1'000 inhabitants*. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_eqs_carhab. Accessed: 2021-08-12.

EVANS, L. C. [2010]: *Partial differential equations*. Providence, R.I.: American Mathematical Society. ISBN: 978-0-8218-4974-3.

FRIESZ, T. L.; HAN, K.; NETO, P. A.; MEIMAND, A. & YAO, T. [2013]: Dynamic user equilibrium based on a hydrodynamic model. In: *Transportation Research Part B: Methodological* 47, pp. 102–126.

GAN, Q.-J.; JIN, W.-L. & GAYAH, V. V. [2017]: Analysis of traffic statics and dynamics in signalized networks: a poincaré map approach. In: *Transportation Science* 51.3, pp. 1009–1029. DOI: 10.1287/trsc.2017.0740.

GAYAH, V. V. & DAGANZO, C. F. [2011]: Effects of turning maneuvers and route choice on a simple network. In: *Transportation Research Record* 2249, pp. 15–19. ISSN: 03611981. DOI: 10.3141/2249-03.

GAYAH, V. V.; ILGIN GULER, S. & GU, W. [2016]: On the impact of obstructions on the capacity of nearby signalised intersections. In: *Transportmetrica B* 4.1, pp. 48–67. ISSN: 21680582. DOI: 10.1080/21680566.2015.1052111.

GE, Q.; CIUFFO, B. & MENENDEZ, M. [2014]: An exploratory study of two efficient approaches for the sensitivity analysis of computationally expensive traffic simulation models. In: *IEEE Transactions on Intelligent Transportation Systems* 15.3, pp. 1288–1297.

GE, Q.; CIUFFO, B. & MENENDEZ, M. [2015]: Combining screening and metamodel-based methods: An efficient sequential approach for the sensitivity analysis of model outputs. In: *Reliability Engineering & System Safety* 134, pp. 334–344.

GEROLIMINIS, N. & DAGANZO, C. F. [2007]: Macroscopic modeling of traffic in cities. In: *Presented at the 86th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

GEROLIMINIS, N. & DAGANZO, C. F. [2008]: Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. In: *Transportation Research Part B: Methodological* 42.9, pp. 759–770. ISSN: 01912615. DOI: 10.1016/j.trb.2008.02.002.

GEROLIMINIS, N. & SUN, J. [2011a]: Hysteresis phenomena of a Macroscopic Fundamental Diagram in freeway networks. In: *Transportation Research Part A: Policy and Practice* 45.9, pp. 966–979. ISSN: 09658564. DOI: 10.1016/j.tra.2011.04.004.

GEROLIMINIS, N. & SUN, J. [2011b]: Properties of a well-defined macroscopic fundamental diagram for urban traffic. In: *Transportation Research Part B: Methodological* 45.3, pp. 605–617. ISSN: 01912615. DOI: 10.1016/j.trb.2010.11.004.

GEROLIMINIS, N. & BOYACI, B. [2012]: The effect of variability of urban systems characteristics in the network capacity. In: *Transportation Research Part B: Methodological* 46.10, pp. 1607–1623. ISSN: 01912615. DOI: 10.1016/j.trb.2012.08.001.

GEROLIMINIS, N.; HADDAD, J. & RAMEZANI, M. [2013]: Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. In: *IEEE Transactions on Intelligent Transportation Systems* 14.1, pp. 348–359. DOI: 10.1109/tits.2012.2216877.

GEROLIMINIS, N.; ZHENG, N. & AMPOUNTOLAS, K. [2014]: A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. In: *Transportation Research Part C: Emerging Technologies* 42, pp. 168–181. DOI: 10.1016/j.trc.2014.03.004.

GIRAULT, J.-T.; GAYAH, V. V.; GULER, I. & MENENDEZ, M. [2016]: Exploratory analysis of signal coordination impacts on macroscopic fundamental diagram. In: *Transportation Research Record: Journal of the Transportation Research Board* 2560.2560 (1), pp. 36–46. DOI: 10.3141/2560-05.

GODFREY, J. [1969]: The mechanism of a road network. In: *Traffic Engineering & Control* 8.8, pp. 323–327.

GREENSHIELDS, B.; BIBBINS, J.; CHANNING, W. & MILLER, H. [1935]: A study of traffic capacity. In: *Highway Research Board Proceedings*. Vol. 1935. National Research Council (USA), Highway Research Board.

GRIGOROPOULOS, G.; KELER, A.; KATHS, J.; KATHS, H.; SPANGLER, M.; HOFFMANN, S. & BUSCH, F. [2018]: Evaluation of the traffic efficiency of bicycle highways: A microscopic traffic simulation study. In: *Presented at the 7th Symposium of the European Association for Research in Transportation*. Athens, Greece.

GULER, S. I.; GAYAH, V. V. & MENENDEZ, M. [2016]: Bus priority at signalized intersections with single-lane approaches: A novel pre-signal strategy. In: *Transportation Research Part C: Emerging Technologies* 63, pp. 51–70.

HADDAD, J. & ZHENG, Z. [2020]: Adaptive perimeter control for multi-region accumulation-based models with state delays. In: *Transportation Research Part B: Methodological* 137, pp. 133–153. ISSN: 0191-2615. DOI: 10.1016/j.trb.2018.05.019.

HAITAO, H.; YANG, K.; LIANG, H.; MENENDEZ, M. & GULER, S. I. [2019]: Providing public transport priority in the perimeter of urban networks: A bimodal strategy. In: *Transportation Research Part C: Emerging Technologies* 107, pp. 171–192. DOI: 10.1016/j.trc.2019.08.004.

HALL, F. L. [1996]: Traffic stream characteristics. In: *Traffic Flow Theory. US Federal Highway Administration* 36.

HAMM, L.; LODER, A.; TILG, G.; MENENDEZ, M. & BOGENBERGER, K. [2022a]: Network inefficiency - Empirical findings for six European cities. In: *Presented at the 101th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

HAMM, L. S.; LODER, A.; TILG, G.; MENENDEZ, M. & BOGENBERGER, K. [2022b]: Network Inefficiency: Empirical Findings for Six European Cities. In: *Transportation Research Record* 0.0, pp. 1–13. DOI: 10.1177/03611981221082588.

HAN, K.; PICCOLI, B. & SZETO, W. [2015]: Continuous-time link-based kinematic wave model: formulation, solution existence, and well-posedness. In: *Transportmetrica B: Transport Dynamics* 4.3, pp. 187–222.

HANKE, H.; BRAAM, W.; BREITENSTEIN, J.; HABERMEHL, K.; HERBER, F.-R.; STEPHAN, R. & WETTERLING, K. [2012]: *Begriffsbestimmungen. Teil: Verkehrsplanung, Straßenentwurf und Straßenbetrieb*. Tech. rep. Forschungsgesellschaft für Straßen- und Verkehrswesen.

HANS, E.; CHIABAUT, N. & LECLERCQ, L. [2015]: Applying variational theory to travel time estimation on urban arterials. In: *Transportation Research Part B: Methodological* 78, pp. 169–181. ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2015.04.004.

HERMAN, R. & PRIGOGINE, I. [1979]: A two-fluid approach to town traffic. In: *Science* 204.4389 (4389), pp. 148–151. DOI: 10.1126/science.204.4389.148.

HERMAN, R. & ARDEKANI, S. [1984]: Characterizing traffic conditions in urban areas. In: *Transportation Science* 18.2, pp. 101–140.

HERZ, R.; SCHLICHTER, G. & SIEGENER, W. [1976]: Angewandte Statistik für Verkehrs- und Regionalplaner. In: *Verkehrs-und Regionalplaner, Werner-Ingenieur-Texte 42, Werner-Verlag, Düsseldorf*, p. 131.

INGOLE, D.; MARIOTTE, G. & LECLERCQ, L. [2020]: Perimeter gating control and citywide dynamic user equilibrium: A macroscopic modeling framework. In: *Transportation Research Part C: Emerging Technologies* 111, pp. 22–49.

INRIX [2020]: *Global traffic Scorecard*. https://inrix.com/scorecard/. Accessed: 2021-08-02.

JI, Y.; DAAMEN, W.; HOOGENDOORN, S.; HOOGENDOORN-LANSER, S. & QIAN, X. [2010]: Investigating the shape of the macroscopic fundamental diagram using simulation data. In: *Transportation Research Record: Journal of the Transportation Research Board* 2161.1, pp. 40–48. ISSN: 0361-1981. DOI: 10.3141/2161-05.

JIN, W.-L.; GAN, Q.-J. & GAYAH, V. V. [2013]: A kinematic wave approach to traffic statics and dynamics in a double-ring network. In: *Transportation Research Part B: Methodological* 57, pp. 114–131. DOI: 10.1016/j.trb.2013.09.004.

JIN, W.-L. [2015]: Continuous formulations and analytical properties of the link transmission model. In: *Transportation Research Part B: Methodological* 74, pp. 88–103.

JOHARI, M.; KEYVAN-EKBATANI, M.; LECLERCQ, L.; NGODUY, D. & MAHMASSANI, H. S. [2021]: Macroscopic network-level traffic models: Bridging fifty years of development toward the next era. In: *Transportation Research Part C: Emerging Technologies* 131, p. 103334.

KAWAI, K.; TAKENOUCHI, A.; IKAWA, M. & KUWAHARA, M. [2019]: Traffic state estimation using traffic measurement from the opposing lane–error analysis based on fluctuation of input data. In: *Intelligent Transport Systems for Everyone's Mobility*. Springer, pp. 247–263.

KAWASAKI, Y.; HARA, Y. & KUWAHARA, M. [2017]: Real-time monitoring of dynamic traffic states by state-space model. In: *Transportation Research Procedia* 21, pp. 42–55. ISSN: 23521465. DOI: 10.1016/j.trpro.2017.03.076.

KERNER, B. S. & LIEU, H. [2005]: The physics of traffic: empirical freeway pattern features, engineering applications; and theory. In: *Physics Today* 58.11, pp. 54–56.

KEYVAN-EKBATANI, M.; KOUVELAS, A.; PAPAMICHAIL, I. & PAPAGEORGIOU, M. [2012]: Exploiting the fundamental diagram of urban networks for feedback-based gating. In: *Transportation Research Part B: Methodological* 46.10, pp. 1393–1403. DOI: 10.1016/j.trb.2012.06.008.

KNOOP, V. L.; JONG, D. DE & HOOGENDOORN, S. P. [2014]: The influence of the road layout on the network fundamental diagram. In: *Transportation Research Record: Journal of the Transportation Research Board* 2421, pp. 22–30. DOI: 10.3141/2421-03.

KNOOP, V. L.; LINT, H. VAN & HOOGENDOORN, S. P. [2015]: Traffic dynamics: Its impact on the macroscopic fundamental diagram. In: *Physica A: Statistical Mechanics and its Applications* 438, pp. 236–250. DOI: 10.1016/j.physa.2015.06.016.

KNOOP, V. L. & DAAMEN, W. [2017]: Automatic fitting procedure for the fundamental diagram. In: *Transportmetrica B: Transport Dynamics* 5.2, pp. 129–144.

KNOOP, V. L. & DAGANZO, C. F. [2018]: The effect of crosswalks on traffic flow. In: *European Journal of Transport and Infrastructure Research* 18.2, pp. 145–157. ISSN: 15677141. DOI: 10.18757/ejtir.2018.18.2.3227.

KOSHI, M. [1983]: Some findings and an overview on vehicular flow characteristics. In: *Proceedings of the 8th International Symposium on Transportation and Traffic Theory*.

KOUVELAS, A.; SAEEDMANESH, M. & GEROLIMINIS, N. [2017]: Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. In: *Transportation Research Part B: Methodological* 96, pp. 26–45. ISSN: 0191-2615. DOI: doi.org/10.1016/j.trb.2016.10.011.

KRAUSS, S. [1998]: Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics. PhD thesis.

LAVAL, J. A. & DAGANZO, C. F. [2006]: Lane-changing in traffic streams. In: *Transportation Research Part B: Methodological* 40, pp. 251–264.

LAVAL, J. A. & LECLERCQ, L. [2013]: The Hamilton–Jacobi partial differential equation and the three representations of traffic flow. In: *Transportation Research Part B: Methodological* 52, pp. 17–30. ISSN: 01912615. DOI: 10.1016/j.trb.2013.02.008.

LAVAL, J. A. & CASTRILLÓN, F. [2015]: Stochastic approximations for the macroscopic fundamental diagram of urban networks. In: *Transportation Research Part B: Methodological* 81, pp. 904–916. ISSN: 01912615. DOI: 10.1016/j.trb.2015.09.002.

LAVAL, J. A.; COSTESEQUE, G. & CHILUKURI, B. [2016]: The impact of source terms in the variational representation of traffic flow. In: *Transportation Research Part B: Methodological* 94, pp. 204–216. ISSN: 01912615. DOI: 10.1016/j.trb.2016.09.011.

LEBACQUE, J. P. & KHOSHYARAN, M. M. [2013]: A variational formulation for higher order macroscopic traffic flow models of the GSOM family. In: *Transportation Research Part B: Methodological* 57, pp. 245–265. ISSN: 01912615. DOI: 10.1016/j.trb.2013.07.005.

LEBLANC, L. J. [1975]: An algorithm for the discrete network design problem. In: *Transportation Science* 9.3, pp. 183–199.

LECLERCQ, L.; LAVAL, J. & CHEVALLIER, E. [2007]: The Lagrangian coordinate system and what it means for first order traffic flow model. In: *Proceedings of the 17th International Symposium on Transportation and Traffic Theory*. New York, USA.

LECLERCQ, L. [2005]: Calibration of flow–density relationships on urban streets. In: *Transportation Research Record: Journal of the Transportation Research Board* 1934.1, pp. 226–234. DOI: 10.1177/0361198105193400124.

LECLERCQ, L. & BECARIE, C. [2012]: Meso lighthill-whitham and richards model designed for network applications. In: *Presented at the 91th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

LECLERCQ, L. & GEROLIMINIS, N. [2013]: Estimating MFDs in simple networks with route choice. In: *Transportation Research Part B: Methodological* 57, pp. 468–484. ISSN: 01912615. DOI: 10.1016/j.trb.2013.05.005.

LECLERCQ, L.; CHIABAUT, N. & TRINQUIER, B. [2014]: Macroscopic fundamental diagrams: A cross-comparison of estimation methods. In: *Transportation Research Part B: Methodological* 62, pp. 1–12. DOI: 10.1016/j.trb.2014.01.007.

LECLERCQ, L.; PARZANI, C.; KNOOP, V. L.; AMOURETTE, J. & HOOGENDOORN, S. P. [2015]: Macroscopic traffic dynamics with heterogeneous route patterns. In: *Transportation Research Part C: Emerging Technologies* 59, pp. 292–307. DOI: 10.1016/j.trc.2015.05.006.

LECLERCQ, L. & PAIPURI, M. [2019]: Macroscopic traffic dynamics under fast-varying demand. In: *Transportation Science* 53.6, pp. 1526–1545. ISSN: 0041-1655. DOI: 10.1287/trsc.2019.0908.

LEVY, J. I.; BUONOCORE, J. J. & VON STACKELBERG, K. [2010]: Evaluation of the public health impacts of traffic congestion: a health risk assessment. In: *Environmental health* 9.1, pp. 1–12.

LI, J. & ZHANG, H. M. [2015]: Bounding tandem queuing system performance with variational theory. In: *Transportation Research Part B: Methodological* 81, pp. 848–862. ISSN: 01912615. DOI: 10.1016/j.trb.2015.07.013.

LIGHTHILL, M. J. & WHITHAM, G. B. [1955]: On kinematic waves II. A theory of traffic flow on long crowded roads. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 229.1178, pp. 317–345.

LITMAN, T. [2011]: Victoria Transport Policy Institute. In: *The Future Isn't What It Used To Be: Changing Trends and their Implications For Transport Planning* 6.

LODER, A.; AMBÜHL, L.; MENENDEZ, M. & AXHAUSEN, K. W. [2017]: Empirics of multi-modal traffic networks – Using the 3D macroscopic fundamental diagram. In: *Transportation Research Part C: Emerging Technologies* 82, pp. 88–101. DOI: 10.1016/j.trc.2017.06.009.

LODER, A.; DAKIC, I.; BRESSAN, L.; AMBÜHL, L.; BLIEMER, M. C.; MENENDEZ, M. & AXHAUSEN, K. W. [2019a]: Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram. In: *Transportation Research Part B: Methodological* 129, pp. 1–19.

LODER, A.; AMBÜHL, L.; MENENDEZ, M. & AXHAUSEN, K. W. [2019b]: Understanding traffic capacity of urban networks. In: *Scientific Reports* 9.1, pp. 1–10. ISSN: 2045-2322. DOI: 10.1038/s41598-019-51539-5.

LONG, J.; GAO, Z.; ZHAO, X.; LIAN, A. & ORENSTEIN, P. [2011]: Urban traffic jam simulation based on the cell transmission model. In: *Networks and Spatial Economics* 11.1, pp. 43–64.

LOPEZ, P. A.; BEHRISCH, M.; BIEKER-WALZ, L.; ERDMANN, J.; FLÖTTERÖD, Y.-P.; HILBRICH, R.; LÜCKEN, L.; RUMMEL, J.; WAGNER, P. & WIESSNER, E. [2018]: Microscopic traffic simulation using SUMO. In: *2018 21th International Conference on Intelligent Transportation Systems (ITSC)*. Ieee.

MA, T.; ANTONIOU, C. & TOLEDO, T. [2020]: Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. In: *Transportation Research Part C: Emerging Technologies* 111, pp. 352–372.

MAHMASSANI, H.; WILLIAMS, J. C. & HERMAN, R. [1987]: Performance of urban traffic networks. In: *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, pp. 1–20.

MAHMASSANI, H. S.; SABERI, M. & ZOCKAIE, A. [2013]: Urban network gridlock: Theory, characteristics, and dynamics. In: *Transportation Research Part C: Emerging Technologies* 36, pp. 480–497. DOI: 10.1016/j.trc.2013.07.002.

MAHUT, M.; FLORIAN, M. & TREMBLAY, N. [2003]: Space-time queues and dynamic traffic assignment: A model, algorithm and applications. In: *Presented at the 82nd Annual Meeting of the Trabsportation Research Board*. Washington, D.C., USA.

MARIOTTE, G.; LECLERCQ, L. & LAVAL, J. A. [2017]: Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. In: *Transportation Research Part B: Methodological* 101.0, pp. 245–267. ISSN: 01912615. DOI: 10.1016/j.trb.2017.04.002.

MARIOTTE, G. & LECLERCQ, L. [2019]: Flow exchanges in multi-reservoir systems with spillbacks. In: *Transportation Research Part B: Methodological* 122, pp. 327–349. ISSN: 0191-2615. DOI: doi.org/10.1016/j.trb.2019.02.014.

MAZARÉ, P. E.; DEHWAH, A. H.; CLAUDEL, C. G. & BAYEN, A. M. [2011]: Analytical and grid-free solutions to the Lighthill-Whitham-Richards traffic flow model. In: *Transportation Research Part B: Methodological* 45.10, pp. 1727–1748. ISSN: 01912615. DOI: 10.1016/j.trb.2011.07.004.

MAZLOUMIAN, A.; GEROLIMINIS, N. & HELBING, D. [2010]: The spatial variability of vehicle densities as determinant of urban network capacity. In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 368.1928, pp. 4627–4647. DOI: 10.1098/rsta.2010.0099.

MEHRAN, B.; KUWAHARA, M. & NAZNIN, F. [2012]: Implementing kinematic wave theory to reconstruct vehicle trajectories from fixed and probe sensor data. In: *Transportation Research Part C: Emerging Technologies* 20.1, pp. 144–163.

MEHRAN, B. & KUWAHARA, M. [2013]: Fusion of probe and fixed sensor data for short-term traffic prediction in urban signalized arterials. In: *International Journal of Urban Sciences* 17.2, pp. 163–183. ISSN: 12265934. DOI: 10.1080/12265934.2013.776291.

MOSKOWITZ, K. [1965]: Discussion of 'freeway level of service as influenced by volume and capacity characteristics' by D.R. Drewand C. J. Keese. In: *Highway Research Record* 99, pp. 43–44.

MÜHLICH, N.; GAYAH, V. V. & MENENDEZ, M. [2014]: An examination of MFD hysteresis patterns for hierarchical urban street networks using micro-simulation. In: *Presented at the 94th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

NAGLE, A. & GAYAH, V. [2014]: Accuracy of networkwide traffic states estimated from mobile probe data. In: *Transportation Research Record: Journal of the Transportation Research Board* 2421, pp. 1–11. ISSN: 0361-1981. DOI: 10.3141/2421-01.

NEWELL, G. F. [1993]: A simplified theory of kinematic waves in highway traffic, part I: General theory, part II: Queueing at freeway bottlenecks, part III: Multi-destination flows. In: *Transportation Research Part B: Methodological* 27.4, pp. 281–313.

NI, W. & CASSIDY, M. [2020]: City-wide traffic control: modeling impacts of cordon queues. In: *Transportation Research Part C: Emerging Technologies*, pp. 164–175. DOI: 10.1016/j.trc.2019.04.024.

OPENSTREETMAP CONTRIBUTORS [2019]: *Munich dump retrieved from https://planet.osm.org*.

ORTIGOSA, J.; MENENDEZ, M. & GAYAH, V. V. [2015]: Analysis of network exit functions for various urban grid network configurations. In: *Transportation Research Record* 2491.1, pp. 12–21.

OSORIO, C. & BIERLAIRE, M. [2009]: An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. In: *European Journal of Operational Research* 196.3, pp. 996–1007.

PAYNE, H. [1971]: Models of freeway traffic and control. In: *Mathematical Models of Public Systems* 1, pp. 51–60.

POLSON, N. & SOKOLOV, V. [2015]: Bayesian analysis of traffic flow on interstate I-55: The LWR model. In: *Annals of Applied Statistics* 9.4, pp. 1864–1888. ISSN: 19417330. DOI: 10.1214/15-aoas853. eprint: 1409.6034.

RAADSEN, M. P.; BLIEMER, M. C. & BELL, M. G. [2016]: An efficient and exact event-based algorithm for solving simplified first order dynamic network loading problems in continuous time. In: *Transportation Research Part B: Methodological* 92, pp. 191–210.

RAKHA, H.; HELLINGA, B.; VAN AERDE, M. & PEREZ, W. [1996]: Systematic verification, validation and calibration of traffic simulation models. In: *Presented at the 75th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

REMPE, F.; LODER, A. & BOGENBERGER, K. [2021]: Estimating motorway traffic states with data fusion and physics-informed deep learning. In: *2021 24th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2208–2214. DOI: 10.1109/ITSC48978.2021.9565096.

RICHARDS, P. I. [1956]: Shock waves on the highway. In: *Operations Research* 4.1, pp. 42–51.

ROCA-RIU, M.; MENENDEZ, M.; DAKIC, I.; BUEHLER, S. & ORTIGOSA, J. [2020]: Urban space consumption of cars and buses: An analytical approach. In: *Transportmetrica B: Transport Dynamics* 8.1, pp. 237–263.

SABERI, M. & MAHMASSANI, H. [2013]: Hysteresis and capacity drop phenomena in freeway networks empirical characterization and interpretation. In: *Transportation Research Record: Journal of the Transportation Research Board* 2391, pp. 44–55. DOI: 10.3141/2391-05.

SABERI, M.; MAHMASSANI, H.; HOU, T. & ZOCKAIE, A. [2014]: Estimating network fundamental diagram using three-dimensional vehicle trajectories. In: *Transportation Research Record: Journal of the Transportation Research Board* 2422, pp. 12–20. ISSN: 0361-1981. DOI: 10.3141/2422-02.

SAEEDMANESH, M. & GEROLIMINIS, N. [2016]: Clustering of heterogeneous networks with directional flows based on "Snake" similarities. In: *Transportation Research Part B: Methodological* 91, pp. 250–269. DOI: 10.1016/j.trb.2016.05.008.

SAEEDNIA, M. & MENENDEZ, M. [2016]: A decision support system for real-time platooning of trucks. In: *2016 19th International Conference on Intelligent Transportation Systems (ITSC)*. Ieee, pp. 1792–1797.

SAFFARI, E.; YILDIRIMOGLU, M. & HICKMAN, M. [2020]: A methodology for identifying critical links and estimating macroscopic fundamental diagram in large-scale urban networks. In: *Transportation Research Part C: Emerging Technologies* 119, p. 102743.

SEO, T.; BAYEN, A. M.; KUSAKABE, T. & ASAKURA, Y. [2017]: Traffic state estimation on highway: A comprehensive survey. In: *Annual Reviews in Control* 43, pp. 128–151.

SHI, R.; MO, Z. & DI, X. [2021]: Physics–informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 1, pp. 540–547.

SHIM, J.; YEO, J.; LEE, S.; HAMDAR, S. H. & JANG, K. [2019]: Empirical evaluation of influential factors on bifurcation in macroscopic fundamental diagrams. In: *Transportation Research Part C: Emerging Technologies* 102, pp. 509–520.

SIRMATEL, I. I. & GEROLIMINIS, N. [2019]: Nonlinear Moving Horizon Estimation for Large-Scale Urban Road Networks. In: *IEEE Transactions on Intelligent Transportation Systems* 21.12, pp. 4983–4994. DOI: 10.1109/tits.2019.2946324.

SIRMATEL, I. I.; TSITSOKAS, D.; KOUVELAS, A. & GEROLIMINIS, N. [2021]: Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics. In: *Transportation Research Part C: Emerging Technologies* 128, pp. 1–20. ISSN: 0968-090x. DOI: 10.1016/j.trc.2021.103157.

SMEED, R. J. [1967]: The road capacity of city centers. In: *Highway Research Record* 169.1, pp. 22–29.

SMEED, R. J. [1968]: Traffic studies and urban congestion. In: *Journal of Transport Economics and Policy* 2.1, pp. 33–70.

SUN, Z. & BAN, X. J. [2013]: Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors. In: *Transportation Research Part C: Emerging Technologies* 36, pp. 268–283.

SYSTEMATICS, C. [2005]: *Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation*. Tech. rep. United States. Federal Highway Administration.

TAKENOUCHI, A.; KAWAI, K. & KUWAHARA, M. [2019]: Traffic state estimation and its sensitivity utilizing measurements from the opposite lane. In: *Transportation Research Part C: Emerging Technologies* 104.April, pp. 95–109. ISSN: 0968090x. DOI: 10.1016/j.trc.2019.04.016.

TAMPÈRE, C. M.; CORTHOUT, R.; CATTRYSSE, D. & IMMERS, L. H. [2011]: A generic class of first order node models for dynamic macroscopic simulation of traffic flows. In:

Transportation Research Part B: Methodological 45.1, pp. 289–309. ISSN: 01912615. DOI: 10.1016/j.trb.2010.06.004.

THOMSON, J. [1967]: Speeds and flow of traffic in Central London. 2: Speed-flow relations. In: *Traffic Engineering & Control* 8.12, pp. 721–725.

TILG, G.; YANG, K. & MENENDEZ, M. [2018]: Evaluating the effects of automated vehicle technology on the capacity of freeway weaving sections. In: *Transportation Research Part C: Emerging Technologies* 96, pp. 3–21. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2018.09.014.

TILG, G.; AMINI, S. & BUSCH, F. [2019]: Arterial macroscopic fundamental diagram: A comparison of analytical approximations and empirical data from Munich. In: *Presented at the 98th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

TILG, G.; AMINI, S. & BUSCH, F. [2020a]: Evaluation of analytical approximation methods for the macroscopic fundamental diagram. In: *Transportation Research Part C: Emerging Technologies* 114, pp. 1–19. ISSN: 0968-090x. DOI: doi.org/10.1016/j.trc.2020.02.003.

TILG, G.; UL ABEDIN, Z.; AMINI, S. & BUSCH, F. [2020b]: Simulation-based design of urban bi-modal transport systems. In: *Frontiers in Future Transportation* 1. ISSN: 2673-5210. DOI: 10.3389/ffutr.2020.581622.

TILG, G.; SKRECKI, M.; BAHR, F.; KELER, A.; TSAKARESTOS, A.; GÖBEL, R. & BUSCH, F. [2021a]: Assessing the spatial impacts of unreliable public transport systems: A quasi real-time data-driven approach. In: *Presented at the 8th International Symposium on Transport Network Reliability*.

TILG, G.; AMBÜHL, L.; BATISTA, S. F.; MENENDEZ, M. & BUSCH, F. [2021b]: On the application of variational theory to urban networks. In: *Transportation Research Part B: Methodological* 150, pp. 435–456. DOI: 10.1016/j.trb.2021.06.019.

TILG, G.; AMBÜHL, L.; BATISTA, S.; MENENDEZ, M.; LECLERCQ, L. & BUSCH, F. [2021c]: Semi-analytical estimation of macroscopic fundamental diagrams: From corridors to networks. In: *Presented at the 100th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA.

TILG, G.; PAWLOWSKI, A. & BOGENBERGER, K. [2021d]: The impact of data characteristics on the estimation of the three-dimensional passenger macroscopic fundamental diagram. In: *2021 24th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2111–2117. DOI: 10.1109/ITSC48978.2021.9564430.

TILG, G.; AMBÜHL, L.; BATISTA, S. F.; MENENDEZ, M.; LECLERCQ, L. & BUSCH, F. [n.d.]: From corridor to network macroscopic fundamental diagrams: A semi-analytical estimation approach. Submitted for publication.

TRB [2016]: *Highway Capacity Manual Sixth Edition: A Guide for Multimodal Mobility Analysis*. 6th ed. Transportation Research Board of the National Academies of Science.

TREIBER, M. & KESTING, A. [2013]: *Traffic flow dynamics*, pp. 239–255. ISBN: 978-3-642-32459-8. DOI: 10.1007/978-3-642-32460-4.

TSUBOTA, T.; BHASKAR, A. & CHUNG, E. [2014]: Macroscopic fundamental diagram for Brisbane, Australia. In: *Transportation Research Record: Journal of the Transportation Research Board* 2421, pp. 12–21. ISSN: 0361-1981. DOI: 10.3141/2421-02.

VICKREY, W. [2020]: Congestion in midtown Manhattan in relation to marginal cost pricing. In: *Economics of Transportation* 21, p. 100152. ISSN: 2212-0122. DOI: 10.1016/j.ecotra.2019.100152.

VLAHOGIANNI, E. I.; KARLAFTIS, M. G. & GOLIAS, J. C. [2014]: Short-term traffic forecasting: Where we are and where we're going. In: *Transportation Research Part C: Emerging Technologies* 43, pp. 3–19.

WADA, K.; USUI, K.; TAKIGAWA, T. & KUWAHARA, M. [2018]: An optimization modeling of coordinated traffic signal control based on the variational theory and its stochastic extension. In: *Transportation Research Part B: Methodological* 117. Trb:isttt-22, pp. 907–925. ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2017.08.031.

WARDROP, J. G. [1952]: Some theoretical aspects of road traffic research. In: *Institution of Civil Engineering* 1, pp. 325–362. DOI: 10.1680/ipeds.1952.11259.

WARDROP, J. [1968]: Journey speed and flow in central urban areas. In: *Traffic Engineering & Control* 8.8.

WHITHAM, G. B. [2011]: *Linear and nonlinear waves*. Vol. 42. John Wiley & Sons.

WU, K.; GULER, S. I. & GAYAH, V. V. [2017]: Estimating the impacts of bus stops and transit signal priority on intersection operations: Queuing and variational theory approach. In: *Transportation Research Record* 2622.1, pp. 70–83. ISSN: 21694052. DOI: 10.3141/2622-07.

WU, K. & GULER, S. I. [2018]: Optimizing transit signal priority implementation along an arterial. In: *Transportation Research Record* 2672.20, pp. 215–227. DOI: 10.1177/0361198118790324.

WU, N. [2002]: A new approach for modeling of fundamental diagrams. In: *Transportation Research Part A: Policy and Practice* 36.10, pp. 867–884.

WU, X.; LIU, H. X. & GEROLIMINIS, N. [2011]: An empirical analysis on the arterial fundamental diagram. In: *Transportation Research Part B: Methodological* 45.1, pp. 255–266. ISSN: 0191-2615. DOI: 10.1016/j.trb.2010.06.003.

XIE, X.; CHIABAUT, N. & LECLERCQ, L. [2013]: Macroscopic fundamental diagram for urban streets and mixed traffic: Cross comparison of estimation methods. In: *Transportation Research Record: Journal of the Transportation Research Board* 2390 (1), pp. 1–10. DOI: 10.3141/2390-01.

XU, G.; YU, Z. & GAYAH, V. V. [2020]: Analytical method to approximate the impact of turning on the macroscopic fundamental diagram. In: *Transportation Research Record* 2674.9, p. 0361198120933274. DOI: 10.1177/0361198120933274.

YANG, K.; ZHENG, N. & MENENDEZ, M. [2018]: Multi-scale perimeter control approach in a connected-vehicle environment. In: *Transportation Research Part C: Emerging Technologies* 94, pp. 32–49. ISSN: 0968-090x. DOI: doi.org/10.1016/j.trc.2017.08.014.

YANG, K.; MENENDEZ, M. & ZHENG, N. [2019]: Heterogeneity aware urban traffic control in a connected vehicle environment: A joint framework for congestion pricing and perimeter control. In: *Transportation Research Part C: Emerging Technologies* 105, pp. 439–455. ISSN: 0968-090x. DOI: doi.org/10.1016/j.trc.2019.06.007.

YILDIRIMOGLU, M.; SIRMATEL, I. I. & GEROLIMINIS, N. [2018]: Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route

guidance. In: *Transportation Research Part B: Methodological* 118, pp. 106–123. ISSN: 0191-2615. DOI: doi.org/10.1016/j.trb.2018.10.007.

YPERMAN, I. [2007]: The link transmission model for dynamic network loading. PhD thesis. Katholieke Universiteit Leuven.

ZAHAVI, Y. [1972a]: Traffic performance evaluation of road networks by the α-relationship: part 1. In: *Traffic Engineering & Control* 14.5, pp. 228–231.

ZAHAVI, Y. [1972b]: Traffic performance evaluation of road networks by the α-relationship: part 2. In: *Traffic Engineering & Control* 14.6, pp. 292–293.

ZHENG, N.; RÉRAT, G. & GEROLIMINIS, N. [2016]: Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. In: *Transportation Research Part C: Emerging Technologies* 62, pp. 133–148. DOI: 10.1016/j.trc.2015.10.015.

ZHENG, N.; DANTSUJI, T.; WANG, P. & GEROLIMINIS, N. [2017]: Macroscopic approach for optimizing road space allocation of bus lanes in multimodal urban networks through simulation analysis. In: *Transportation Research Record* 2651.1, pp. 42–51. DOI: 10.3141/2651-05.

ZHONG, R.; CHEN, C.; HUANG, Y.; SUMALEE, A.; LAM, W. & XU, D. [2018]: Robust perimeter control for two urban regions with macroscopic fundamental diagrams: A control-Lyapunov function approach. In: *Transportation Research Part B: Methodological* 117, pp. 687–707. ISSN: 0191-2615. DOI: doi.org/10.1016/j.trb.2017.09.008.

ZOCKAIE, A.; SABERI, M. & SAEDI, R. [2018]: A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. In: *Transportation Research Part C: Emerging Technologies* 86.October 2017, pp. 245–262. ISSN: 0968090x. DOI: 10.1016/j.trc.2017.11.017.

List of Terms and Abbreviations

CTM cell transmission model 19, 57, 74, 82, 85, 97, 109–113, 115–118, 121, 125, 126, 134, 136, 164

ELDD empirical loop detector data 46

FCD floating car data 10, 33, 34, 38, 123, 129

FD fundamental diagram 4, 7, 8, 17, 19–25, 27, 38, 39, 52, 56, 59, 60, 63, 70, 71, 75, 78, 81, 92, 93, 98, 100, 107, 108, 116, 118, 122, 123, 129, 135, 137

FIFO first in first out 13, 59, 69, 73, 88, 93, 96, 103, 105, 122, 123

FS full spillback 87, 88, 91, 95, 98, 108, 111, 113, 115–117, 123

KWT kinematic wave theory 4, 11–15, 17–20, 22–24, 45, 52, 55–61, 63, 70, 71, 74, 76, 81, 91, 99, 106, 111, 113, 121, 122, 125–129, 133–135, 137, 138

LDD loop detector data 7, 32–35, 38, 40, 45, 46, 52, 129

LOESS locally estimated scatter plot smoothing 40, 46

LS limited spillback 87, 88, 91, 95, 108, 111, 113, 115, 116, 123

LTM link transmission model 19, 74, 82, 121, 126, 134, 136

MC method of cuts 12–14, 16, 23, 26–29, 31, 32, 34–36, 38–42, 44, 45, 47–53, 57, 82, 83, 85–93, 99–101, 110, 118, 123–125, 127, 132–138, 163, 167

MFD macroscopic fundamental diagram 4–16, 23, 25–29, 31–53, 55, 57, 58, 82–89, 91, 92, 96, 99–102, 107–115, 117–119, 123–129, 131–139, 163

nMC network method of cuts 14, 16, 86, 87, 91, 99, 106, 108, 109, 113, 115, 117, 118, 123–125, 127–129, 131, 134–138, 164

nVT network variational theory 14–16, 59, 61, 70–82, 87, 88, 91, 99, 100, 102, 106, 108, 110–113, 115, 116, 121–123, 125–129, 133–138, 164

OD origin-destination 7, 27, 33, 40, 84, 101, 113, 128, 138

PDE partial differential equation 17–20, 121

QP queue propagation 87, 106, 111, 113, 115–117, 123

RMSE root mean square error 36, 41, 45–49, 51, 53

SA stochastic approximation 11, 13, 14, 16, 26–29, 31, 32, 34–36, 39–41, 44, 45, 49–53, 83, 110, 118, 123, 125, 132, 133, 135–137, 163

VFCD virtual floating car data 46

VLDD virtual loop detector data 46

VT variational theory 11, 12, 14–17, 19–27, 29, 55–64, 66, 67, 70, 73–75, 81, 82, 85, 88, 97, 99, 121, 122, 125–128, 133–138, 163

List of Symbols

$\alpha \in A$	Turning ratio in set of turning ratios
$\tilde{\alpha}$	Factor impacting the queue growth due to turning flows
β	Time-dependent capacity constraint representing stationary bottlenecks
δ	Average coefficient of variation
δN	Cycle-based difference in supply- and demand-related cumulative count at a signal phase change
ΔN	Cycle-based difference in supply- and demand-related cumulative count
$\Delta_{P'}$	Level of congestion
Δr	Remaining red time
Δg	Remaining green time
Δt	Numerical step length in the temporal dimension
Δx	Numerical step length in the spatial dimension
γ	Factor for approximating congestion propagation
κ	Traffic density
κ_{max}	Link jam density
κ_{opt}	Optimal traffic density
κ'	Transformed traffic density
K	Network-wide average traffic density
K_k	Maximum traffic density on the link k
K_{max}	Network-wide jam density
λ	Inflow factor at origin links
λ_B	Mean dimensionless block length
μ_g	Mean green time
μ_l	Mean block lengths
μ_r	Mean red time
$\Omega = \{s_1, s_2\}$	Strategies related to $v_m = u, w$

ρ	Long-run red to green ratio
σ	Duration of spillback impact
τ	Time between beginning of active phase and t_e
θ	Absolute value of the ratio of u and w
B	Number of intersections
$\mathcal{B}$	Curve in time and space along which the boundary data $N_{\mathcal{B}}$ is given
c	Cycle length
c_f	Cycle when the final vehicle reaches the intersection
$C \in \mathcal{C}$	Corridor in set of corridors
$F_{q(\kappa')}(q)$	Cumulative density function of the MFD
g	Duration of a green phase
G	Lopsided variational graph
$\mathcal{G}$	Hypernetwork
H	Fundamental diagram
$I \in \mathcal{I}$	Intersection in set of intersections
i, j	Indices for corridors
k	Index for intersections and links
l	Link length
$L \in \mathcal{L}$	Link in set of links
m	Index for moving observer
n	Number of sample variations for Monte Carlo Simulation
N	Cumulative count/Moskowitz function
N_c	Maximum number of vehicles passing an intersection during a cycle
N_d	Cumulative count at the position right downstream of an intersection
N_e	Cumulative count at the end of a green phase
N_s	Cumulative count at the start of a green phase
N_u	Cumulative count at the position right upstream of an intersection
$\mathcal{N}$	Physical road network

o	Offset
$P(x,t)$	Generic point in time and space
$p \in \mathcal{P}$	Valid path in set of valid paths from $\mathcal{B}$ to P
p_s	Probability that a bus stops within a link
q	Traffic flow
q_f	Traffic flow during the cycle c_f
q_{max}	Maximum of n results for $Q(k)$
q_{mean}	Mean of n results for $Q(k)$
q_{min}	Minimum of n results for $Q(k)$
q_{opt}	Link capacity
$\tilde{q}$	Maximum cycle-average traffic flow during green phase
$\tilde{q}_d$	Traffic flow $\tilde{q}$ only constrained by demand
$\tilde{q}_s$	Traffic flow $\tilde{q}$ only constrained by supply
$\tilde{q}_{k,s}^{j \rightarrow i}$	Traffic flow $\tilde{q}$ on C^i constrained by spillbacks from C^j
Q	Network-wide average traffic flow
Q_{max}	Network-wide capacity
r	Duration of a red phase
$R(v_m)$	Average maximum passing traffic flow for a moving observer m
s_0	Strategy related to $v_m = 0$
t	Temporal dimension
t_b	Beginning of spillback impact
t_d	Average dwell times of buses
t_e	End of spillback impact
t_f	Time instant when the final vehicle leaves the intersection
$t_{k,dq}$	Time instant when the queue starts to grow at the downstream end of the link k
$t_{k,uq}$	Time instant when the queue reaches the upstream end of the link k
u	Free-flow speed
u_b	Average bus free-flow speed
w	Backward wave speed
v	Speed
$\overline{v}$	Instantaneous average speed

v_b	Average speed of buses
v_m	Speed of moving observer m
x	Spatial dimension
x_Φ	Position of an inter-corridor connection along a corridor
Z_m	Sum of all costs along the path of a moving observer m
z_p	Cost of a path p

List of Figures

List of Tables

Appendix A

Derivation of $\tilde{\alpha}$

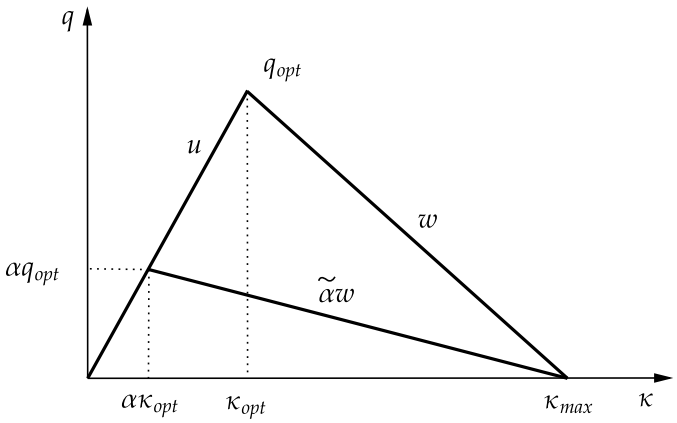


Figure A.1: Illustration of $\tilde{\alpha}$ in the FD.

The capacity q_{opt} can be written as $q_{opt} = u \cdot \kappa_{opt}$ and $q_{max} = w \cdot (\kappa_{max} - \kappa_{opt})$. From this follows that:

$$u = w \cdot \frac{\kappa_{max} - \kappa_{opt}}{\kappa_{opt}}. \tag{A.1}$$

Furthermore, from the fundamental diagram as depicted in Figure A.1 we can write:

$$\alpha q_{max} = (\kappa_{max} - \alpha \cdot \kappa_{opt}) \cdot \tilde{\alpha} \cdot w. \tag{A.2}$$

Then, we can reformulate this equation to:

$$\tilde{\alpha} = \frac{\alpha \cdot q_{max}}{w \cdot (\kappa_{max} - \alpha \cdot \kappa_{opt})} \tag{A.3}$$

By putting $q_{max} = u \cdot \kappa_{opt}$, diving by k_{opt}, and reformulating the denominator, and finally

utilizing eq.(A.1), we get the following:

$$\begin{aligned}
\tilde{\alpha} &= \frac{\alpha \cdot q_{max}}{w \cdot (\kappa_{max} - \alpha \cdot \kappa_{opt})} \\
&= \frac{\alpha \cdot u \cdot \kappa_{opt}}{w \cdot (\kappa_{max} - \alpha \cdot \kappa_{opt})} \\
&= \frac{\alpha \cdot u}{\frac{w}{\kappa_{opt}} \cdot (\kappa_{max} - \alpha \cdot \kappa_{opt})} \\
&= \frac{\alpha \cdot u}{\frac{w}{\kappa_{opt}} \cdot (\kappa_{max} - \kappa_{opt} + (1-\alpha) \cdot \kappa_{opt})} \\
&= \frac{\alpha \cdot u}{\frac{w}{\kappa_{opt}} \cdot (\kappa_{max} - \kappa_{opt}) + \frac{w}{\kappa_{opt}} \cdot (1-\alpha) \cdot \kappa_{opt}} \\
&= \frac{\alpha \cdot u}{\frac{w}{\kappa_{opt}} \cdot (\kappa_{max} - \kappa_{opt}) + w \cdot (1-\alpha)} \\
&= \frac{\alpha \cdot u}{u + w \cdot (1-\alpha)}
\end{aligned} \tag{A.4}$$

This concludes the derivation of $\tilde{\alpha}$ as used in the Chapter 5.

Appendix B

Scientific contributions

This section lists all journal and conference papers, held presentations, and presented posters that originated during the doctorate.

B.1 Journal papers

G. TILG, L. AMBÜHL, S. F. BATISTA, M. MENENDEZ, L. LECLERCQ & F. BUSCH [n.d.]: From corridor to network macroscopic fundamental diagrams: A semi-analytical estimation approach. Submitted for publication.

S. BATISTA, G. TILG & M. MENENDEZ [2021a]: Exploring the potential of aggregated traffic models for estimating network-wide emissions. In: *Transportation Research Part D: Transport and Environment*. Submitted for publication.

L. S. HAMM, A. LODER, G. TILG, M. MENENDEZ & K. BOGENBERGER [2022b]: Network Inefficiency: Empirical Findings for Six European Cities. In: *Transportation Research Record* 0.0, pp. 1–13. DOI: 10.1177/03611981221082588

G. TILG, L. AMBÜHL, S. F. BATISTA, M. MENENDEZ & F. BUSCH [2021b]: On the application of variational theory to urban networks. In: *Transportation Research Part B: Methodological* 150, pp. 435–456. DOI: 10.1016/j.trb.2021.06.019

G. TILG, S. AMINI & F. BUSCH [2020a]: Evaluation of analytical approximation methods for the macroscopic fundamental diagram. In: *Transportation Research Part C: Emerging Technologies* 114, pp. 1–19. ISSN: 0968-090x. DOI: doi.org/10.1016/j.trc.2020.02.003

F. DANDL, R. ENGELHARDT, M. HYLAND, G. TILG, K. BOGENBERGER & H. S. MAHMASSANI [2021]: Regulating mobility-on-demand services: Tri-level model and Bayesian optimization solution approach. In: *Transportation Research Part C: Emerging Technologies* 125,

p. 103075

G. TILG, Z. UL ABEDIN, S. AMINI & F. BUSCH [2020b]: Simulation-based design of urban bi-modal transport systems. In: *Frontiers in Future Transportation* 1. ISSN: 2673-5210. DOI: 10.3389/ffutr.2020.581622

G. TILG, K. YANG & M. MENENDEZ [2018]: Evaluating the effects of automated vehicle technology on the capacity of freeway weaving sections. In: *Transportation Research Part C: Emerging Technologies* 96, pp. 3–21. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2018.09.014

B.2 Conference papers

G. TILG, A. PAWLOWSKI & K. BOGENBERGER [2021d]: The impact of data characteristics on the estimation of the three-dimensional passenger macroscopic fundamental diagram. In: *2021 24th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2111–2117. DOI: 10.1109/ITSC48978.2021.9564430

F. DANDL, G. TILG, M. ROSTAMI-SHAHRBABAKI & K. BOGENBERGER [2020]: Network fundamental diagram based routing of vehicle fleets in dynamic traffic simulations. In: *2020 23th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–8. DOI: 10.1109/itsc45102.2020.9294204

S. AMINI, G. TILG & F. BUSCH [2018]: Evaluating the impact of real-time traffic control measures on the resilience of urban road networks. In: *2018 21th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 519–524. DOI: 10.1109/itsc.2018.8569678

B.3 Conference presentations

G. TILG, M. SKRECKI, F. BAHR, A. KELER, A. TSAKARESTOS, R. GÖBEL & F. BUSCH [2021a]: Assessing the spatial impacts of unreliable public transport systems: A quasi real-time data-driven approach. In: *Presented at the 8th International Symposium on Transport Network Reliability*

S. AMINI, L. AMBÜHL, G. TILG, K. BOGENBERGER & M. MENENDEZ [2020a]: Generating and calibrating large-scale, mesoscopic SUMO networks. In: *SUMO User Conference 2020*

S. AMINI, G. TILG, A. CAGRI TEKIN, D. TIDDI & F. BUSCH [2019a]: Automatic calibration of the link fundamental diagram for macroscopic traffic simulation models. In: *mobil.TUM 2019*. Munich, Germany

S. AMINI, G. TILG & F. BUSCH [2019b]: Calibration of mesoscopic simulation models for urban corridors based on the macroscopic fundamental diagram. In: *Presented at the 8th Symposium of the European Association for Research in Transportation*

S. AMINI & G. TILG [2018]: Estimating the effects of temporary bottlenecks on the capacity of urban arterial: An MFD-approach. In: *Presented at the 7th Symposium of the European Association for Research in Transportation*

B.4 Conference posters

L. HAMM, A. LODER, G. TILG, M. MENENDEZ & K. BOGENBERGER [2022a]: Network inefficiency - Empirical findings for six European cities. In: *Presented at the 101th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA

G. TILG, L. AMBÜHL, S. BATISTA, M. MENENDEZ, L. LECLERCQ & F. BUSCH [2021c]: Semi-analytical estimation of macroscopic fundamental diagrams: From corridors to networks. In: *Presented at the 100th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA

S. AMINI, G. TILG & F. BUSCH [2020b]: Macroscopic traffic dynamics in urban networks during incidents. In: *Presented at the 99th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA

G. TILG, S. AMINI & F. BUSCH [2019]: Arterial macroscopic fundamental diagram: A comparison of analytical approximations and empirical data from Munich. In: *Presented at the 98th Annual Meeting of the Transportation Research Board*. Washington, D.C., USA

Appendix C

Key papers for the dissertation

C.1 Short summaries and authors' contributions

Below, short summaries based on the original abstracts and specifications of the authors' contributions are provided.

Evaluation of analytical approximation methods for the macroscopic fundamental diagram

The macroscopic fundamental diagram (MFD) describes the relation of average network flow, density, and speed in urban networks. It can be estimated based on empirical or simulation data or approximated semi-analytically. Two main semi-analytical approximation methods to derive the MFD for arterial roads and urban networks exist at the moment. These are the method of cuts and related approaches as well as the stochastic approximation. This paper systematically evaluates these methods including their most recent advancements for the case of an urban arterial MFD. Both approaches are evaluated based on a traffic data set for a segment of an arterial in the city of Munich, Germany. This data set includes loop detector and signal data for a typical working day. It is found that the deterministic method of cuts finds a more accurate upper bound for the MFD for the studied case. The estimation error of the stochastic method is about three times higher than the one of the deterministic method. However, the stochastic approximation outperforms the method of cuts in approximating the free-flow branch of the MFD. The analysis of the discrepancies between the empirical and the analytical MFDs includes an investigation of the measurement bias and an in-depth sensitivity study of signal control and public transport operation-related input parameters. This study is conducted as a Monte-Carlo-Simulation based on a Latin Hypercube sampling. Interestingly, it is found that applying the method of cuts for a high number of feasible green-to-cycle ratios predicts the empirical MFD well. Overall, it is concluded that the availability of signal data can improve the analytical approximation of the MFD even for a highly inhomogeneous arterial.

G. Tilg conducted the majority of the work. In particular, he identified the research gap, designed the analysis concept and methodology, implemented the method of cuts and the stochastic approximation, conducted the case study as well as the sensitivity analysis

based on the Monte Carlo Simulation, and wrote, reviewed, and edited the manuscript. S. Amini contributed to the case study by processing the empirical data and performing the microscopic simulation, reviewed, and edited the manuscript. F. Busch reviewed the concept and supervised the study completion.

On the application of variational theory to urban networks

The well-known Lighthill–Whitham–Richards theory is the fundamental pillar for most macroscopic traffic models. In the past, many methods were developed to numerically derive solutions for problems described according to this theory. Examples of such numerical solution schemes are the cell transmission model, the link transmission model, and the variational theory (VT) of traffic flow. So far, the eulerian formulation of VT found applications in the fields of traffic modeling, macroscopic fundamental diagram approximation, multi-modal traffic analyses, and data fusion. However, these studies apply VT only at the link or corridor level. To the best of our knowledge, there is no methodology yet to apply VT at the network level. We address this gap by developing a VT-based framework applicable to networks. Our model allows us to account for source terms (e.g. inflows and outflows at intersections) and the propagation of spillbacks between adjacent corridors consistent with kinematic wave theory. We show that the trajectories extracted from a microscopic simulation fit the predicted traffic states from our model for a simple intersection with both source terms and spillbacks. We also use this simple example to illustrate the accuracy of the proposed model, and the ability to model complex bottlenecks. Additionally, we apply our model to the Sioux Falls network and again compare the results to those from a microscopic simulation in line with kinematic wave theory. Our results indicate a close fit of traffic states, but with substantially lower computational cost. The developed methodology is useful for extending existing VT applications to the network level, for network-wide traffic state estimations in real-time, or other applications within a model-based optimization framework.

G.Tilg conducted the majority of the work. In particular, he had the original idea of extending variational theory to networks, designed the mathematical concept of the network variational theory, implemented the model, conducted the case study and the small examples, analyzed the results, and wrote, reviewed, and edited the manuscript. L. Ambühl, S. Batista, M. Menendez reviewed the concept, the case study, contributed conceptually to the results analysis, and reviewed the manuscript. F. Busch reviewed the concept and supervised the study completion.

From corridor to network macroscopic fundamental diagrams: A semi-analytical estimation approach

The design of network-wide traffic management schemes or transport policies for urban areas requires computationally efficient traffic models. The MFD is a promising tool for such applications. Unfortunately, empirical MFDs are not always available, and semi-analytical approximation methods require a reduction of the network to a corridor which introduces substantial inaccuracies. We propose a semi-analytical methodology to estimate the MFD for

realistic urban networks, without the information loss induced by the reduction of networks to corridors. The methodology is based on the method of cuts, but applies to networks with irregular topologies, accounts for different spatial demand patterns, and determines the upper bound of network flow. Thereby, we consider flow conservation and the effects of spillbacks, both at the network level. Our framework decomposes the network into a set of corridors, creates a hypernetwork including the impacts of source terms, and then treats the dependencies across corridors (e.g. due to turning flows and spillbacks). Based on this hypernetwork, we derive the free-flow and capacity branch of the MFD. The congested branch is estimated by considering gridlock characteristics and utilizing recent advancements in MFD research. We showcase the applicability of the proposed methodology in a case study with a realistic setting based on the Sioux Falls network. We then compare the results to the original method of cuts, and a ground truth derived from the cell transmission model. The results analysis reveals that our method is over five times more accurate than the state of the art in estimating the network-wide capacity and jam density. Moreover, they clearly indicate the MFD's dependency on spatial demand patterns. Compared to simulation-based MFD estimation approaches, the potential of the proposed framework lies in the modeling flexibility, explanatory value, and reduced computational cost.

G.Tilg conducted the majority of the work. In particular, he designed the theoretical framework, designed the semi-analytical models, and integrated the network variational theory. Moreover, he implemented the framework, conducted the case study, analyzed the results, and wrote, reviewed, and edited the manuscript. L. Ambühl, S. Batista, M. Menendez, L. Leclercq reviewed the concept, contributed to the case study by suggesting specific aspects to be analyzed, contributed to the results analysis itself, and reviewed the manuscript. F. Busch reviewed the concept and supervised the study completion.

C.2 Publishers' agreements

Tilg, Gabriel

Von: Permissions Helpdesk <permissionshelpdesk@elsevier.com>
Gesendet: Freitag, 3. September 2021 17:05
An: Tilg, Gabriel
Betreff: Re: Permission to use published and submitted papers within my dissertatio [210902-021324]

Dear Gabriel Tilg

Thank you for writing to us.

You certainly can include your published papers and submitted manuscript in your thesis. As long as the papers are embedded inside your thesis you can include them anywhere in your thesis. And you are allowed to publish your thesis online.

You do not need to obtain formal permission for this.

However, if you wish you have permission license I am happy to send you one.

Kind regards,

Roopa Lingayath
Senior Copyrights Coordinator
ELSEVIER | HCM - Health Content Management

Visit Elsevier Permissions

From: Administrator
Date: Thursday, September 02, 2021 02:16 PM GMT

Dear Customer

Thank you for contacting Elsevier's Permissions Helpdesk.

This is an automated acknowledgement to confirm we have received your query. Ticket number 210902-021324 has been opened on your behalf and we aim to respond within five business days.

Regards,

Permissions Helpdesk

From: Gabriel Tilg
Date: Thursday, September 02, 2021 02:16 PM GMT

My name is Gabriel Tilg and I am the first authors of the papers entitled “Evaluation of analytical approximation methods for the macroscopic fundamental diagram” (https://doi.org/10.1016/j.trc.2020.02.003) and “On the application of variational theory to urban network” (https://doi.org/10.1016/j.trb.2021.06.019) published in Transportation Research part C and Part B, respectively. Also, I am the first author of the manuscript entitled “From corridor to network macroscopic fundamental diagrams: A semi-analytical estimation approach” which was submitted for publication to Transportation

Research Part B.

I would like to include these published papers as well as the submitted manuscript in my dissertation. Three chapters would be heavily based on the contents of these papers (only minor changes to increase the consistency would be applied). It will be clearly stated that the dissertation is publication-based, and references to the original papers will be made. Moreover, the original papers would be put into the appendix of the dissertation. The dissertation would be published online by the University library.

I hereby ask for your permission to include the published and submitted papers as described above.

I look forward to your answer.

Many thanks in advance and best regards.
Gabriel Tilg

C.3 Original paper versions

Transportation Research Part C 114 (2020) 1–19

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Evaluation of analytical approximation methods for the macroscopic fundamental diagram

Gabriel Tilg*, Sasan Amini, Fritz Busch

Chair of Traffic Engineering and Control, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Germany

ARTICLE INFO

Keywords:
Macroscopic Fundamental Diagram (MFD)
Traffic flow theory
Empirical analysis
Sensitivity analysis
Monte-Carlo-simulation
Microscopic simulation

ABSTRACT

The Macroscopic Fundamental Diagram (MFD) describes the relation of average network flow, density and speed in urban networks. It can be estimated based on empirical or simulation data, or approximated analytically. Two main analytical approximation methods to derive the MFD for arterial roads and urban networks exist at the moment. These are the method of cuts (MoC) and related approaches, as well as the stochastic approximation (SA). This paper systematically evaluates these methods including their most recent advancements for the case of an urban arterial MFD. Both approaches are evaluated based on a traffic data set for a segment of an arterial in the city of Munich, Germany. This data set includes loop detector and signal data for a typical working day. It is found that the deterministic MoC finds a more accurate upper bound for the MFD for the studied case. The estimation error of the stochastic method is about three times higher than the one of the deterministic method. However, the SA outperforms the MoC in approximating the free-flow branch of the MFD. The analysis of the discrepancies between the empirical and the analytical MFDs includes an investigation of the measurement bias and an in-depth sensitivity study of signal control and public transport operation related input parameters. This study is conducted as a Monte-Carlo-Simulation based on a Latin Hypercube sampling. Interestingly, it is found that applying the MoC for a high number of feasible green-to-cycle ratios predicts the empirical MFD well. Overall, it is concluded that the availability of signal data can improve the analytical approximation of the MFD even for a highly inhomogeneous arterial.

1. Introduction

Modeling urban traffic has been a major concern since the early days of traffic research. Urbanization and the associated growth in demand increase the need for research on efficient models of urban traffic on a network-wide scale. Early studies date back to the 1960s (Godfrey, 1969; Herman and Prigogine, 1979; Mahmassani et al., 1987) in which the idea of a macroscopic network-level relation of traffic flow and accumulation was introduced. Later, Daganzo (2007) formulated a macroscopic relationship between the network-wide outflow and the aggregated accumulation. This was verified in Geroliminis and Daganzo (2008) that investigated the existence of the macroscopic fundamental diagram (MFD) by linking the vehicle accumulation to the space-mean flow in a network. They found a well-defined and low-scatter relationship for homogeneous regions which is insensitive to small changes in demand for empirical data from Yokohama. Such a property makes the MFD a promising tool for a wide range of applications in urban traffic control and modeling. Studies analyzing such potential vary from gating (Keyvan-Ekbatani et al., 2012; Aboudolas and Geroliminis, 2013; Girault et al., 2016; Yang et al., 2018), pricing (Zheng et al., 2016), perimeter control (Geroliminis et al., 2012; Kouvelas et al.,

* Corresponding author.
E-mail addresses: gabriel.tilg@tum.de (G. Tilg), sasan.amini@tum.de (S. Amini), fritz.busch@tum.de (F. Busch).

https://doi.org/10.1016/j.trc.2020.02.003
Received 28 May 2019; Received in revised form 28 January 2020; Accepted 2 February 2020

2017; Zhong et al., 2018; Sirmatel and Geroliminis, 2018), routing (Yildirimoglu et al., 2018; Amini et al., 2018), the investigation of the impact of a network's topology on its performance (Knoop et al., 2014b; Ortigosa et al., 2017), to the analysis of the effects of parking on urban traffic dynamics (Cao and Menendez, 2015; Leclercq et al., 2017) and traffic control in a connected environment (Yang et al., 2019).

Inspired by the findings of Geroliminis and Daganzo (2008) several studies investigated the existence of a similar relationship in a number of cities around the world. This includes Yokohama, Japan (Geroliminis and Daganzo, 2008), Toulouse, France (Buisson and Ladier, 2009), Brisbane, Australia (Tsubota et al., 2014), Zurich, Switzerland (Ambühl et al., 2017), Lucerne, Switzerland and London in the United Kingdom (Ambühl et al., 2018), and many other cities (Loder et al., 2019). The data sets for these MFDs came mostly from loop detectors and probe vehicles. However, these data have to be treated with care as several biases are included. Methods to reduce the impact of these biases are reported in the literature (Ambühl et al., 2017; Du et al., 2016; Leclercq et al., 2014). Moreover, a heterogeneous density distribution might lead to hysteresis in the MFD (Geroliminis and Sun, 2011; Gayah and Daganzo, 2011; Saberi and Mahmassani, 2013; Shim et al., 2019; Ambühl et al., 2018). New technologies and associated data, such as mobile phone traces, can help to reduce the data measurement related biases.

Even though empirical observations are essential, they come with a number of limitations. First and foremost, they do not allow a systematic analysis of the characteristics of the MFD and the influential factors on its shape. Secondly, they require extensive data processing as real sensors are prone to failure, bias and incomplete coverage of the network. Thirdly, many cities do not possess empirical data and, hence, are not able to empirically derive the MFD. Thus, some simulation-based studies on MFD have been conducted. Generally speaking, the results of these studies (e.g. Duruisseau and Leclercq, 2018; Gayah et al., 2014; Knoop et al., 2014a; Ji et al., 2010) show that the MFD depends on link characteristics, the network topology, traffic control, and route choice.

Despite the fact that simulation studies show promising results, the application of large-scale simulations might be limited due to the burdensome calibration. Additionally, if the application of the MFD is aiming towards real-time evaluation of traffic control measures, the high computational cost of simulation-based methods might reduce their suitability. However, existing analytical methods rectify these drawbacks of simulation-based and empirical estimation of the MFD. These methods are mainly based on the framework of the variational theory (VT) (Daganzo, 2005a; Daganzo, 2005b; Daganzo and Menendez, 2005). The corresponding concept is described by Daganzo and Geroliminis (2008), who introduced the method of cuts (MoC) that is able to find an upper bound for the arterial MFD. Leclercq and Geroliminis (2013) continued this line of research, improved the MoC and showed the effects of route choice on a simple network. The concept was further enhanced by Daganzo and Lehe (2016), who formulated the method as a linear program. Contrarily to this deterministic approach, Laval and Castrillón (2015) developed a stochastic analytical method based on the MoC and the renewal theory, called stochastic approximation (SA) hereafter. The method enables to approximate the MFD for urban arterial roads as well as networks. The authors found the MFD mainly to be influenced by the mean-block-length-to-green ratio and the mean-red-to-green ratio.

Analytical methods are particularly suitable to run extensive sensitivity analyses which provide insights on the impact of certain parameters on the shape of the MFD. For example, Geroliminis and Boyacı (2012) investigated the effects of different traffic signal offsets and link lengths on the MFD. Daganzo et al. (2018) studied the impact of adaptive signal offsets on the MFD. Moreover, these methods allow the performance of an ex-ante prediction of the MFD. Such a prediction is crucial for the applicability of MFD-based modeling techniques. For instance, Ambühl et al. (2018) developed a method to derive a functional form with a physical meaning for the MFD. They state that next to empirical data, also analytically derived MFDs can deliver essential input to the estimation of the functional form. Moreover, Mariotte et al. (2017) analyzed different ways to derive the outflow from regions in which traffic dynamics are described based on an MFD. This input MFD can be derived based on analytical approximations. Mariotte and Leclercq (2019) proposed a new flow transfer scheme that considers multiple trip lengths in one region. Batista et al. (2019) developed a methodology for adapting trip lengths for multi-regional MFD-based applications. The methodology of both works requires MFDs as input which can be derived analytically.

These works confirm the importance of analytical approximation methods in MFD-based modeling of urban traffic. Naturally, this leads to the question of how well empirically derived MFDs can be approximated. The original method of practical cuts (Daganzo and Geroliminis, 2008) and the SA (Laval and Castrillón, 2015) have both been validated with data from Yokohama, Japan. However, to the best knowledge of the authors, no such verification has been reported for the MoC for inhomogeneous arterials, nor has the SA been directly compared to the MoC for such an arterial. Additionally, many advances on empirical MFD estimation methods have been achieved to date, which were not considered in these studies. Signal data is an input for both analytical approximation methods. Yet, no information on detailed signal data is available for the case of Yokohama. Therefore, it remains unclear to what extent the existence of such data can improve the estimation accuracy of the analytical methods. Furthermore, extensions to consider the effects of public transport operation exist for both methods (Xie et al., 2013; Castrillon and Laval, 2017). Nonetheless, no reports about any evaluation of them based on comparison to empirical data are known.

This paper aims to address the mentioned research gaps by comparing the MoC and the SA including the most recent extension to consider public transport operation to an empirically derived arterial MFD. Even though the SA applies to networks, the MoC was originally developed for arterials. Thus, we choose an arterial for this study. The empirical data at hand consists of loop detector data (LDD) and signal data. The LDD come from an inhomogeneous road segment in Munich, Germany, with multi-modal traffic, active actuated signal control, and transit signal priority. The occurring modes include private cars and buses, trams on physically separated tracks and bicycle traffic on bike lanes. This allows investigating the effects of infrastructure complexity on differences between empirically and analytically derived MFDs. The three dimensional passenger MFD (e.g. Geroliminis et al., 2014; Loder et al., 2017) extends the idea of the MFD to multi-modal systems. It relates the passenger production to the network bus and car accumulation. However, such an analysis requires additional data and lies beyond the scope of this paper. Thus, we focus on the 2D-MFD

Table 1
List of abbreviations.

Abbreviation	Term
BVP	Boundary value problem
CDF	Cumulative distribution function
ELDD	Empirical loop detector data
FCD	Floating car data
FD	Fundamental diagram
IVP	Initial value problem
KWT	Kinematic wave theory
LDD	Loop detector data
LOESS	Locally estimated scatter plot smoothing
MFD	Macroscopic fundamental diagram
MoC	Method of cuts
OD	Origin–Destination
RMSE	Root Mean Square Error
SA	Stochastic approximation
VFCD	Virtual floating car data
VLDD	Virtual loop detector data
VT	Variational theory

approximated by the MoC and the SA including their public transport extensions. Thereby, we compare the according results to empirical data from the arterial in Munich. The results of this comparison are thoroughly analyzed. We set the focus on the measurement bias inevitably included in LDD, the signal data and the impact of public transport operation. No comparison of the analytical methods including the public transport extension has been conducted so far. Also, no investigation of the role and benefit of a rich signal data set to accurately estimate the MFD has been reported. Lastly, to the best knowledge of the authors, no attempts to quantify the effects of these elements on the difference between empirical and analytically approximated MFDs have been presented. Interestingly, the results of this paper indicate that applying the MoC for a wide range of signal data leads to an aggregated traffic pattern close to the empirical MFD. This is the main contribution of this paper. As an overall result, this study sheds light on the performance of the studied analytical approximation methods, as well as on the importance of corresponding assumptions and simplifications.

The remainder of the paper is structured as follows. Section 2 describes the underlying theory of the MoC and the SA. Moreover, the current state of research regarding the empirical MFD estimation is presented. Section 3 specifies the general methodology of the paper including a case study, the applied methods, the investigation of the measurement bias and the sensitivity analysis of input parameters related to signal control and public transport operation. The results of the case study and their discussion are presented in Section 4. Furthermore, the section includes the results of the investigation of the measurement bias based on a microscopic simulation in SUMO (Lopez et al., 2018), and of the sensitivity analysis of the analytical approximations. Finally, Section 5 draws closing conclusions, highlights limitations and outlines possible future research. To increase the readability of the paper, we provide a list of abbreviations used throughout the paper in alphabetical order in Table 1.

2. MFD estimation

This section describes analytical approximation and empirical estimation methods to derive the MFD. More specifically, the frameworks of the MoC and the SA are introduced as main representatives of analytical approximations. Besides, important aspects and related research of empirical MFD estimation are presented. This section as well as the following one include many variables. For the sake of clarity, a nomenclature including these variables is provided in Table 2.

2.1. Analytical approximation methods

In general, the herein described analytical approximations of the MFD are based on the VT. The VT enables to calculate the exact solutions of an initial or boundary value problem (IVP or BVP, respectively) according to the kinematic wave theory (KWT) (Lighthill and Whitham, 1955; Richards, 1956). The VT states that a kinematic wave problem can be solved as shortest path problem in the time–space plane. The corresponding costs are related the boundary data and the maximum number of vehicles that can pass a moving observer traveling along a given path. The mathematical formulation is:

$$N(P) = \inf_{B \in \mathscr{B}_P} \{N(B) + \Delta_{BP}\} \tag{1}$$

where $N(P)$ is the cumulative count of vehicles at the generic point $P(x, t)$, $\mathscr{B}_P$ is the boundary data in the domain of dependence of P, Δ_{BP} is the maximum number of vehicles that can pass a moving observer which travels from the point B on the boundary to P. For more details, please refer to Daganzo (2005a,b) and Daganzo and Menendez (2005).

Table 2
Nomenclature.

Fundamental diagram		
u	[km/h]	Free flow speed
w	[km/h]	Backward wave speed
k_j	[veh/km]	Jam density
q_{cap}	[veh/h]	Capacity
Variational theory		
x	[km]	Spatial dimension
t	[h]	Temporal dimension
$P(x, t)$	[–]	Generic point at (x, t)
$N(P)$	[veh]	Cumulative vehicle count at $P(x, t)$
$\mathscr{B}_P$	[–]	Boundary data in the domain of dependence of P
Δ_{BP}	[veh]	Maximum number of vehicles that can pass a moving observer which travels from B to P
Method of cuts		
q	[veh/h]	Flow
k	[veh/km]	Density
$Q(k)$	[veh/h]	MFD
j	[–]	Index of cut C
C_j	[–]	Cut
v_j	[km/h]	Mean speed of cut C_j
$R(v_j)$	[veh/h]	Costs of cut C_j
v_b	[km/h]	Average speed of buses
Stochastic approximation		
v_O	[km/h]	Speed of moving observer
q	[veh/h]	Flow
k	[veh/km]	Density
k'	[–]	Dimensionless density
μ_g	[s]	Mean green time
μ_r	[s]	Mean red time
ρ	[–]	Long-run red to green ratio
μ_l	[m]	Mean block lengths
λ	[–]	Mean dimensionless block length
δ	[–]	Average coefficient of variation
$F_{q(k')}(q)$	[veh/h]	CDF of the MFD
s_0	[–]	Strategy related to $v_O = 0$
$\Omega = \{s_1, s_2\}$	[–]	Strategies related to $v_O = [u, w]$
B	[–]	Number of intersections
p_s	[–]	Probability that bus stops within a link
u_b	[km/h]	Average bus free flow speed
t_d	[s]	Average dwell times of buses
Sensitivity analysis		
n	[–]	Number of sample variations for Monte-Carlo-Simulation
q_{min}	[veh/h]	Minimum of n results for $Q(k)$
q_{mean}	[veh/h]	Mean of n results for $Q(k)$
q_{max}	[veh/h]	Maximum of n results for $Q(k)$

2.1.1. Method of cuts

Daganzo and Geroliminis (2008) introduced a methodology to analytically approximate the MFD for homogeneous arterials with identical block lengths and signal control parameters. This methodology was further extended by Leclercq and Geroliminis (2013) to be applicable to heterogeneous topologies.

In order to set up an IVP several input parameters need to be specified. These include the number of links in the arterial, the number of lanes on and the length of each link, signal parameters and the fundamental diagram (FD). This defines the solution space and consequently the shortest path problem. The solution space in the (t, x)-plane consists of horizontal edges representing intersections including both green and red phases. Note that green times and cycle lengths are not allowed to vary within a specific intersection. In other words, only fixed time signal control can be modeled. Additionally, slanted edges with extremal slopes (free flow speed u or backward wave speed w) starting at bottleneck termini connect the horizontal edges related to different intersections. Whenever such a slanted edge hits a red phase, it is terminated. Based on the slope of these edges, costs for each edge can be defined. The corresponding values are derived based on the VT and relate to the maximum passing rate a moving observer would observe

traveling along such an edge. This procedure leads to a numerical representation of a kinematic wave problem which is called global variational graph (Leclercq and Geroliminis, 2013).

Depending on the initial density, the flow at a certain point approaches a location-dependent limit for the case of a steady state. Daganzo and Geroliminis (2008) describe this flow based on Eq. (1) as $q = N(P)/t$, $t = t_0 \rightarrow \infty$. Thus, an upper bound of the flow q can be derived for each k by evaluating Eq. (2) based on the global variational graph as follows:

$$q = \inf_{v_j} \{k v_j + R(v_j)\} \tag{2}$$

with $q = Q(k)$ as the MFD, k as the density, v_j as the average speed of a moving observer j and $R(v_j) = \Delta_{BP}/t$, $t = t_0 \rightarrow \infty$ as the related maximum passing rate.

Evaluating Eq. (2) for several k and a specific moving observer with average speed v_j leads to a constraint of the flow q along with the evaluated range k. Repeating this procedure for several different average moving observer speeds v_j results in a family of cuts. The lower envelope of all cuts is the MFD $Q(k)$.

The original method (Daganzo and Geroliminis, 2008) is only applicable on homogeneous arterials which herein means that block lengths, green times and cycle lengths do not differ between intersections. Even though no signal data were available, the results of the analytical approximation fit an empirical MFD for the city of Yokohama. The authors stated that differences in the predicted and the measured MFD were due to inhomogeneity in the network and errors in the data. However, as Daganzo and Lehe (2016) pointed out, this method does not necessarily lead to a tight bound for stationary cuts, which relate to an average moving observer speed of $v_j = 0$. The authors rectified this drawback and formulated the overall method as a linear program. Nevertheless, their approach to estimating a tight bound for stationary observers is transferable to the MoC.

Xie et al. (2013) further extended the MoC to approximate the effects of public transport vehicles on the capacity of an arterial. The additional input parameters consist of the average bus speed v_b and the headways. The effects of buses are modeled based on the numerical representation of moving bottlenecks. A moving bottleneck is approximated by reducing the capacity of a certain link. That is, the costs of the corresponding green phase edges at the upstream and downstream intersection are adjusted. Thus, the effects of public transport can be included in the variational graph and the remaining procedure to derive cuts does not need to be adjusted.

2.1.2. Stochastic approximation

A different approach is presented by Laval and Castrillón (2015). They introduced the SA of the MFD for urban arterials and networks. Again, the core of the method is based on the VT and the MoC. Three types of cuts are evaluated. They are associated with different strategies which are explained in the following. Strategy s_0 is related to a stationary observer with a mean moving observer speed $v_O = 0$. Strategy s_1 relates to moving observers traveling with speed $v_O = [u, w]$ which stop when they cross a red phase. There they travel horizontally to the beginning of the current red phase. From there they resume the speed of $v_O = [u, w]$ until the next red phase is hit. Strategy s_2 is similar except for the fact that moving observers stop at every intersection independently if they cross a green or red phase. The fundamental difference of this method compared to the MoC is that the duration of green phases and cycle lengths do not have to be regular for its application. Thus, not only fixed time control settings can be analyzed but also networks or arterials where adaptive or actuated signal control is active.

An additional element of this framework is the transformation of the FD from the (q, k)-plane to a (q, k')-plane, where k' is the normalized density. This procedure leads to a symmetric and normalized FD. Thus, the related parameters are only required for the transformation back to the (q, k)-plane but not for the general application of the method. The only necessary input parameters are the mean green μ_g and mean red μ_r times, the mean block length μ_l and the related coefficients of variation δ which are assumed to be equal for simplicity. Based on these input parameters and the renewal reward theory, a framework is built to estimate the mean and coefficient of variation for cuts related to the described strategies. Assuming a normal distribution for the cuts, this framework enables to calculate a cumulative distribution function (CDF) for each cut and eventually for the MFD. Thus, it allows specifying certain percentiles of the MFD at (q, k)-pairs and therefore to derive its stochastic representation. Mathematically, the CDF of the MFD is expressed as shown in Eq. (3).

$$F_{q(k')}(q) = 1 - \left(1 - F_{s_0}\left(q\right)\right)^B \prod_{s \in \Omega} \left(1 - F_{s,k'}(q)\right) \tag{3}$$

where $F_{q(k')}(q)$ denotes the CDF of the MFD, q the flow, k' the normalized density, s_0 and $\Omega = \{s_1, s_2\}$ the different strategies and B the number of intersections. More details on the mathematical framework can be found in Laval and Castrillón (2015).

Major arterials in urban networks are often frequented by public transport vehicles. Possible impacts of multimodal traffic on the stochastic representation of the MFD for homogeneous arterials were reported in Castrillon and Laval (2017). In this case, homogeneity refers to equal signal parameters and block lengths amongst the modeled intersections. The authors modified the SA to analytically approximate these impacts and introduce three new parameters, the average bus free flow speed u_b, the stop probability p_s and the average bus dwell time t_d. Castrillon (2015) presented an extension of this method for heterogeneous arterials. This is done by adjusting model parameters to account for the stationary bottleneck effect of buses as well as for their role as moving bottleneck.

2.2. Empirical estimation methods

Next to analytical approaches, the MFD can be estimated based on data from fixed sensors or probe vehicles. This section presents related estimation techniques, and which drawbacks come with associated data sources. However, potential ways of dealing with

such drawbacks are not listed here. Instead, we refer to the corresponding studies. Eventually, we discuss the assumption of homogeneously distributed congestion as it has substantial effects on the shape of the MFD.

2.2.1. Data types

Loop detectors measure vehicle counts and the duration they are occupied. The data are often aggregated to 3- or 5-min intervals. Thus, average flows during these intervals, as well as occupancy values in percent can be derived. LDD might be faulty since malfunctions might occur during operation. Data recording systems apply basic algorithms which help to exclude the majority of technology-induced measurement errors. The first MFD based on LDD was shown in Geroliminis and Daganzo (2008). It was derived by averaging the link-weighted network density and flow. Even though a well-defined MFD was found for that case, empirical MFDs have to be treated with care as several biases were identified associated with LDD in the literature. For example, Wu et al. (2011) showed that high scatter in the MFD for an arterial road can result from the fact that loop detectors also capture transient states. Ambühl et al. (2017) described the loop detector placement bias. It refers to the fact that loop detectors only measure traffic states at a specific position, which is then often regarded as representative for the entire link. In addition, Courbon and Leclercq (2011) argued that loop detector positions need to be distributed uniformly across the links of the network to represent average traffic states on a global level. Zockaie et al. (2018) found that not only the position of loop detectors is crucial, but also the link selection plays an important role for the estimation of the MFD. This is because the measured traffic on selected links is related to Origin–Destination (OD) pairs. For a valid MFD estimation, these ODs have to represent the overall OD distribution in the network. Another bias associated with the link selection was described in Ambühl et al. (2017). The authors argued that loop detectors are usually installed on links more prone to congestion. Thus, this might result in an overestimation of network-wide congestion. Despite these biases, LDD are a commonly utilized data type for the MFD estimation due to their high availability compared to other data types.

As mentioned before, probe vehicles can be an alternative data source to estimate the MFD (e.g. Geroliminis and Daganzo, 2008). These data come either from GPS devices, mobile phones or are the result of processed Bluetooth data (e.g. Tsubota et al., 2014) and are often referred to as floating car data (FCD). They are considered to represent a better spatial coverage of traffic states in the network (Nagle and Gayah, 2014). Leclercq et al. (2014) showed that the MFD estimation based on probe vehicle data is more accurate than based on loop detectors. However, a crucial element is the penetration rate. It describes the relative number of probe vehicles in the total flow which is difficult to estimate for empirical data but essential for the MFD estimation. A potential way to cope with this drawback was shown by Nagle and Gayah (2014) who fused FCD with limited LDD for MFD estimation. However, they assumed that probe vehicles are uniformly distributed throughout the network. This might not be true since these data often come from certain vehicle types (e.g. taxi data) which might drive only on specific routes and thus biases might be introduced. Consequently, this leads to a heterogeneous distribution of trajectories. Du et al. (2016) investigated the effects of such heterogeneously distributed FCD on the MFD including data from probe vehicles with varying trip lengths. More analyses of the fusion of LDD and FCD for MFD estimation is reported in Zockaie et al. (2018) and Ambühl and Menendez (2016). Furthermore, Dakic and Menendez (2018) showed that data from automated vehicle location devices installed in public transport vehicles can contribute to the empirical MFD estimation.

Another kind of situation occurs for the case of microscopic simulation data. Under these circumstances, all vehicle trajectories are known and can be extracted. This enables to calculate the average flow, density and speed based on Edie's definition (Edie, 1963), which was referred to as Edie method in Leclercq et al. (2014). In such a case, no measurement biases are included in the data. Saberi et al. (2014) showed the existence of the MFD for a microscopic simulation of Chicago, Illinois, and Salt Lake City, Utah. They concluded that even without any measurement biases, the MFD still shows a range of flows for a given density. More specifically, they presented a hysteresis pattern in the MFD which occurs due to inhomogeneous congestion in the network.

2.2.2. Network homogeneity

The previous paragraphs describe the existing data sources for the estimation of the empirical MFD, as well as associated biases. Early works exploring the effects of heterogeneity in urban and freeway traffic networks on the MFD are Buisson and Ladier (2009), Mazloumian et al. (2010), Geroliminis and Sun (2011), Knoop et al. (2015) and Saberi and Mahmassani (2013). Even for the case of simulated and thus bias-free data, scatter in the MFD might still occur (Saberi et al., 2014). Based on empirical data, Ambühl et al. (2018) suggest several reasons for this. First, the heterogeneous spatial distribution of traffic density is supposedly a main factor. The MFD in its original definition is merely found for homogeneously congested networks. In reality, however, this is rarely observed. Flows measured during inhomogeneously distributed densities are lower than in the homogeneous case (e.g. Daganzo et al., 2011, Mühlich et al. (2014), Knoop et al. (2015)). It was found that the partitioning of large-scale networks into regional sub-networks can substantially decrease the heterogeneity of congestion in these smaller networks. Clustering methods for this partitioning are discussed in Saeedmanesh and Geroliminis (2016); Lopez et al., 2017; Saeedmanesh and Geroliminis, 2017; Ambühl et al., 2019. Secondly, the MFD is merely well-defined for steady states. However, changes in traffic states do not propagate instantaneously throughout the network. These dynamics can lead to multiple flows for specific density measurements (Mariotte et al., 2017). Lastly, public transport operation might have different impacts on traffic flow throughout the day (e.g. Arnet et al., 2015, Castrillon and Laval, 2017).

3. Methodology

This paper aims to perform an up-to-date evaluation of the existing methods to analytically approximate the MFD. A preliminary analysis for a data set of one working day of a road segment of Munich showed that the occurring differences cannot be explained by a

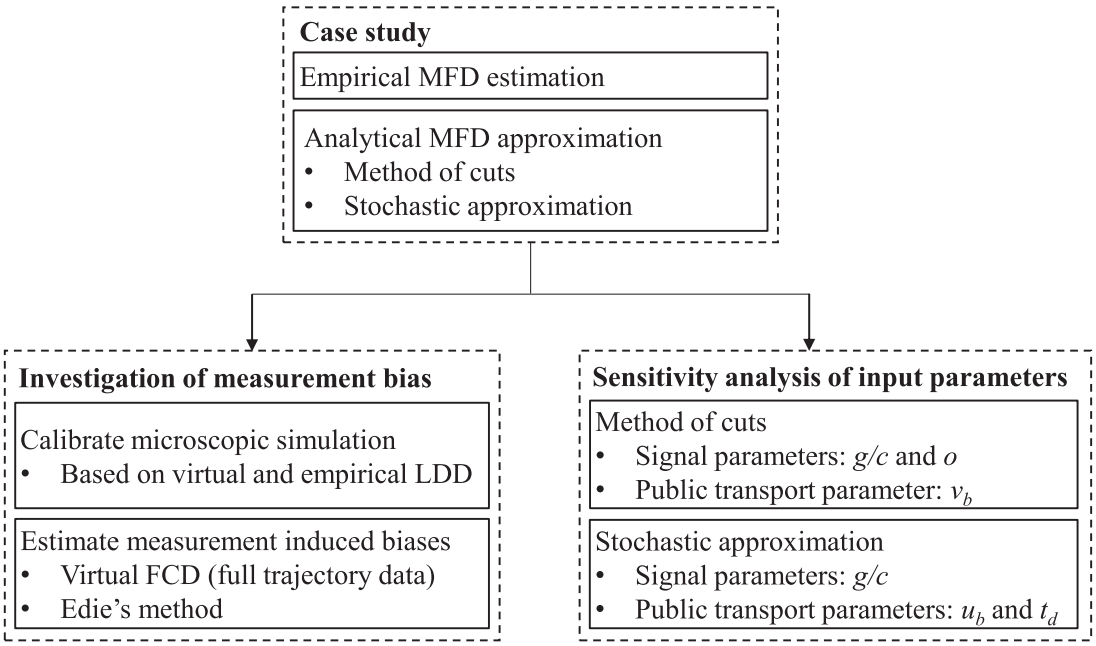


Fig. 1. General methodological framework.

simple scenario-specific evaluation (Tilg et al., 2019). Thus, a more sophisticated approach is proposed in this paper. The framework for the general methodology is shown in Fig. 1.

First, a case study based on a segment of the urban arterial Leopoldstraße located in the city of Munich, Germany, is conducted. The choice of an arterial road seems to be appropriate for this study since the MoC was originally developed for arterials. Thus, we minimize possible error sources by focusing on such an arterial. The analytical approximation (MoC and SA) and empirical estimation methods are applied based on data recorded on this road. The purpose of this case study is to investigate the fit of the analytical MFD approximations to the empirical MFD. Furthermore, it is the starting point for a systematic study of the appearing differences between the resulting MFDs.

As mentioned before LDD include a measurement bias that can affect the empirical derived MFD and thus also influence the difference to analytical approximated MFDs. Thus, a second step is to investigate this measurement bias. This is done by the means of microscopic simulation. The calibration of the simulation is carried out based on empirical and virtual LDD. To estimate the bias-free MFD, virtual trajectory data are used to calculate average flow and density according to Edie's method as shown in Leclercq et al. (2014). Based on the resulting bias-free MFD, the nature of the measurement bias can be analyzed.

Lastly, the role of signal data and public transport operation on the approximated MFD shape is investigated. The analytical methods and related input parameters imply certain simplifications of reality. The explicit studied parameters are explained below. The study is conducted as a sensitivity analysis based on a Monte-Carlo-Simulation for both analytical approximation methods. This study allows investigating the role of input parameters and analyzing the corresponding results in comparison to the empirical MFD.

The remaining section describes the layout of the case study, applied analytical approximation methods as well as the empirical MFD estimation, the investigation of the measurement bias, and the structure of the sensitivity analysis.

3.1. Case study

The Leopoldstraße leads from the center to the very north of Munich and serves as a critical corridor that connects the city center to the city's ring road. The empirical data containing raw LDD and signal data were collected from a segment of the northbound Leopoldstraße on a typical working day in October 2017. Comparable results for the time series of occupancies are found for other weekdays at the Leopoldstraße. Additionally, the distribution of green times and cycle lengths in a second available data set do not substantially differ to the one shown in the paper. Thus, we assume that analyzing the data presented in the paper is representative for the given road segment. Fig. 2 shows the layout of the studied road segment with a total length of 1.1 km. It includes five signalized intersections, labeled from one to five in the figure. The studied segment mostly consists of two lanes except for some of the intersections where a dedicated turning lane exists for right or left turning vehicles. The loop detectors on the turning lanes are stop-line detectors, which cannot be directly used to derive the MFD, and therefore, are not considered in this study. As it can be seen in Fig. 2, there are some minor unsignalized intersections along the arterial from which no data, i.e. number of turning and incoming vehicles, were available. For the sake of simplicity, the impact of these intersections is neglected.

Each of the five intersections is controlled by actuated traffic signals i.e. adaptive green time and cycle lengths as well as prioritized public transit vehicles. Right at the north-east corner of interSection 3 an important mobility hub, called *Münchner Freiheit*, i.e. a transfer point for public transport users between bus, tram, metro and car-sharing vehicles, is located. Moreover, another on-street bus stop is located between interSection 1 and 2. Two bus lines operate on the arterial, each with a 10-min headway during

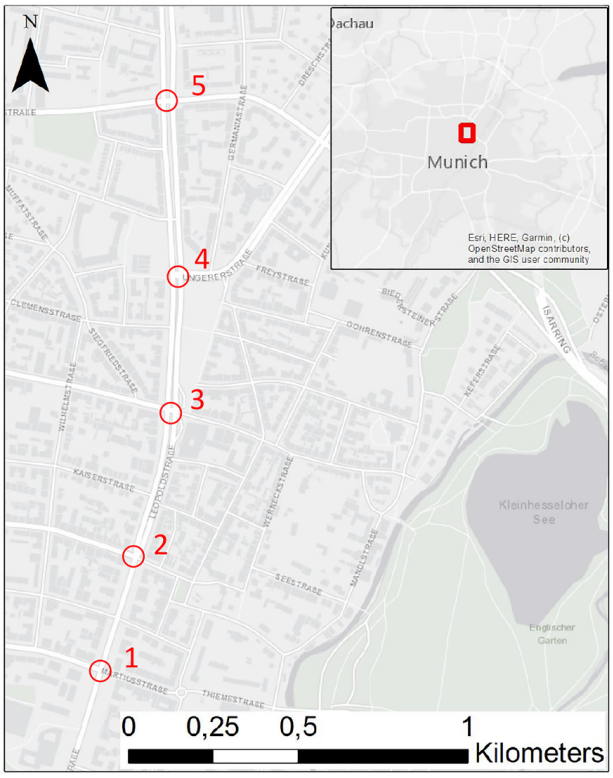


Fig. 2. The urban arterial Leopoldstraße in Munich, Germany. The numbers 1–5 refer to the examined intersections and increase in direction of travel.

most times of the day. Additionally, a tram line operates from interSection 3 to 5. However, the tram tracks are physically separated from the other lanes and thus do not impact traffic flow on the neighbouring lanes.

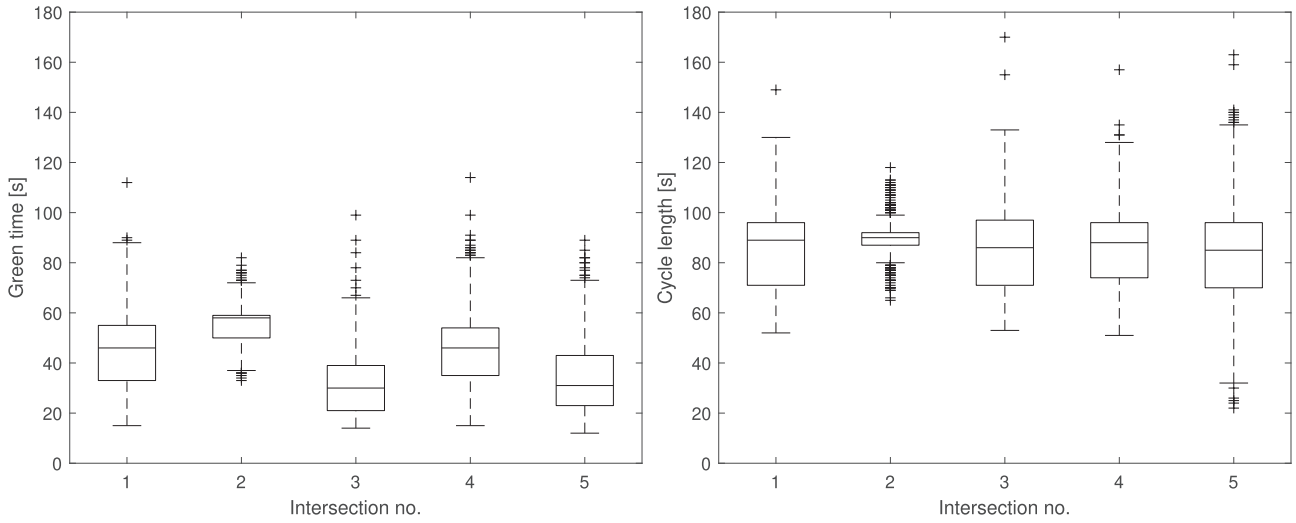
3.2. MFD estimation

The MoC and the SA are applied on the described road segment of the Leopoldstraße. Furthermore, an empirical MFD is estimated as described below. The quantitative comparison is made based on the root mean squared error (RMSE), which was successfully applied in the terms of MFD-related research (e.g. Ambühl et al., 2018).

3.2.1. Empirical estimation

The estimation of the empirical MFD for the case study is based on data recorded from loop detectors. Upstream of each intersection a loop detector is installed on each lane. In total, that sums up to 10 loop detectors for the northbound travel direction. All considered loop detectors are installed approximately 35 meters upstream of the stop-line. The average length of a loop detector is 1.5 m. Each observation consists of a time-stamp and an occupancy duration. A set of plausibility tests such as searching for duplicate detections, negative headways, and extreme occupancy values is performed. Subsequently, the data are aggregated in 5-min intervals. The result is a vehicle count and an occupancy value for each detector for each interval. In order to reduce the noise and errors due to aggregation, as well as to exclude outliers, we smooth the data using time series analysis as suggested in Ambühl et al. (2018). Then, the obtained occupancies are used to calculate densities assuming an average vehicle length of 5 meters. The density is calculated based on the sum of the mean vehicle and detector length (see Ambühl et al. (2018)).

Another data type available is signal data. This includes the time-stamp for each phase change, as well as the phase type (green, amber or red) for all intersections. Since the traffic signals are not pre-time controlled, the cycle lengths, green times and red times are strongly skewed. Fig. 3 shows boxplots for green times and cycle lengths of the northbound leg at each intersection for the measurement day. The ordinate shows the duration of the corresponding parameter in seconds. The intersection labels are displayed on the abscissa. As it is shown in the plots, substantial variance of both parameters can be observed. The reasons for this variance are the actuated traffic signal control, the transit signal priority, and the fact that the green times and cycle lengths of the entire day and thus different signal plans are included in the illustration. Additionally, it can be observed that the scatter is higher for longer green times. This is possibly due to the required minimum green times at the studied intersection legs. On the contrary, the distribution of cycle lengths seems to be symmetric. Moreover, the majority of green times of interSection 3 and 5 are substantially lower than the ones at the other intersections. Regarding interSection 3, this could be explained by the fact that it is located next to the *Münchner Freiheit* mobility hub, where many bus and tram lines stop. Thus, the effect of transit signal priority is very strong. Possible reasons for a lower mean green time at interSection 5 are the that a tram line crosses the intersection, and a bus line is running between the east-



a) Green times [s] b) Cycle lengths [s]

Fig. 3. Green times and cycle lengths for the studied intersections at Leopoldstraße (October, 2017).

and southbound leg of the intersection. Both are having priority and thus green times are reduced for private vehicles.

Furthermore, limited historical FCD are available. The data are too scarce to benefit from merging it with LDD for the empirical MFD estimation. However, it can be used to estimate the free flow speed u as explained below.

3.2.2. Method of cuts

The first analytical approximation for the MFD which is evaluated is the MoC. As described in Section 2, a global variational graph needs to be constructed to apply this method. For this purpose, the link FD, the arterial topology, the signal timings and the average bus speed v_b need to be defined.

The link FD plays an important role for the MoC. To define it, three out of the four related parameters, the free flow speed u, the capacity q_{cap}, the backward wave speed w and the jam density k_j have to be specified. It has a crucial impact on the approximation of the MFD. The free flow speed u is chosen according to FCD which were available for the entire analyzed segment of the Leopoldstraße. The found value for the free flow speed equals $u = 45$ km/h. The jam density is assumed to be $k_j = 150$ veh/km. This is based on visual inspection of a number of queues captured by high-resolution aerial pictures. The backward wave speed is assumed to be $w = 15$ km/h. Leclercq (2005) report similar values for urban roads in Toulouse, France. In this study, however, the jam density is found to be substantially higher. Thus, we choose a slightly larger value for w to calculate realistic values for the capacity q_{cap}. Given these three parameters and assuming a triangular FD, the capacity is calculated as about $q_{cap} = 1690$ veh/h.

The block lengths are measured from OpenStreetMap® (OpenStreetMap contributors, 2019) and are defined as the segment length between two consecutive signalized intersections. The effects of the minor intersection along the arterial are neglected for this study. This is due to the fact that the analytical approximation methods are not capable of considering the impact of such intersections. Also, the corresponding streets are of a minor level and partly only one-lane streets. Furthermore, no flow or occupancy data from these streets are available. As shown in Fig. 3, the green times and the cycle lengths are strongly skewed for all intersections. However, the MoC allows only for one value per intersection. In other words, only fixed time signal control can be modelled. Thus, we choose the average for both green times and cycle lengths for the case study. For the sensitivity study, a different approach is chosen as described below.

The last family of input parameters regards to public transport operation. The multimodal extension of the MoC (Xie et al., 2013) only needs an exogenous average bus speed as input parameter. As a first assumption, we choose 75% of the car free flow speed which equals to $v_b = 34$ km/h since no more accurate data is available. This value is based on a study of free flow speeds of different vehicle classes for Great Britain (Department for Transport, 2015). The ratio of car and bus free flow speeds in urban areas with a speed limit of about 50 km/h is 90%. Taking into account that there exists one on-street bus stop on the studied arterial, and assuming a dwell time of 30 s, a bus speed v_b of 75% of the car free flow speed u seems to be appropriate. However, the impact of this assumption is investigated as part of the sensitivity study. The dwell time of 30 s on average are chosen according to limited data from other bus lines which were recorded with GPS trackers.

Based on these input parameters a variational graph can be constructed. It allows deriving cuts C_j with speed v_j and costs $R(v_j)$ to estimate an upper bound for the arterial MFD. Since the Leopoldstraße represents a highly inhomogeneous arterial, an appropriate calculation time window is chosen to generate a sufficiently large solution space. Next, all possible moving observers are taken into account. For cuts with the same speeds, the ones with the lowest costs are chosen. The large size of the calculation time window ensures that all possible paths are found. Finally, taking the lower envelope of all cuts leads to an estimate of the upper bound for the MFD on the given arterial.

3.2.3. Stochastic approximation

The SA is the second method being evaluated. Again, the input parameters are derived based on the road topology, signal data and available public transport data.

The FD is not necessarily required to apply the SA. However, in order to compare the resulting MFD to the empirical and the MoC-based one, an FD is needed. The same values for the free flow speed u, the backward wave speed w, the jam density k_j and consequently the capacity q_{cap} are applied.

Furthermore, the block length is an input for the SA as well. The specific input parameter is the mean dimensionless block length $\lambda = 6.8$. The number of blocks in the studied road segment is $B = 5$. Theoretically, the SA allows to consider distributed green times and cycle lengths. The related input parameters are the average green time $\mu_g = 42$ s, and the long-run red to green ratio $\rho = 1.2$. Averaging the coefficient of variation of the green times, the red times and the block length results in an overall average coefficient of variation equals $\delta = 0.21$. No offset parameter is necessary as input for the SA.

The public transport extension of the SA requires additional input parameters. In the contrary to the MoC, the bus free flow speed is not assumed to account for potential stops. They are modelled explicitly via the stop probability p_s. Since there is a on-street bus stop only on one of the five segments, we choose $p_s = 1/5$. The average free flow bus speed u_b is chosen as 90% of the car free flow speed, see the explanation in Section 3.2.2. Again, the average dwell time is $t_d = 30$ s.

Based on these parameters the mean and coefficient of variation for the cuts related to three different strategies can be calculated (see Section 2). Assuming normally distributed cuts, the CDF for each cut is defined. This enables to compute the CDF and thus the percentiles of interest for the MFD for the herein studied arterial road.

3.3. Investigation of the measurement bias

A measurement bias inevitably exists in LDD, as the data are recorded at fixed positions. There are certain procedures to minimize this bias, as described in Section 2.2. For example, measurements from several fixed sensors on different positions could be projected to a virtual link. These projected and then accordingly weighted measurements could lead to a reduced bias in the data. However, this is infeasible for the case studied in this paper since all loop detectors are positioned at nearly the same distance from the stop-line. Another way to exclude measurement induced biases is to gain complete knowledge about the traffic situation, i.e. all trajectories. Since this is not possible for the empirical data of the Leopoldstraße, a microscopic simulation environment is built in SUMO. The layout of the simulation assembles the real arterial. In total, 129 OD pairs are defined and the route choice is designed based on recorded turning rates at each intersection. Additionally, two public transport bus lines each with a headway of ten minutes are represented in the simulation. The estimation of OD pairs and input demand curves for each pair is based on the measured travel times. More details regarding that are described in Grigoropoulos et al. (2018).

In order to investigate the measurement bias, a calibrated microscopic simulation is needed. This is achieved by adapting the Krauss car-following model (Krauß, 1998) parameters with the objective of minimizing the difference between the two MFDs from real loop detectors and virtual detectors from the simulation. For this purpose, flow and density measurements are recorded at positions that correspond to the loop detector positions at the real Leopoldstraße. Based on these virtual LDD measurements a locally estimated scatter plot smoothing (LOESS) regression (Cleveland, 1979) is applied. Additionally, a LOESS regression is applied to the empirical data. The comparison of both curves is the base for the calibration procedure which is based on RMSE minimization. This would not be possible without using regression techniques. Secondly, the calibrated simulation is further used to estimate the MFD based on full trajectory data following Edie's definitions. By doing so, the impact of measurement induced biases on the shape of the MFD for the present case of the Leopoldstraße can be inspected.

3.4. Sensitivity analysis of signal and public transport parameters

The sensitivity analysis is conducted to study the impact of important input parameters of the analytical approximations on the resulting MFDs. The MoC does not allow for implementing varying green phases and cycle lengths within an intersection. The SA assumes green phases and cycle lengths which are independent and identically distributed. Both assumptions are violated by using the signal data at hand. Furthermore, the Leopoldstraße is frequented by buses of the public transport operator. To estimate the impacts of multimodal flows on the MFD by applying the approximation methods, additional data on public transport operation are required. However, no reliable data for the specific input parameters are currently available. Consequently, realistic assumptions are made.

We conduct a Monte-Carlo-Simulation to study the violation of assumptions concerning the signal related input parameters, as well as to analyze the assumption regarding public transport related input parameters. In order to replicate the real-world scenarios in the solution space, all possible input parameter combinations have to be considered. However, this is infeasible due to the high computational effort. To overcome this challenge, we apply a Latin Hypercube sampling to generate a representative range of values for each input parameter. Several scenarios are studied. In each scenario a different number and type of input parameter are varied. The studied input parameters are related to signal control and public transport operation. The signal related input parameters are drawn from the distribution of measured values. For the MoC, this includes the green time and cycle length ratio g/c, as well as the offset o. For the SA, the parameters considered for sampling are the average green time μ_g, the long-run red to green ratio ρ and the corresponding standard deviations. In order to derive according mean and standard deviation values, small samples of five consecutive cycles are considered. The sample size of five is chosen as a trade-off between having a large enough number of mean and standard deviation values for the Monte-Carlo-Simulation and avoiding extreme values for each mean and standard deviation. For the

Table 3
Input parameter configuration for the sensitivity study.

	Method of cuts								Stochastic approximation					
Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14
g/c	✓			✓	✓		✓	✓			✓	✓		✓
o		✓		✓		✓	✓							
v_b			✓		✓	✓	✓							
u_b									✓		✓		✓	✓
t_d										✓		✓	✓	✓

sake of simplicity, these parameters are also labeled g/c further on in the paper as they essentially refer to green times and cycle lengths. The public transport related input parameters consist of the bus frequency, the average speed v_b of buses for the MoC, and the average bus free flow speed u_b and dwell times t_d for the SA. However, no reliable data for these specific input parameters are available. Thus, the bus frequency is assumed to follow the schedule and is therefore not further investigated within the sensitivity study. Moreover, v_b and u_b are varied between 50% and 100% of the car free flow speed u. In addition, t_d is varied between 10 s and 60 s. The interval for t_d is chosen based on a limited number of on-site measurements.

For each input parameter, a sample with a size of $n = 10000$ is drawn. Based on a set of input parameters, the MFDs can be approximated for each scenario. The effects of varying input parameters are studied for both analytical methods, the MoC and the SA. To give an overview of the conducted scenarios, Table 3 shows the investigated parameter combinations. As mentioned above, g/c represents the green phase and cycle length parameters for both the MoC and the SA. The input parameters o and v_b are only required for the MoC. The average bus free flow speed u_b and average dwell time t_d are only required for the SA.

4. Results and discussion

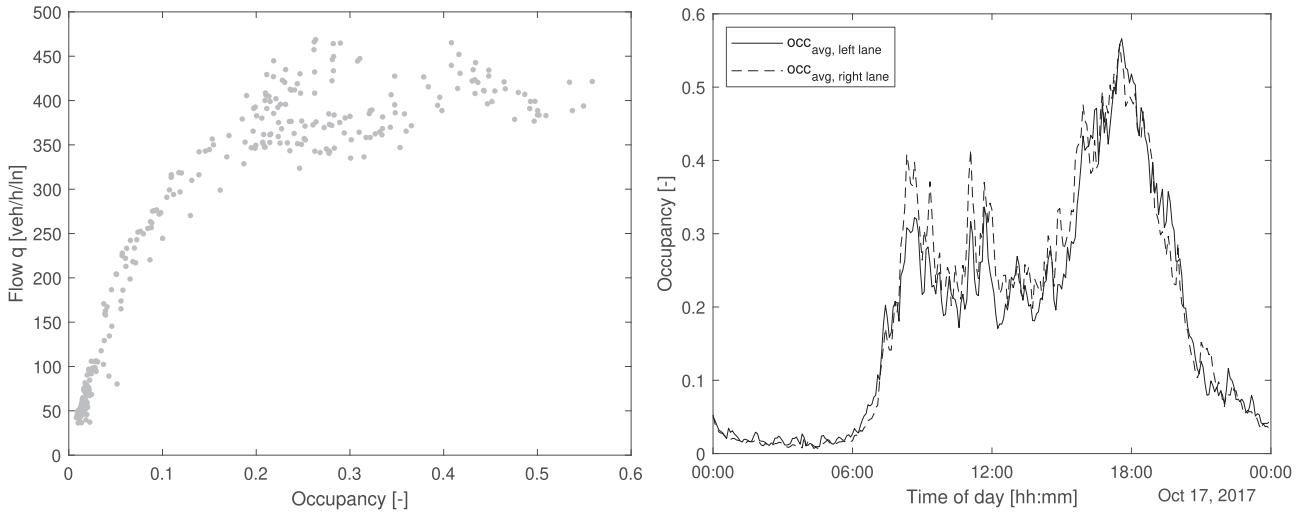
This section shows and discusses the results for the case study, the investigation of the measurement bias and the sensitivity analysis. Many results are presented as MFDs, such as in Fig. 4a, and Fig. 5a-8. If not stated otherwise, the abscissa shows the average density in veh/km/ln and the ordinate shows the average flow in veh/h/ln.

4.1. Case study

The results for the case study include the empirical MFD, and the MFDs approximated by the MoC the SA, respectively.

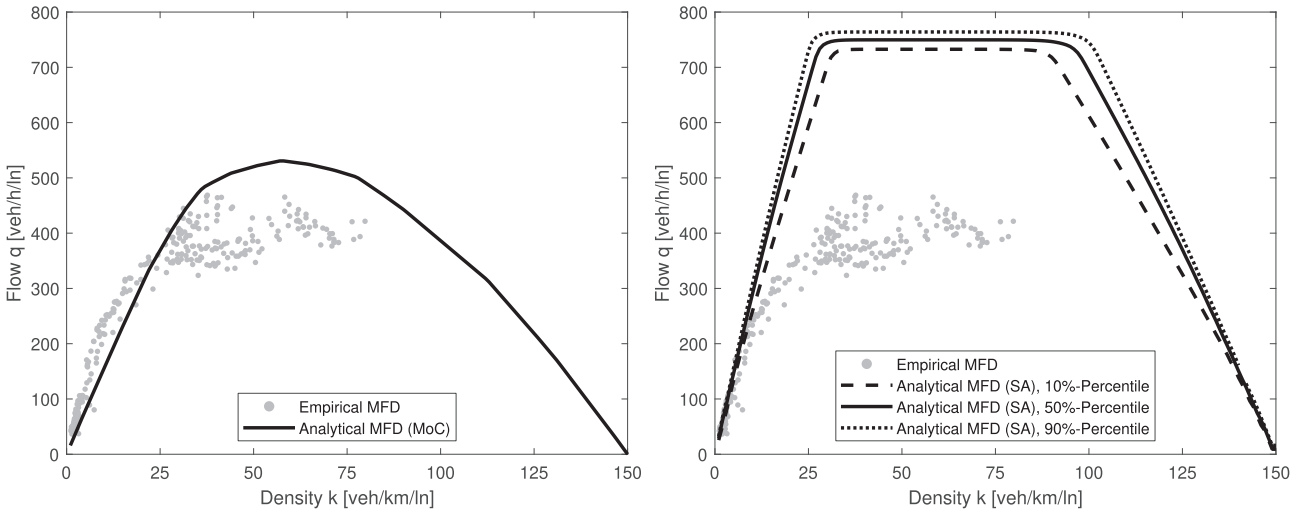
4.1.1. Empirical MFD

The empirical MFD is estimated based on the methods explained and the data described in the previous section. The result is shown in Fig. 4a. The abscissa shows the average dimensionless occupancy per lane which ranges from 0 to 1. The figure shows a well-defined MFD with a clear shape. Especially, the data related to free flow states are little scattered. Maximum flows are observed around 480 veh/h/ln. These observations are similar to MFDs reported for other cities (e.g. Ambühl et al., 2017). The occupancy reaches values up to 0.6. At occupancies between 0.2 to 0.4, a clockwise hysteresis loop can be observed. Its investigation reveals that the upper part of the loop occurred during the loading period, while the lower part occurred during the unloading period in the



a) Empirically estimated MFD. b) Average lane-specific occupancies.

Fig. 4. Empirical results for the Leopoldstraße on October 17, 2017.



a) The MFD based on the MoC. b) The MFD based on the SA.

Fig. 5. Comparison of analytical approximated and empirically derived MFDs.

evening. These periods correspond to the onset and offset of congestion. This pattern can be explained as follows. The direction of travel points towards the outer parts of the city, including the city's ring road. In other words, commuters are traveling along the Leopoldstraße to leave the city. This explains the time of the onset and offset of congestion. Similar patterns were observed in other studies (Saberi and Mahmassani, 2013; Buisson and Ladier, 2009). Reasons for this phenomenon are presented in Section 2.

The capacity of the arterial could be reduced due to constraining effects on the supply side (such as signals), but also due to demand-related matters. For example, if there were a strong favoring of one lane, the other lane's capacity might be underutilized and thus the MFD would not show the effective capacity for the whole road segment. To investigate this issue, we plot the lane specific average of occupancies for the analyzed road segment. The results are displayed in Fig. 4b. The abscissa and the ordinate show the time of day and the dimensionless occupancy per lane, respectively. The corresponding values for the left lane in direction of travel are shown as solid black curve, and as dashed black line for the right lane. The figure demonstrates that there are only minor differences between the two lanes. In general, occupancy values for the right lane are slightly higher than the ones for the left lane. Reasons for that might be the potentially higher frequency of flow-disturbing elements such as buses leaving a bus bay, cars looking for parking spots, and logistic vehicles while loading and unloading goods. However, the differences are minor and we thus conclude that the capacity of both lanes is utilized. Another reason for low flows apart from bottlenecks within the studied road segment can be constraining conditions upstream and downstream of the road segment. We suppose there are no upstream constraining conditions since we observe a congested branch in the empirical MFD. If there would be any constraining conditions upstream, the observed congested branch would be very limited. This is not the case, as can be seen in Fig. 4a. Additionally, traffic flow could be constrained by spillbacks from downstream intersections. Unfortunately, the available data does not include measurements from such locations. However, qualitative on-site evaluation from several days suggests that spillbacks are rare and do not systematically impact the capacity of the most downstream intersection of the studied road segment of the Leopoldstraße. Comparable results for the time series of occupancies are found for other weekdays at the studied road segment. Additionally, the distribution of green times and cycle lengths in a second available data set do not substantially differ to the one shown in the paper. Thus, we assume that analyzing the data presented in the paper is representative for the given road segment. In summary, we assume that the observed capacity in the MFD is due to bottlenecks within the investigated system and not due to underutilized lanes or constraining impacts from up- or downstream.

4.1.2. Analytical approximation

The evaluation of the analytical approximations for a given set of input parameters leads to an estimate for the upper bound of the MFD, shown in Fig. 5. The curves show the MFDs estimated based on the MoC and the SA. For comparison, the scatter plot is shown too, representing the empirically derived MFD with occupancies converted to densities as described in Section 3. The general trapezoidal shape results from the fact that the MoC and SA approximate the MFD based on different families of cuts, as explained in Section 2.

A qualitative comparison of the analytical MFD based on the MoC (see Fig. 5)a and the empirical MFD leads to the following observations. The analytically derived free flow cut matches the data rather well, even though the flows are in general underestimated for densities ranging up to 25 veh/km/ln. For densities greater than that, flows are overestimated by the MoC. Moreover, the maximum flow is reached around 470 veh/h/ln in reality. The estimated capacity of the MoC is around 530 veh/h/ln. Therefore, the MoC overestimates the capacity of the arterial by roughly 13%. Possible reasons for such differences between the empirically estimated and the analytically approximated MFD are the assumptions related to the KWT such as instantaneous acceleration and stationary traffic flow. Short green times and cycle lengths could lead to an enhancement of the effects of these assumptions. Furthermore, taking the average of green times and cycle lengths as input to the MoC might decrease its accuracy. Moreover, the MoC

does not allow for consideration of signal actuation or transit signal priority, which are in operation in reality and affect the traffic flow dynamics. Furthermore, the empirical MFD is based on LDD and thus includes a measurement bias. This bias is analyzed within this paper too. The respective results are presented in the next section. All these points potentially contribute to the observed discrepancies in the empirical estimated and analytical approximated MFD.

Please note that the SA computes a CDF for the MFD, i.e. a distribution of flows q for each density k. The results of the evaluation of the SA for the 10%-, the 50%- and the 90%-percentile are shown in Fig. 5b. Additionally, the figure shows the empirical data for comparison. The dashed, solid and dotted curves are the 10%-, the 50%- and the 90%-percentile of the CDF of the MFD, respectively. Generally, the free flow branch of the analytical approximations fit the empirical data well up to a flow of about 250 veh/h/ln. For higher flows, differences arise. The largest discrepancy occurs for the capacity which is about 300 veh/h/ln, or about 63% in relative terms. This discrepancy clearly exceeds the one observed for the MoC. Moreover, the difference between the percentiles of the analytical approximations are minor and cannot represent the scatter in the empirical data. The largest differences between them occur close to capacity and for the backward waves. We suppose that there exist several possible reasons for the large differences in the estimated and observed capacities. The actuated traffic signal control and transit signal priority violate the assumption of an independent distribution of green and red times. Moreover, there exist substantial differences in green-time-to-cycle-length ratios between intersections as can be concluded from Fig. 3. In other words, intersection No. 3 has the lowest g/c ratio and, thus, is determining the overall capacity of the arterial. However, the SA derives the capacity cut based on an average of capacities related to each intersection. In the present case of an arterial with substantial differences of capacities within the intersections, this might lead to a wrong estimation of the overall capacity. Lastly, it is again noteworthy to mention that the empirical data includes a measurement bias which impacts the general fit. All in all, these aspects could lead to the fact that the capacity based on stationary cuts is overestimated.

To quantify the comparison of the MoC and the SA to the empirical MFD, the corresponding RMSE are calculated. For the MoC an RMSE of 83.0, and for the SA an RMSE of 271.2 is calculated. These values indicate that the overall approximation based on the MoC is superior to one based on the SA for the input parameter configuration of the case study. As mentioned above, probable reasons are related to the measurement bias included in empirical data, the representation of signal control and public transport operation in the models, as well as assumptions inherited from the KWT. The results of a thorough analysis of these aspects are investigated in the next subsections.

4.2. Investigation of the measurement bias

Since only limited empirical data are available, the associated measurement bias can only be estimated. This is done based on a microscopic simulation, as described in Section 3. Fig. 6 shows the results of this investigation. Three different MFDs are presented. First, the MFD based on empirical LDD (ELDD) as presented in Fig. 4a is displayed again as grey scatter plot for comparison. Additionally, the curve resulting from a LOESS regression based on this data is shown as black solid line. Secondly, the curve resulting from a LOESS regression based on virtual loop detector measurements (VLDD) is shown as black dotted line. The curve suggests flows $q > 0$ veh/h/ln for densities $k = 0$ veh/km/ln. Please note that this is a result of the regression method, and physically impossible. Lastly, the LOESS regression curve based on trajectories from the simulation, virtual floating car data (VFCD), is presented as black dashed line.

The regression results based on ELDD and VLDD were used for calibration. The RMSE of the LOESS (ELDD) and LOESS (VLDD) is 22 veh/h/ln. Considering the impact of day-to-day variance in traffic flow, measurement errors and other factors we assume that this value is low enough for stating that the microscopic simulation is calibrated well enough for our purposes.

Since the trajectories contain all information, the curve based on virtual trajectories represents an estimate for an MFD without the measurement bias. It can be seen that the free flow branch is slightly shifted towards higher speeds and the capacity is increased

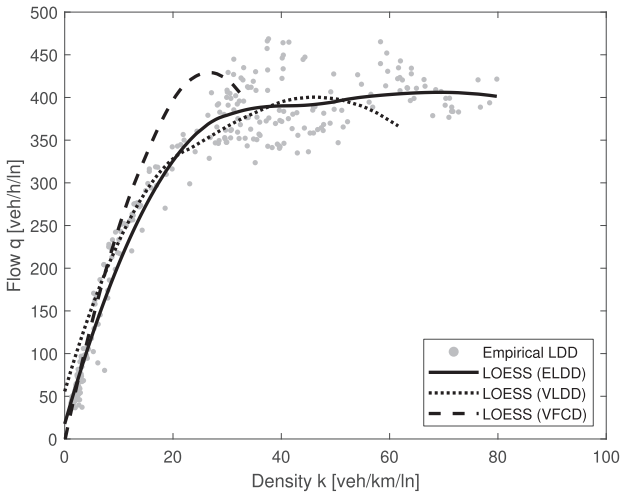
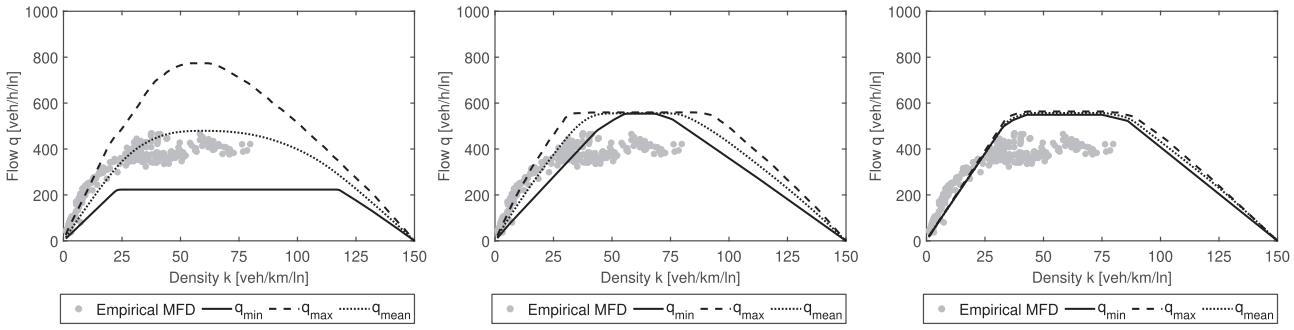


Fig. 6. Illustration of the measurement bias estimation.



a) Scenario 1: Sensitivity to g/c. b) Scenario 2: Sensitivity to o. c) Scenario 3: Sensitivity to v_b.

Fig. 7. Results of the sensitivity analysis of the MoC.

by roughly 40 veh/h/ln. This means that the measurement bias could lead to an underestimation of the actual measured capacity, and thus the error of the analytical approximation could be smaller as observed by comparison with the empirical LDD based MFD. Furthermore, the results indicate that the average free flow speed in the network is underestimated by simply relying on the empirical data.

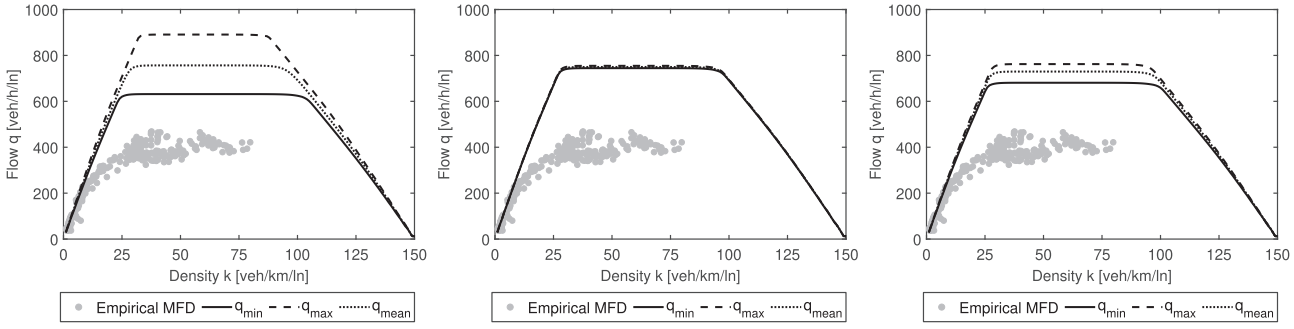
4.3. Sensitivity analysis of signal and public transport parameters

As discussed before based on the results from the case study, the representation of signal control schemes and public transport operation in the analytical approximations are potential reasons for the differences of the corresponding MFDs and the empirical MFD. In order to study the impact of the related input parameters, a Monte-Carlo-Simulation with $n = 10000$ draws is conducted as described in Section 3. Thus, for each density value a range of 10000 flow values is calculated. Based on these results, a corresponding minimum q_{min}, mean q_{mean} and maximum flow q_{max} can be derived. For the whole range of densities, these flows can be plotted as curves. Each curve is compared to the empirical MFD and an RMSE is calculated. The following subsections show and discuss the results for both investigated analytical approximations. The data plotted in Fig. 7 and Fig. 8 show the empirical MFD as scatter plot, and q_{min}, q_{mean} and q_{max} as dashed, solid and dotted curve, respectively.

4.3.1. Method of cuts

Three input parameters of the MoC are related to signal control and public transport operation. This is the green time and cycle length which is represented by g/c, the offsets o and the average bus speed v_b. To investigate all parameters and the possible combinations of them, seven scenarios are studied. Presenting all plots within this paper is not expedient since the generally observed patterns can be explained with the three diagrams shown in Fig. 7.

As expected, varying the parameter g/c related to cycle length and green times has a large impact on the shape of the resulting MFDs (see Fig. 7a). The capacity ranges between roughly 200 veh/h/ln and 800 veh/h/ln. Additionally, the free flow and congested branches of the MFD are varying strongly. Interestingly, the mean flows q_{mean} approximate the empirical MFD well. Fig. 7b shows the results of the sensitivity analysis of the input parameter related to the offset o. As expected, the impact on capacity is negligible. However, the free flow and congested branches are highly impacted by this input parameter. In general, the offset is challenging to specify, since it is adaptive in reality and this cannot be reproduced by the MoC. Lastly, Fig. 7c presents the results for the scenario related to the average bus speed v_b. It can be seen that the impact is minor. Thus, for the current representation of public transport in the MoC, this parameter plays only a minor role for the given case. Thus, assuming a corresponding value instead of measuring it is not crucial for the final result.



a) Scenario 8: Sensitivity to g/c. b) Scenario 9: Sensitivity to t_d. c) Scenario 10: Sensitivity to u_b.

Fig. 8. Results of the sensitivity analysis of the SA.

Table 4
RMSE results for the sensitivity study of the MoC.

Scenario	RMSE		
	q_{min}	q_{mean}	q_{max}
1	138.1	59.2	224.8
2	88.1	104.3	123.8
3	106.5	111.2	117.7
4	141.1	59.3	228.6
5	138.8	58.5	225.0
6	83.7	104.8	127.7
7	140.8	58.7	230.1

The figures for the other four scenarios (4 to 7) are not shown explicitly. Generally, visual inspection shows that the parameter g/c has the major influence also when it is combined with the other studied input parameters. The combination of o and v_b reveals that their impact on the MFD is still minor in comparison to g/c. Thus, the main shapes of the results are similar to the ones shown in Fig. 7. For the sake of comprehensiveness, the numerical results in the form of RMSE for all scenarios are presented in Table 4.

The results for the RMSE reveal further insights. First, it can be seen that the calculated RMSE for q_{mean} are the smallest except for scenario 2 and 6. q_{mean} for sceanrios 1, 4, 5 and 7 have a lower RMSE than the one calculated for the MFD of the case study. Thus, using q_{mean} would lead to an improved estimation of the empirical MFD for these scenarios. Another interesting result is that the RMSE of q_{mean} differ from the one related to the case study. In the case study, the MoC was applied with the average of each input parameter. Comparing the RMSE from resulting MFD from the case study to the RMSE of q_{mean} resulting from the sensitivity study reveals a difference. That is that the average input values do not lead to the mean output values. This finding confirms the purposefulness of the sensitivity study. Moreover, the ranges between the RMSE of q_{min} and q_{max} differ greatly. They are especially large when the parameter g/c is varied, such as in scenario 1, 4, 5 and 7. This lets us conclude that the impact of g/c is the largest on the analytical approximation by the MoC. This is in line with what was concluded from Fig. 7a. Lastly, the low impact of the average bus speed v_b is evident when the corresponding RMSE values are studied for scenario 3. Additionally, the values for scenario 5 are similar to the ones of scenario 1, so the interactive effect of v_b with g/c is low. The same applies to scenarios 2 and 6, where the interactive effect of v_b with o is examined.

In summary, it can be observed that the approximated capacity is highly sensitive to g/c. This is an expected result. However, averaging the flows for each density based on the computed MFD approximations reveals a well-fitting curve to the empirical MFD. This is interesting, since possible g/c values can be estimated based on signal control programs and thus an ex-ante prediction of the capacity of the arterial is feasible. The high sensitivity to g/c is due to the direct relation of this parameter to the stationary cut which affects the capacity for the present case. The choice of offsets o has a substantial impact on the free flow and backward wave branch but no effect on the capacity of the resulting MFDs. Reasons for that are varying offsets o do not affect the g/c ratio and thus not the capacity, as long as the stationary cuts are constraining. However, different offsets change the variational graph in such a way that the impact on shortest paths within the framework of the MoC leads to differing free flow and backward wave branches in the resulting MFDs. The fit of these branches to empirical data or potentially bias-free data is limited. However, this is not surprising since only fixed-time control can be implemented into the MoC, whereas in reality traffic signal actuation and transit signal priority are in operation. Additionally, it is clearly shown that the public transport operation as currently implemented in the methods has nearly no effect on the resulting MFD. This can be due to the low impact on the variational graph for the given headway and average bus speed.

4.3.2. Stochastic approximation

Similar to the MoC, three input parameters are related to signal control and public transport operation for the SA. Similar to the MoC, g/c reflects the green time and cycle length, the dwell times of buses t_d and the average bus free flow speed u_b. Again, seven different scenarios are studied to investigate all parameters and their possible combinations. Since the SA calculates a CDF of the MFD, a percentile has to be specified to derive a single MFD for each run within the sensitivity analysis. We choose the 50%-percentile as this seems to be the most representative.

Visual inspection of all plots shows again that it is sufficient to present only the diagrams of the first three scenarios in Fig. 8. The main results can be explained through these figures.

Fig. 8a shows that the parameter g/c related to cycle length and green times has again a large impact on the shape of the resulting MFDs. The capacity ranges roughly between 600 veh/h/ln and 900 veh/h/ln. Contrarily to the results of the MoC, the free flow and congested branch of the MFD are barely affected. Fig. 8b shows the results for varying the bus dwell times. It becomes clear that this input parameter has a negligible influence on the MFD for the investigated case. Thus, it seems that the choice of this parameter is not important for applying the SA for the case of the Leopoldstraße. Lastly, Fig. 8c presents the results for the scenario related to the average bus free flow speed u_b. The parameter has a substantial influence on the shape of the MFD. More specifically, it can be seen that the capacity ranges between 650 and 750 veh/h/ln. For all figures, it is observed that the free flow and backward wave regions are substantially less impacted by varying the input parameters compared to the MoC. Since the free flow cut matches the empirical data well, this is an advantage of the SA.

Again, the figures for the other four scenarios (10 to 14) are not shown explicitly. Visual inspection reveals that no unexpected

Table 5
RMSE results for the sensitivity study of the SA.

Scenario	RMSE		
	q_{min}	q_{mean}	q_{max}
8	267.5	271.2	274.8
9	187.2	275.7	367.5
10	221.6	256.8	280.2
11	204.3	256.6	293.7
12	148.9	261.0	369.7
13	174.6	275.5	372.0
14	133.7	260.9	369.7

interaction effects between the studied input parameters exist. The consideration of multiple variable input parameters results in adding their impacts. For example, t_d has still very limited effects when it is combined with g/c or u_b. The main shapes of the results are similar to the ones shown in Fig. 8. For the sake of comprehensiveness, the numerical results in the form of RMSE for all scenarios are presented in Table 5.

The results for the RMSE reveal further insights. First, it can be seen that the calculated RMSE for q_{min} are always the smallest. Additionally, all q_{min} have a lower RMSE than the one calculated for the MFD of the case study. Thus, using q_{min} would lead to an improved estimation of the empirical MFD for all scenarios. Another interesting result is that the RMSE of q_{mean} differ from the one related to the case study. Similarly to the results for the MoC, the mean input to the SA does not lead to the mean output. Moreover, the ranges between the RMSE of q_{min} and q_{max} differ greatly. They are especially large when the parameter g/c is varied, such as in scenario 8, 11, 12 and 14. This lets us conclude that the impact of g/c is the largest on the analytical approximation by the SA. This is in line with what was concluded from Fig. 8a and with similar results for the MoC. Lastly, the low impact of the bus dwell times t_d is evident when the corresponding RMSE values are studied for scenario 10. Additionally, the values for scenario 12 are similar to the ones of scenario 10, so the interactive effect of u_b with g/c is low. The same applies to scenarios 10 and 13, where the interactive effect of u_b with t_d is examined.

Generally, the SA shows a good and robust fit to the free flow branch. However, the capacity is clearly overestimated, even by taking the minimum of all computed MFDs in the corresponding scenarios. Reasons for such an overestimation are discussed in Section 4.1.2. Nevertheless, the method seems to capture public transport operation more effectively, since the related parameters have a substantial impact on the resulting MFDs. This could be due to the endogenous consideration of effective bus speeds in the model (Castrillon and Laval, 2017). Extending the SA to account for single intersections which dominate the overall capacity might be an interesting line of research. Such an extension could lead to a better fit to empirical data for cases similar to the Leopoldstraße, i.e. arterials with a comparably low number of intersections and high variability in signal timings.

This study aimed to investigate the differences between the empirically estimated and analytically approximated MFDs. The results shed light on possible reasons, such as the measurement bias, model assumptions and simplified consideration of signal control and public transport operation. Remaining unexplored reasons for discrepancies are likely the assumptions of the KWT, and the hysteresis phenomenon. For example, the role of infinite acceleration and deceleration implied by KWT could be investigated. First analyses conducted by the authors showed that more effort is necessary to understand the effect of these parameters on the MFD. Setting the acceleration and deceleration to extremely high values in the car-following model of the microscopic simulation does not produce the expected results. Under such conditions, some vehicles are not able to perform lane changes to reach their desired destination. As a result, they remain stopped at intersections. This causes unnecessary congestion, reduces the capacity of the arterial, and increases the differences between the simulation-based and analytical MFDs. Since a more detailed investigation of this phenomenon is beyond the scope of this paper, we suggest future research to study the role of infinite acceleration assumption of KWT. The hysteresis phenomenon can be considered with MFD-based modeling as shown in Mariotte et al. (2017). However, both the MoC and SA are methods to derive accumulation-based MFDs and thus such an investigation lies beyond the scope of this paper. Overall, the conducted analysis helps to gain a better understanding of several potential error sources.

5. Conclusion

This paper compares two different methods which analytically approximate the arterial MFD. The first method is the MoC which is based on the KWT. It is deterministic and approximates the MFD based on the link FD, road topology, signal settings and public transport operation parameters. The second method is the SA. It considers the stochasticity of input parameters and estimates a CDF for the MFD. Within a case study, both methods are evaluated by comparing the resulted MFDs against the MFD derived from empirical data collected at five intersections of an urban arterial, namely the Leopoldstraße in Munich, Germany. The empirical data set consists mainly of LDD and time-stamps for signal phase changes. The case study reveals differences between the empirical and analytical MFDs. Subsequently, the paper investigates potential reasons for the occurring discrepancies between the empirical estimation of the MFD and corresponding analytical approximations. This includes the analysis of the measurement bias which is induced by LDD. The results suggest that the actual capacity could be larger as the empirical MFD proposes for the studied case. Moreover, the free flow branch of the estimated MFD appears to be steeper without the bias. Eventually, an in-depth investigation of

the input parameters regarding signal control and public transport operation for both methods and their impacts on the resulting MFD approximation is conducted. The analysis is performed as Monte-Carlo-Simulation based on a Latin Hypercube sampling. The results indicate that the capacity predicted by both methods is highly sensitive to the green-to-cycle ratio related input parameters. Interestingly, the mean flow resulting from calculating a large sample of MFDs based on the MoC by varying the green-to-cycle ratio fits the empirical data very well. Moreover, the offsets have a strong impact on the free flow branch of MFDs predicted by the MoC, which is generally underestimated by this method. Contrarily, the SA overestimates the capacity for all cases. However, this method predicts the free flow branch accurately. Input parameters related to public transport operation have a comparably minor impact. Overall, the MoC seems to be more promising for the case studied within this paper as indicated by the corresponding RMSE results. By knowing possible green-to-cycle ratios, a general shape of the MFD seems to be predictable for highly inhomogeneous urban arterials. This opens the door to model-predictive traffic control based on the MFD.

The contributions of this paper include an in-depth analysis of the role of signal data for analytical MFD estimation methods. Moreover, we compare the performance of two prominent analytical MFD estimation methods to empirical data from an arterial in Munich, Germany. The results indicate that the MoC can be applied to estimate the capacity for highly inhomogeneous arterials, whereas the SA delivers a good fit for the free-flow branch.

Future research regards to investigate more urban arterials to generalize the findings for the MoC and the SA. Furthermore, extending the MoC to be applicable to road networks and arterials with traffic actuated signal control would be an interesting task. This could lead to a fast and efficient way to analyze the capacity of an arterial with actuated signal control. Thus, it could be used to study different signal control strategies regarding maximum arterial capacity. However, it is unclear if the mathematical framework holds for hyperlinks which represent varying signal timings between intersections and at each intersection to represent more complex signal control schemes. Thus, completely new approaches might be necessary. Nevertheless, more accurate implementation of public transport can be incorporated. This might already improve the accuracy of the methods. Moreover, only the SA applies to networks. No deterministic method such as the MoC is available to analytically estimate the MFD for networks in general. The MoC was only applied to simplified networks, which can be approximated by a single arterial. Therefore, corresponding developments could open the door to network-wide deterministic ex-ante MFD prediction.

Acknowledgments

The authors would like to express their gratitude to Prof. Ludovic Leclercq and Prof. Jorge Laval for providing the MATLAB® and Mathematica® codes for their methods. Furthermore, the authors would like to thank the KVR of Munich for providing the data this study is based on. Lastly, we would like to thank the anonymous reviewers for their valuable comments which helped to substantially improve the paper.

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. Transport. Res. Part B: Methodol. 55, 265–281.
Ambühl, L., Loder, A., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2018. A functional form with a physical meaning for the macroscopic fundamental diagram. Transport. Res. Part B: Methodol..
Ambühl, L., Loder, A., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2018. Introducing a resampling methodology for the estimation of empirical macroscopic fundamental diagrams. Transport. Res. Rec.: J. Transport. Res. Board 2676, 239–248.
Ambühl, L., Loder, A., Menendez, M., Axhausen, K.W., 2017. Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data. In: Presented at the 96th Annual Meeting of the Transportation Research Board.
Ambühl, L., Loder, A., Zheng, N., Axhausen, K.W., Menendez, M., 2019. Approximative network partitioning for mfds from stationary sensor data. Transport. Res. Rec.: J. Transport. Res. Board 2673, 94–103.
Ambühl, L., Menendez, M., 2016. Data fusion algorithm for macroscopic fundamental diagram estimation. Transport. Res. Part C: Emerg. Technol. 71, 184–197.
Amini, S., Tilg, G., Busch, F., 2018. Evaluating the impact of real-time traffic control measures on the resilience of urban road networks. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 519–524.
Arnet, K., Guler, S.I., Menendez, M., 2015. Effects of multimodal operations on urban roadways. Transport. Res. Rec.: J. Transport. Res. Board 2533, 1–7.
Batista, S., Leclercq, L., Geroliminis, N., 2019. Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. Transport. Res. Part B: Methodol. 122, 192–217.
Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transp. Res. Rec. 2124, 127–136.
Cao, J., Menendez, M., 2015. System dynamics of urban traffic based on its parking-related-states. Transport. Res. Part B: Methodol. 81, 718–736.
Castrillon, F., 2015. Theoretical analysis of the effects of bus operations on urban corridors and networks. Ph.D. thesis. Institute of Technology, Georgia.
Castrillon, F., Laval, J., 2017. Impact of buses on the macroscopic fundamental diagram of homogeneous arterial corridors. Transport. B: Transport Dyn. 6, 286–301.
Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74, 829–836.
Courbon, T., Leclercq, L., 2011. Cross-comparison of macroscopic fundamental diagram estimation methods. Procedia-Social Behav. Sci. 20, 417–426.
Daganzo, C.F., 2005a. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transport. Res. Part B: Methodol. 39, 187–196.
Daganzo, C.F., 2005b. A variational formulation of kinematic waves: Solution methods. Transport. Res. Part B: Methodol. 39, 934–950.
Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transport. Res. Part B: Methodol. 41, 49–62.
Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transport. Res. Part B: Methodol. 45, 278–288.
Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transport. Res. Part B: Methodol. 771–781.
Daganzo, C.F., Lehe, L.J., 2016. Traffic flow on signalized streets. Transport. Res. Part B: Methodol. 90, 56–69.
Daganzo, C.F., Lehe, L.J., Argote-Cabanero, J., 2018. Adaptive offsets for signalized streets. Transport. Res. Part B: Methodol. 117, 926–934.
Daganzo, C.F., Menendez, M., 2005. A variational formulation of kinematic waves: Bottleneck properties and examples. In: Proceedings of the 16th International Symposium on Transportation and Traffic Theory.
Dakic, I., Menendez, M., 2018. On the use of lagrangian observations from public transport and probe vehicles to estimate car space-mean speeds in bi-modal urban networks. Transport. Res. Part C: Emerg. Technol. 91, 317–334.

Department for Transport, 2015. Free flow vehicle speeds statistics: Great Britain 2014. Technical Report October. Department for Transport, UK.
Du, J., Rakha, H., Gayah, V.V., 2016. Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. Transport. Res. Part C: Emerg. Technol. 66, 136–149.
Duruisseau, C., Leclercq, L., 2018. A global sensitivity analysis of dynamic loading and route selection parameters on network performances. J. Adv. Transport. 2018.
Edie, L., 1963. Discussion of traffic stream measurements and definitions. In: Almond, J. (Ed.), Proceedings of the 2nd International Symposium on the Theory of Traffic Flow, OECD, Paris, France, pp. 139–154.
Gayah, V., Daganzo, C.F., 2011. Clockwise hysteresis loops in the Macroscopic Fundamental Diagram: An effect of network instability. Transport. Res. Part B: Methodol. 45, 643–655.
Gayah, V.V., Gao, X., Nagle, A.S., 2014. On the impacts of locally adaptive signal control on urban network stability and the Macroscopic Fundamental Diagram. Transport. Res. Part B: Methodol. 70, 255–268.
Geroliminis, N., Boyacı, B., 2012. The effect of variability of urban systems characteristics in the network capacity. Transport. Res. Part B: Methodol. 46, 1607–1623.
Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transport. Res. Part B: Methodol. 42, 759–770.
Geroliminis, N., Haddad, J., Ramezani, M., 2012. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. IEEE Trans. Intell. Transp. Syst. 14, 348–359.
Geroliminis, N., Sun, J., 2011. Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. Transport. Res. Part A: Policy Practice 45, 966–979.
Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transport. Res. Part C: Emerg. Technol. 42, 168–181.
Girault, J.T., Gayah, V.V., Guler, I., Menendez, M., 2016. Exploratory analysis of signal coordination impacts on macroscopic fundamental diagram. Transport. Res. Rec.: J. Transport. Res. Board 36–46.
Godfrey, J., 1969. The mechanism of a road network. Traffic Eng. Control 8.
Grigoropoulos, G., Keler, A., Kaths, J., Kaths, H., Spangler, M., Hoffmann, S., Busch, F., 2018. Evaluation of the traffic efficiency of bicycle highways: A microscopic traffic simulation study. In: Presented at the 7th Symposium of the European Association for Research in Transportation.
Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204, 148–151.
Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S., Qian, X., 2010. Investigating the shape of the macroscopic fundamental diagram using simulation data. Transport. Res. Rec.: J. Transport. Res. Board 2161, 40–48.
Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., Papageorgiou, M., 2012. Exploiting the fundamental diagram of urban networks for feedback-based gating. Transport. Res. Part B: Methodol. 46, 1393–1403.
Knoop, V.L., De Jong, D., Hoogendoom, S.P., 2014a. Influence of road layout on network fundamental diagram. Transp. Res. Rec. 2421, 22–30.
Knoop, V.L., de Jong, D., Hoogendoorn, S.P., 2014b. The influence of the road layout on the network fundamental diagram. Transport. Res. Rec.: J. Transport. Res. Board 2421, 22–30.
Knoop, V.L., Van Lint, H., Hoogendoorn, S.P., 2015. Traffic dynamics: Its impact on the macroscopic fundamental diagram. Physica A 438, 236–250.
Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. Transport. Res. Part B: Methodol. 96, 26–45.
Krauß, S., 1998. Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics. Ph.D. thesis.
Laval, J.A., Castrillón, F., 2015. Stochastic approximations for the macroscopic fundamental diagram of urban networks. Transport. Res. Part B: Methodol. 81, 904–916.
Leclercq, L., 2005. Calibration of flow–density relationships on urban streets. Transport. Res. Rec.: J. Transport. Res. Board 1934, 226–234.
Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. Transport. Res. Part B: Methodol. 62, 1–12.
Leclercq, L., Geroliminis, N., 2013. Estimating mfds in simple networks with route choice. Transport. Res. Part B Methodol. 57, 468–484.
Leclercq, L., Sénécat, A., Mariotte, G., 2017. Dynamic macroscopic simulation of on-street parking search: A trip-based approach. Transport. Res. Part B: Methodol. 101, 268–282.
Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves ii. a theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. A 229, 317–345.
Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2017. Empirics of multi-modal traffic networks – using the 3d macroscopic fundamental diagram. Transport. Res. Part C: Emerg. Technol. 82, 88–101.
Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2019. Understanding traffic capacity of urban networks. Sci. Rep. 9, 1–10.
Lopez, C., Leclercq, L., Krishnakumari, P., Chiabaut, N., van Lint, H., 2017. Revealing the day-to-day regularity of urban congestion patterns with 3d speed maps. Sci. Rep. 7, 1–11.
Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wießner, E., 2018. Microscopic traffic simulation using sumo. In: The 21st IEEE International Conference on Intelligent Transportation Systems. IEEE.
Mahmassani, H., Williams, J.C., Herman, R., 1987. Performance of urban traffic networks. In: Proceedings of the 10th International Symposium on Transportation and Traffic Theory, pp. 1–20.
Mariotte, G., Leclercq, L., 2019. Flow exchanges in multi-reservoir systems with spillbacks. Transport. Res. Part B: Methodol. 122, 327–349.
Mariotte, G., Leclercq, L., Laval, J.A., 2017. Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. Transport. Res. Part B: Methodol. 101, 245–267.
Mazloumian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. Philosoph. Trans. Roy. Soc. A: Mathe., Phys. Eng. Sci. 368, 4627–4647.
Mühlich, N., Gayah, V.V., Menendez, M., 2014. An examination of mfd hysteresis patterns for hierarchical urban street networks using micro-simulation. In: Presented at the 94th Annual Meeting of the Transportation Research Board.
Nagle, A., Gayah, V., 2014. Accuracy of networkwide traffic states estimated from mobile probe data. Transport. Res. Rec.: J. Transport. Res. Board 2421, 1–11.
OpenStreetMap contributors, 2019. Munich dump retrieved from https://planet.osm.org.
Ortigosa, J., Gayah, V.V., Menendez, M., 2017. Analysis of one-way and two-way street configurations on urban grid networks. Transport. B: Transport Dyn. 1270, 1–21.
Richards, P.I., 1956. Shock waves on the highway. Oper. Res. 4, 42–51.
Saberi, M., Mahmassani, H., 2013. Hysteresis and capacity drop phenomena in freeway networks empirical characterization and interpretation. Transport. Res. Rec.: J. Transport. Res. Board 2391, 44–55.
Saberi, M., Mahmassani, H., Hou, T., Zockaie, A., 2014. Estimating network fundamental diagram using three-dimensional vehicle trajectories. Transport. Res. Rec.: J. Transport. Res. Board 2422, 12–20.
Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on “snake” similarities. Transport. Res. Part B: Methodol. 91, 250–269.
Saeedmanesh, M., Geroliminis, N., 2017. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. Transport. Res. Part B: Methodol. 105, 193–211.
Shim, J., Yeo, J., Lee, S., Hamdar, S.H., Jang, K., 2019. Empirical evaluation of influential factors on bifurcation in macroscopic fundamental diagrams. Transport. Res. Part C: Emerg. Technol. 102, 509–520.
Sirmatel, I.I., Geroliminis, N., 2018. Mixed logical dynamical modeling and hybrid model predictive control of public transport operations. Transport. Res. Part B: Methodol. 114, 325–345.
Tilg, G., Amini, S., Busch, F., 2019. Arterial macroscopic fundamental diagram: A comparison of analytical approximations and empirical data from munich. In: Presented at the 98th Annual Meeting of the Transportation Research Board.

Tsubota, T., Bhaskar, A., Chung, E., 2014. Macroscopic fundamental diagram for brisbane, australia. Transport. Res. Rec.: J. Transport. Res. Board 2421, 12–21.
Wu, X., Liu, H.X., Geroliminis, N., 2011. An empirical analysis on the arterial fundamental diagram. Transport. Res. Part B: Methodol. 45, 255–266.
Xie, X., Chiabaut, N., Leclercq, L., 2013. Macroscopic fundamental diagram for urban streets and mixed traffic: Cross comparison of estimation methods. Transport. Res. Rec.: J. Transport. Res. Board 2390, 1–10.
Yang, K., Menendez, M., Zheng, N., 2019. Heterogeneity aware urban traffic control in a connected vehicle environment: A joint framework for congestion pricing and perimeter control. Transport. Res. Part C: Emerg. Technol. 105, 439–455.
Yang, K., Zheng, N., Menendez, M., 2018. Multi-scale perimeter control approach in a connected-vehicle environment. Transport. Res. Part C: Emerg. Technol. 94, 32–49 ISTTT22.
Yildirimoglu, M., Sirmatel, I.I., Geroliminis, N., 2018. Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. Transport. Res. Part B: Methodol. 118, 106–123.
Zheng, N., Rérat, G., Geroliminis, N., 2016. Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. Transport. Res. Part C: Emerg. Technol. 62, 133–148.
Zhong, R., Chen, C., Huang, Y., Sumalee, A., Lam, W., Xu, D., 2018. Robust perimeter control for two urban regions with macroscopic fundamental diagrams: A control-lyapunov function approach. Transport. Res. Part B: Methodol. 117, 687–707 TRB:ISTTT-22.
Zockaie, A., Saberi, M., Saedi, R., 2018. A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. Transport. Res. Part C: Emerg. Technol. 86, 245–262.

Transportation Research Part B 150 (2021) 435–456

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

On the application of variational theory to urban networks

Gabriel Tilg [a,*], Lukas Ambühl [b], Sergio Batista [c], Monica Menendez [c], Fritz Busch [a]

[a] Chair of Traffic Engineering and Control, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Germany
[b] Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zurich, Switzerland
[c] Division of Engineering, New York University Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
Variational theory
Network modelling
Kinematic wave theory
LWR model
Traffic flow theory

ABSTRACT

The well-known Lighthill–Whitham–Richards (LWR) theory is the fundamental pillar for most macroscopic traffic models. In the past, many methods were developed to numerically derive solutions for LWR problems. Examples for such numerical solution schemes are the cell transmission model, the link transmission model, and the variational theory (VT) of traffic flow. So far, the eulerian formulation of VT found applications in the fields of traffic modelling, macroscopic fundamental diagram estimation, multi-modal traffic analyses, and data fusion. However, these studies apply VT only at the link or corridor level. To the best of our knowledge, there is no methodology yet to apply VT at the network level. We address this gap by developing a VT-based framework applicable to networks. Our model allows us to account for source terms (e.g. inflows and outflows at intersections) and the propagation of spillbacks between adjacent corridors consistent with kinematic wave theory (KWT). We show that the trajectories extracted from a microscopic simulation fit the predicted traffic states from our model for a simple intersection with both source terms and spillbacks. We also use this simple example to illustrate the accuracy of the proposed model, and the ability to model complex bottlenecks. Additionally, we apply our model to the Sioux Falls network and again compare the results to those from a microscopic KWT simulation. Our results indicate a close fit of traffic states, but with substantially lower computational cost. The developed methodology is useful for extending existing VT applications to the network level, for network-wide traffic state estimations in real-time, or other applications within a model-based optimization framework.

1. Introduction

The Lighthill–Whitham–Richards (LWR) theory (Lighthill and Whitham, 1955; Richards, 1956) is the base for most macroscopic models of traffic dynamics at the link level. The first-order model relates the conservation law (see Eq. (1)) and the fundamental diagram (FD) $q = Q(k)$ of traffic:

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \tag{1}$$

where q is the flow in vehicles per hour, k is the density in vehicles per kilometre, and t, x represent the time and space coordinates, respectively. Newell (1993) formulated the LWR theory in terms of the cumulative vehicle count $N(x,t)$ in his simplified kinematic wave theory (KWT). It describes the number of vehicles N that have passed point x by time t. The cumulative count N across all

* Corresponding author.
E-mail addresses: gabriel.tilg@tum.de (G. Tilg), lukas.ambuehl@ivt.baug.ethz.ch (L. Ambühl), sergio.batista@nyu.edu (S. Batista), monica.menendez@nyu.edu (M. Menendez), fritz.busch@tum.de (F. Busch).

https://doi.org/10.1016/j.trb.2021.06.019
Received 10 June 2020; Received in revised form 13 April 2021; Accepted 30 June 2021

(x,t) forms a surface which is also known as the Moskowitz function (Moskowitz, 1965). The derivatives of this surface at a given point (x,t) are the flow $q(x,t)$ and the density $k(x,t)$. Newell (1993) formulated the problem as follows:

$$\frac{\partial N}{\partial t} - Q(-\frac{\partial N}{\partial x}) = 0. \tag{2}$$

The formulation of the LWR in terms of $N(x,t)$ corresponds to a partial differential equation of the Hamilton–Jacobi type. Over the past decades numerous approaches have been developed to numerically solve the LWR partial differential equation. Generally, the equation can be represented in three different coordinate systems (Laval and Leclercq, 2013): eulerian (t,x), time-lagrangian (t,N), and space-lagrangian (N,x). In the eulerian representation, popular methods are the cell transmission model (CTM) (Daganzo, 1992, 1995), the link transmission model (LTM) (Yperman, 2007; Tampère et al., 2011; Han et al., 2015; Jin, 2015), techniques from partial differential equations theory such as the grid-free method by Mazaré et al. (2011), and the variational theory (VT) of traffic flow (Daganzo, 2005a,b; Daganzo and Menendez, 2005; Claudel and Bayen, 2010a,b). Several works have also explored the time- and space-lagrangian coordinate systems for solving LWR and KWT partial differential equations based on the VT approach (e.g. Mahut et al., 2003; Leclercq et al., 2007; Leclercq and Becarie, 2012; Duret and Yuan, 2017). For a more detailed overview on numerical schemes to solve LWR and KWT problems, we refer the reader to Seo et al. (2017).

The selection of a KWT solution method could be case-specific. For example, the CTM is able to model network traffic including multiclass traffic flow, but is known to exhibit numerical viscosity in which shock waves are inaccurately represented as smooth variations. These numerical errors converge to zero with decreasing time-step sizes. However, this increases the computational effort accordingly. The LTM is numerically precise, and efficient for large-scale networks. Yet, it does not allow to explicitly consider complex heterogeneous KWT problems. Such problems include space–time dependent FDs, in combination with any type and number of stationary and/or moving bottlenecks within the links. Solutions based on VT have been explored for time- and space-lagrangian, as well as eulerian coordinates. They are exact for triangular FDs. Models based on the lagrangian perspective apply to networks and multiclass traffic flow. However, the consideration of continuous inflows and outflows within links, i.e. source terms, is not possible yet (Laval et al., 2016). Consequently, phenomena such as lane changes cannot be readily modelled (Leclercq et al., 2007). In contrary, such source terms can be considered in eulerian VT. However, this VT formulation only applies to corridors where flow conservation holds. On the other hand, it permits the evaluation of KWT problems including intra-link heterogeneities both in time and space, such as changing FDs or any type and number of bottlenecks (Daganzo, 2005b; Daganzo and Menendez, 2005; Mazaré et al., 2011).

The literature related to eulerian VT reveals its wide range of applications apart from pure traffic modelling. This includes multi-modal traffic analysis (Dakic et al., 2020; Guler et al., 2016; Wu et al., 2017; Wu and Guler, 2018; Gayah et al., 2016; Saeednia and Menendez, 2016), MFD estimation (Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013; Laval and Castrillón, 2015; Tilg et al., 2020a), and traffic data imputation (Mehran et al., 2012; Sun and Ban, 2013). Additionally, extensions for consideration of higher order traffic models (Lebacque and Khoshyaran, 2013; Costeseque and Lebacque, 2014), stochasticity (Wada et al., 2018; Dakic et al., 2020), and continuous source terms (Laval et al., 2016) exist. The wide range of applications, the existing framework extensions, as well as the ability to consider complex intra-link heterogeneities makes the eulerian VT a valuable framework to numerically solve KWT problems. As a result, extending its applicability from single corridors to networks benefits all these applications, and allows us to evaluate the impact of complex intra-link heterogeneities at the network level with numerical precision and efficiency. To the best of our knowledge, none of the previous studies have started to explore the application of eulerian VT at the network level accounting for its complexities. To address this gap, we propose an extension of eulerian VT that permits to model the traffic dynamics at the network level. The contributions of this paper are fourfold:

1. We include inflows and outflows in the mathematical framework of eulerian VT. Contrary to Laval et al. (2016), these source terms accept any values from the set of real numbers, most importantly, they also include zero.
2. We model the propagation of spillbacks across the network. This allows us to track the evolution of network-wide congestion patterns.
3. We evaluate the proposed methodology and compare it with a microscopic KWT model implemented in SUMO (Lopez et al., 2018). This enables us to show a proof of our concept, and to demonstrate the inheritance of useful properties from the original VT, including the numerical precision and the ability to model complex bottlenecks.
4. We propose an algorithm to apply our model to any road network, and showcase its applicability using the well-known example of the Sioux Falls network.

In the remainder of this paper, 'VT' refers to the eulerian system of coordinates. This paper is organized as follows. Section 2 provides a brief background on the theoretical concept of VT, its applications and methodological extensions. Section 3 proposes a VT-based methodology to estimate traffic state propagation throughout a network. This includes the consideration of inflows and outflows at intersections and the correct propagation of spillbacks. Section 4 provides a proof of concept. Section 5 demonstrates, with the aid of small networks, that the proposed model inherits the original VT properties, which have proven to be rather useful for a wide range of applications. Section 6 presents the applicability of our methodology to larger networks. Section 7 critically assesses the developed framework including its limitations and capabilities. Finally, Section 8 highlights the conclusion of this study and outlines potential future research topics.

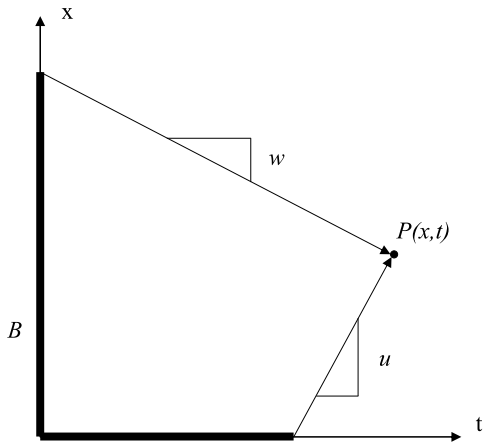


Fig. 1. Concept of VT.

2. Background on VT

2.1. Theoretical concept

Daganzo (2005a) formulated the VT framework the first time to determine the cumulative number of vehicles, i.e. the Moskowitz function, $N(x,t)$ which have passed the location x of a road by time t for given boundary conditions. The surface formed by N across all (x,t) is continuous but not differentiable at shock waves under the assumption of flow conservation. Fig. 1 illustrates the basic concept of VT.

The FD, i.e. $Q(k,x,t)$, is a necessary input for solving a KWT problem. Such an FD is required to be concave and piecewise differentiable by VT. However, the VT framework allows to model inhomogeneous conditions with FDs varying across time t and space x (Daganzo, 2005b). Additionally, boundary data N_B along a curve B are required, as illustrated by a thick black line in Fig. 1. Let us now discuss, how to determine N at point $P(x,t)$. We define a set of valid paths $p \in \mathcal{P}$, that start at the boundary B and end at P. A path p is valid if its slope ranges between the extremal speeds $v \in [w,u]$, where w is the backward wave speed which has a negative value, and u is the free-flow speed. Related to the slope, each path p has a certain cost c_p, corresponding to the maximum traffic rate that can pass a moving observer travelling along that path. Given the set $\mathcal{P}$ with costs c_p for all paths $p \in \mathcal{P}$, and the boundary value $N_{B,p}$ associated to the path p, the cumulative count N_P is found as follows:

$$N_P = \min_{p\in\mathcal{P}}\{N_{B,p} + c_p\}. \tag{3}$$

Assuming a triangular FD further simplifies the procedure. It is characterized by u, w, and the jam density κ. The capacity is derived at the critical density, i.e. $Q(k_c)$. In this case, the formulation becomes exact (Daganzo, 2005a). We also define $\theta = |\frac{u}{w}|$ as the absolute value of the ratio of the free-flow and the backward wave speed. If θ is an integer, the so-called lopsided variational graph can be constructed (see Daganzo and Menendez, 2005), where nodes in the graph align horizontally. This graph represents a discretized grid, where Δt and $\Delta x = u\Delta t$ are the temporal and spatial grid lengths, respectively. Note that θ is not required to be integer in general. Depending on it being an integer or not, different types of variational graphs can be constructed as explained in Daganzo and Menendez (2005). We denote the variational graph as $G(\mathcal{C},\mathcal{I})$, where $\mathcal{C}$ is the set of corridors, and $\mathcal{I}$ is the set of intersections. For the original VT formulation, $|\mathcal{C}| = 1$. The case of a multi-dimensional graph will be explained in the following sections. One can then solve a KWT problem to determine N on each node of G, by conducting a shortest-path search starting at the boundary. In the absence of moving bottlenecks, the necessary paths to consider become straight lines and have either extremal slopes, or slopes equal to zero at signals, i.e. $v \in \{w,0,u\}$. We can find $N(x,t)$ for this case as the minimum of the shortest paths coming from three nodes, one upstream node at $x-\Delta x$, one downstream node at $x+\Delta x$, and one node at the same location x (Daganzo and Menendez, 2005; Leclercq and Paipuri, 2019):

$$\begin{aligned} N(x,t) = \min(&N(x-\Delta x, t-\Delta t),\\ &N(x+\Delta x, t-\theta\Delta t) + \Delta x\kappa,\\ &N(x, t-\Delta t) + \beta). \end{aligned} \tag{4}$$

The first term in Eq. (4) refers to free-flow states travelling from upstream, the second one to congested states travelling from downstream, and the third term accounts for bottleneck induced capacity constraints at the same location as the target node. For free-flow states the cost of the edge, i.e. the maximum rate at which traffic can pass a moving observer travelling at a speed corresponding to the slope of that edge, is zero. For congested states the cost of the edge is $\Delta x\kappa$. For capacity constraints at the same location the cost of the edge is β. If a signal at position x is red during the interval $[t-\Delta t, t]$, $\beta = 0$. Otherwise, $\beta = Q(k_c)$, i.e. the link capacity. This formulation allows to apply VT to any signalized or unsignalized corridor with a triangular FD, where θ is integer.

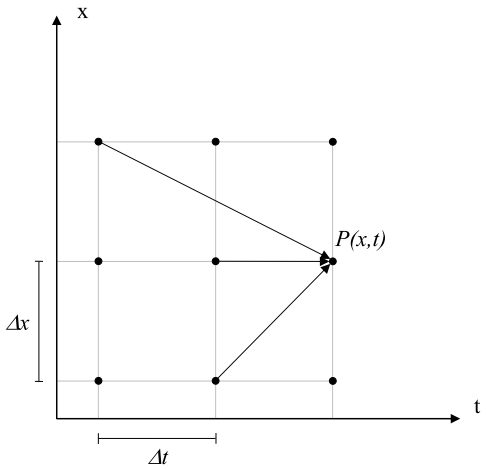


Fig. 2. Excerpt of the variational graph.

Fig. 2 represents an excerpt of the variational graph G which allows for a graphical interpretation of Eq. (4). The graph G is represented as a numerical grid with Δx in the spatial and Δt in the temporal dimension. For this figure, we set $\theta = 2$. Additionally, the figure shows the three "from nodes": upstream, at the same location, and downstream of $P(x,t)$, as well as the corresponding paths along which the costs are calculated.

VT is a numerical solution method for KWT problems which enables the consideration of varying concave and piecewise differentiable FDs over time and space. While any type of bottleneck can be implemented (Daganzo and Menendez, 2005), its trajectory needs to be fully mapped in G in order to be considered. For example, nodes in the variational graph G for single corridors and triangular FDs with non-integer θ do not align horizontally. Thus, additional manual efforts are required to include traffic signals. While this poses a limitation from a practical perspective, the general application of VT is still feasible for such cases. Another limitation is that no continuous multi-modal, multi-commodity traffic streams can be modelled. However, hybrid approaches to model multi-modal aspects were successfully developed in previous studies where slower modes (e.g. trucks or buses) are represented as moving bottlenecks (Dakic et al., 2020; Saeednia and Menendez, 2016). The main limitation of VT so far is that it only applies to single corridors, i.e. where the conservation equation holds.

2.2. Applications

The concept of VT has been applied in several contexts so far, with a few studies advancing the existing VT modelling techniques. Friesz et al. (2013) developed a dynamic user equilibrium framework for networks. Traffic dynamics were modelled based on the LWR theory, and numerically represented with VT. Yet, they did not consider spillbacks across intersections in their study, and thus underestimated delays for heavily congested scenarios. Similarly, Li and Zhang (2015) described the performance of queueing systems with multiple sequential and parallel bottlenecks based on VT. However, their framework did not account for spillover effects. Chow et al. (2015) utilized the ability of VT to model complex traffic dynamics at the corridor level on urban arterials in London (UK) and compared it to CTM. They confirmed the high-quality results of VT for such settings, especially for platoon dispersion and moving bottlenecks. Hans et al. (2015) developed a mesoscopic model based on VT to exactly estimate travel times for urban arterials. While the application of VT was still limited to the link and corridor levels, these studies show that VT has been recognized as a powerful numerical scheme for solving KWT problems.

VT is not restricted to the study of car traffic only. It has also been utilized for multi-modal traffic analysis. Guler et al. (2016) investigated innovative transit signal priority designs and estimated its effects on intersection performance with VT. Similarly, Wu et al. (2017) analysed the impacts of bus stop locations and TSP on intersection operations. Wu and Guler (2018) examined the optimal locations of transit signals in the context of mixed traffic with cars and buses. Gayah et al. (2016) studied the impacts of general obstructions on the capacity of an isolated intersection, and computed the capacity losses using VT. Saeednia and Menendez (2016) evaluated the effect of truck platoons on freeway traffic also using VT.

Additionally, VT has been applied for the approximation of the macroscopic fundamental diagram (MFD). The method of cuts introduced by Daganzo and Geroliminis (2008) estimates the MFD for a homogeneous ring road based on VT. Accordingly, related extensions and studies also refer and exploit VT (Boyaci and Geroliminis, 2011; Leclercq and Geroliminis, 2013; Laval and Castrillón, 2015; Tilg et al., 2020a; Daganzo and Knoop, 2016; Leclercq and Paipuri, 2019; Girault et al., 2016; Ambühl et al., 2018; Loder et al., 2019). Again, these approximation methods apply VT only at the corridor level, even when trying to estimate traffic conditions at the network level. Early works have addressed the inclusion of turning ratios into the MFD estimation, see Tilg et al. (2021) and Xu et al. (2020).

Another field of application is traffic state estimation and data fusion (Mehran et al., 2012; Sun and Ban, 2013). This line of research has explored the application of VT to reconstruct vehicle trajectories based on data from fixed and/or mobile sensors, as well as data fusion algorithms. Mehran and Kuwahara (2013) further extended this approach to predict vehicle trajectories based on

real-time and historical data. Duret and Yuan (2017) proposed a traffic state estimation framework based on eulerian and lagrangian observations. Their data fusion framework includes the concept of VT. Chen et al. (2020) reconstructed trajectories of vehicles using a hybrid approach that integrates VT and Kalman filtering. More works on developing data fusion frameworks exploiting the concept of VT concern real-time applications (Kawasaki et al., 2017) and measurements from vehicles running on opposite lanes (Kawai et al., 2019; Takenouchi et al., 2019).

2.3. Methodological extensions

Not only has VT been applied in several contexts, but a number of attempts to extend the original framework have also been reported. For example, Lebacque and Khoshyaran (2013) showed that generic second-order models admit a Hamilton–Jacobi and variational formulation as an optimal control problem. Costeseque and Lebacque (2014) numerically investigated the VT formulation for higher order traffic models. Additionally, the deterministic nature of VT was questioned and by the inclusion of stochastic shortest path algorithms new fields of application were made accessible. Wada et al. (2018) applied VT for coordinated traffic signal control for both deterministic and stochastic demands. Dakic et al. (2020) applied a stochastic shortest path search within VT, to estimate the capacity of bi-modal corridors.

Another interesting line of research is the impact of source terms in VT, i.e. the existence of inflows and outflows in corridors. Eq. (4) implies there are no inflows nor outflows within the corridor. This is because the occurrence of such source terms violates flow conservation along the corridor, which is the main assumption of the conservation equation (see Eq. (1)). While the assumption of zero net inflows might hold for singular corridors, it represents a major limitation for applying VT at the network scale. Laval et al. (2016) studied the incorporation of such source terms into the VT framework. They developed a framework to consider continuous inflows and outflows (i.e. source terms) which could represent trips starting and ending throughout the link. However, their method is not able to handle source terms at intersections, as they are discrete in space and time, and include non-zero values during green and values equal to zero during red phases.

The wide range of applications as well as the methodological extensions of the basic concept show the general interest in VT. These works indicate that extending its applicability from corridors to networks is a valuable contribution. In the next section, we introduce a framework to apply VT at the network level. Thereby, we account for source terms in VT at intersections and model spillbacks. Hereafter, we refer to our framework as ‘nVT’ as abbreviation for network VT, and ‘original VT’ to the formulation introduced by Daganzo (2005a).

3. Generalizing the VT framework to networks

In this paper, we modify VT and establish a comprehensive model that allows to numerically solve complex heterogeneous KWT problems at the network level. We make the following assumptions:

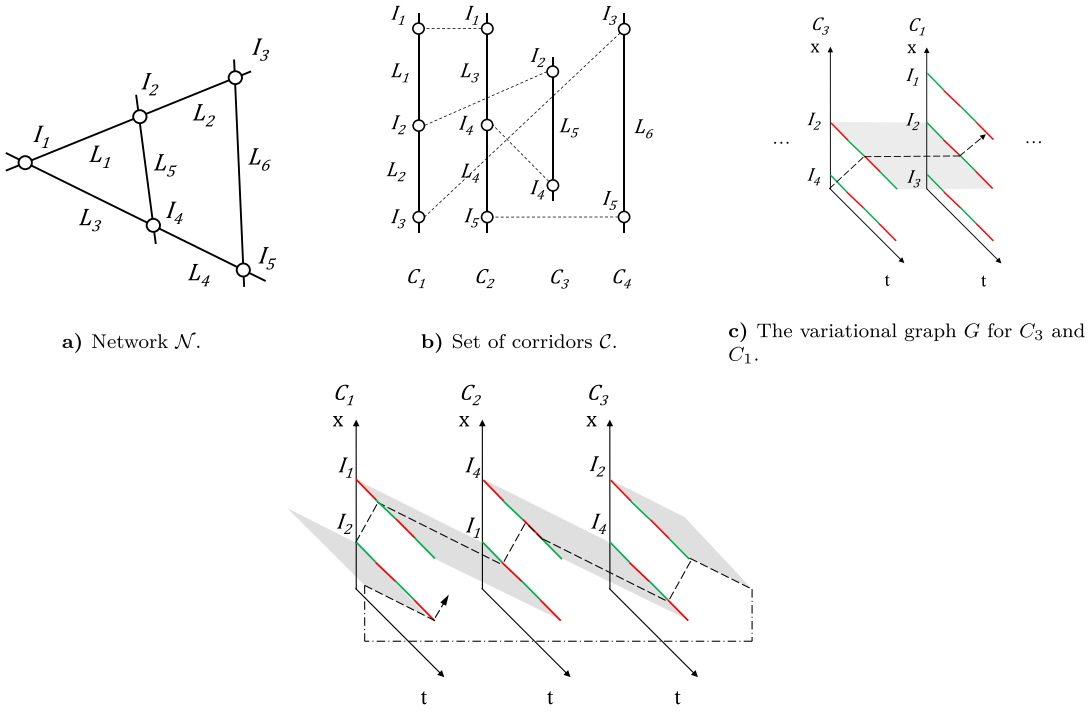
1. *Fundamental diagram*: The application of nVT requires a concave and piecewise differentiable FD. This assumption is inherited from the original VT. Triangular FDs with θ being integer valued facilitate the modelling of traffic in signalized networks.
2. *Variational graph*: In order to consider the effects of bottlenecks as well as source terms, their trajectories need to be fully mapped in G.
3. *Queueing discipline:* Traffic flow follows a first-in-first-out (FIFO) rule.
4. *Network topology*: The proposed framework applies merely to signalized urban networks without any partially conflicting traffic streams. Therefore, the focus lies on networks with signalized intersections with dedicated phases for the conflicting traffic streams.
5. *Turning ratios*: The framework assumes constant turning ratios, which are an exogenous input.

The relaxation of these assumptions is discussed later in Section 7. Below, we describe the overall framework which consists of three steps.

Step 0 initializes the problem by defining the infrastructure and the demand. This consists of the definition of the network topology, the signal control settings, and the temporal and spatial demand patterns. The latter aspect includes the origins as well as turning ratios at each intersection. Note that this implies an indirect definition of destinations. This allows to generate the network $\mathcal{N}$ and the data at the boundaries.

Step 1 involves the decomposition of the network $\mathcal{N}$ into a set of corridors $C \in \mathcal{C}$. This enables us to define a multi-dimensional variational graph in order to solve the given KWT problem. The graph has $(|\mathcal{C}|+1)$ dimensions, consisting of the corridors in $\mathcal{C}$, and one temporal dimension. We provide more details on this step in Section 3.1.

Step 2 applies our VT framework, the nVT, taking the multi-dimensional variational graph as input. Our framework builds an extension that incorporates source terms into the original VT concept. We treat turning flows as source terms at the location of intersections which are discrete in time and space. Moreover, the model replicates spillbacks across intersections in the network. This step is explained in detail in Section 3.2.



a) Network $\mathcal{N}$.

b) Set of corridors $\mathcal{C}$.

c) The variational graph G for C_3 and C_1.

d) The variational graph G for C_1, C_2, and C_3 including circular routes.

Fig. 3. Schematic illustration of the network decomposition and definition of a multi-dimensional variational graph G. The green and red lines illustrate traffic signal phases. The grey planes symbolize inter-corridor connections in $\mathcal{G}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1. Step 1: Decomposition of networks into corridors

The problem initialization results in a network $\mathcal{N}$ consisting of intersections $I \in \mathcal{I}$ and links $L \in \mathcal{L}$. In this paper, we define a link as the road segment between two intersections. Additionally, we specify signal control settings, i.e. red and green times, and turning ratios $\alpha \in A$ at each intersection $I \in \mathcal{I}$, as well as origin flows. As discussed above, the definition of the variational graph requires the decomposition of the network $\mathcal{N}$ into a set of corridors $\mathcal{C}$. We define a corridor as ordered sequence of links in the same direction of travel. Thereby, the topological order including turning ratios α_{ij} between each pair of corridors C_i and C_j as well as control settings for $I \in \mathcal{I}$ have to be stored. Turning ratios are assumed to be constant. The implications of this assumption are discussed in Section 7.

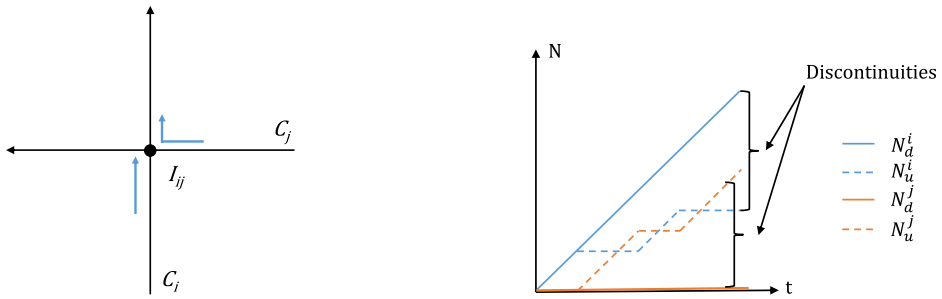
There are two main requirements for decomposing the network $\mathcal{N}$ into a variational graph:

1. The set $\mathcal{C}$ includes all links $L \in \mathcal{L}$.
2. Each link L exists only once in $\mathcal{C}$, i.e. the corridors do not have any overlapping segments.

For small toy networks, we can determine the set $\mathcal{C}$ manually from $\mathcal{N}$. For the most interesting case of real networks, this decomposition can be performed according to the actual layout of roads. From a practical perspective, it is very convenient to decompose real networks according to the network structure as defined by local municipalities. Each link in a real network is part of a 'street', defined by its name, and such streets are non-overlapping. Each street can be represented as a corridor C and as such can be incorporated in G. This will always satisfy both conditions mentioned above. Note that the resulting decomposed network, i.e. the set $\mathcal{C}$, will not affect the result. The KWT solution is determined by the demand (origin flows and turning ratios), as well as the supply (link FDs, link lengths, etc.). As these parameters are not affected by assigning links L to corridors C, the KWT and therefore the VT solution are unaffected as well. In other words, the KWT solution is independent of the decomposition method.

Fig. 3 schematically illustrates the process with an example. Fig. 3a depicts a network $\mathcal{N}$ consisting of five intersections, $\mathcal{I} = \{I_1, I_2, I_3, I_4, I_5\}$, and six unidirectional links, $\mathcal{L} = \{L_1, L_2, L_3, L_4, L_5, L_6\}$. The direction of travel on each link is indicated with the arrows. Note that our method is not limited to unidirectional links, but this assumption simplifies the example. The intersections connecting the different corridors play an important role, as they serve as interfaces where flow is transferred from one corridor to the other, i.e. where source terms apply. We denote such intersections as inter-corridor connections to distinguish them from intersections without any turning flows (e.g. pedestrian crossings). For our example, we assume that $\alpha > 0$ for all intersections. That is, they are all inter-corridor connections.

We manually decompose $\mathcal{N}$ into a set $\mathcal{C} = \{C_1, C_2, C_3, C_4\}$ as shown in Fig. 3b. This set contains only non-overlapping corridors covering the entire network. Both requirements stated above are then fulfilled. The inter-corridor connections are highlighted as dashed lines. As observed, the network's topology is fully retained in this decomposition.



a) Two corridors C_i and C_j and a simple inter-corridor connection.

b) Moskowitz function N for the upstream and downstream demands.

Fig. 4. Free-flow case: network and Moskowitz function. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In Fig. 3c one can see an excerpt of the multi-dimensional variational graph G for corridors C_3 and C_1. Red and green phases are represented by the corresponding colours at the inter-corridor connections. The grey plane illustrates the inter-corridor connection I_2, and a single trajectory from a vehicle changing from one corridor to the other is shown as a dashed line for illustration purposes.

Fig. 3d shows the graph G for corridors C_1, C_2, and C_3 including inter-corridor connections represented as grey planes. These corridors include circular route dependencies. This is highlighted by a trajectory, shown as a bold dashed line, and illustrating a vehicle which starts to drive at intersection I_2, changes the corridor several times, and ends up at intersection I_2 again, but some time later. The figure demonstrates that such trajectories are still feasible within the decomposed network. These trajectories will be reflected by the corresponding cumulative counts. In other words, our framework works independently of demand patterns, whether phenomena such as circular route dependencies exist or not. Such special demand aspects are implicitly included in the turning ratios, which are considered as input in our framework, as described in the following sections.

3.2. Step 2: Integration of inter-corridor connections

Once the multi-dimensional variational graph G is defined, we apply nVT to solve the KWT problem. The original formulation, introduced in Section 2, does not apply as it does not account for discrete source terms. The implications of source terms for variational theory and the underlying concept of the Moskowitz function are shown in Sections 3.2.1 and 3.2.2. Mathematical formulations for upstream and downstream traffic state propagation across inter-corridor connections are presented in Sections 3.2.3 and 3.2.4, respectively.

3.2.1. Discontinuous Moskowitz function

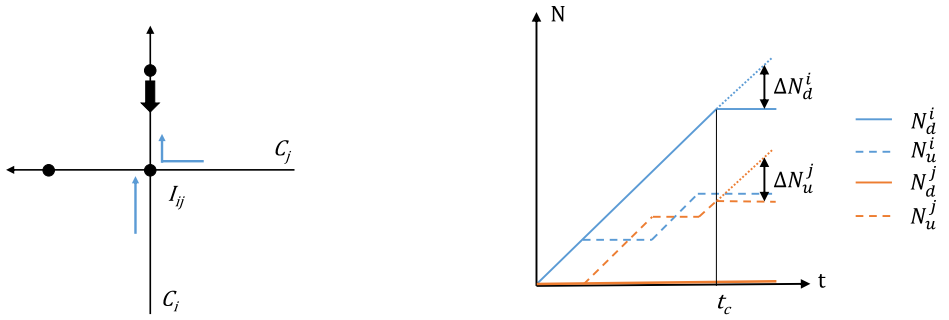
In this section, we highlight the effects of source terms on the Moskowitz function on corridors by means of two examples, a 'free-flow case' and a 'congested case'.

Consider two uni-directional corridors C_i and C_j that intersect each other at the signalized inter-corridor connection I_{ij} as displayed in Fig. 4a. As there exist no bottlenecks on the corridors upstream of I_{ij}, we refer to this example as the 'free-flow case'. Let us denote the Moskowitz function at the position right downstream of I_{ij} on each corridor as N_d^i and N_d^j, respectively. The Moskowitz function for positions right upstream of I_{ij} are N_u^i and N_u^j, respectively. We further assume a high demand on both corridors, such that the inter-corridor connection is saturated. Moreover, we consider a case where all vehicles from C_j merge onto C_i, i.e. the turning ratio $\alpha_{ji} = 1$, and all vehicles on corridor i continue straight at the inter-corridor connection, i.e. $\alpha_{ij} = 0$. Fig. 5b depicts possible time series of N_d^i, N_u^i, N_d^j and N_u^j.

The Moskowitz functions N_u^i and N_u^j at the upstream positions show discharge flows during green phases, and zero flows during red phases. The downstream Moskowitz function for C_i is the sum of both upstream functions, i.e. $N_d^i = N_u^i + N_u^j$, assuming that the travel time between u and d is negligible. The Moskowitz function downstream of the inter-corridor connection on corridor C_j equals to zero, i.e. $N_d^j = 0$, since no vehicles enter from C_i and all flow on C_j turns at the inter-corridor connection I_{ij}.

Recall that the Moskowitz function is in general continuous and respects the flow conservation principle (see Section 2). Since this assumption also applies to the original VT, it requires a continuous surface across space x and time t. However, in our example, a discontinuity of the function occurs at the inter-corridor connection I_{ij}, as N on the same corridor jumps drastically from upstream to downstream of I_{ij} (see Fig. 4b). In other words, no flow conservation applies considering each corridor separately although flow is indeed conserved at the network level, i.e. when flows are aggregated across the two corridors. Note that the discontinuous change in N is positive when a net inflow occurs (e.g. at corridor C_i), and negative when a net outflow occurs (e.g. at corridor C_j). Additionally, the absolute value of the discontinuity increases with time depending on the volumes of the transfer flows.

The second example is the 'congested case' which is depicted in Fig. 5. We examine a simple network with active bottlenecks somewhere downstream of the inter-corridor connection I_{ij}, from which congestion propagates upstream, as schematically illustrated in Fig. 5a. The bottlenecks on each corridor are displayed by the black points, while the bold black arrow represents the



a) Two corridors C_i and C_j, including two bottlenecks, and a simple inter-corridor connection.

b) Moskowitz function N for the upstream and downstream demands.

Fig. 5. Congested case: network and Moskowitz function. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

back-propagation of congestion. Fig. 5b shows possible time-series of the Moskowitz functions on both corridors, upstream and downstream of I_{ij}, i.e. N_d^i, N_u^i, N_d^j and N_u^j.

In the figure, the occurrence of the spillback appears as a reduction in N_d^i growing from $t = t_c$ onward which we denote as ΔN_d^i. To ensure the correct propagation of spillbacks across inter-corridor connections, we have to account for the discontinuity at I_{ij}. Moreover, we have to define how to apportion ΔN_d^i into the N_u of each corridor. As discussed in the previous section, N_d^i is the sum of N_u^i and N_u^j. In order to divide ΔN_d^i into the summands, we have to consider its time-dependency. Flows from the connected corridors can never occur simultaneously for the case of a signalized intersection without partially conflicting flows. This implies that congestion only propagates to the currently discharging corridor. Thus, only the Moskowitz function of this corridor is adapted. In our example, congestion is propagated from N_d^i to N_u^j, i.e. N_u^j is reduced by $\Delta N_u^j = \Delta N_d^i$. Therefore, ΔN_d^i is transferred to N_u^j, while N_u^i stays constant until corridor C_i gets the green light at the inter-corridor connection.

These examples for the 'free-flow case' and the 'congested case' highlight the existence of a discontinuity of the Moskowitz function at inter-corridor connections. Moreover, they show that the propagation of spillbacks necessitates the knowledge of the current signal phase. We have to extend the concept of VT to a discontinuous Moskowitz function to be able to solve the KWT represented by the variational graph G. In the following, we propose a framework to cope with this discontinuity for both the downstream and upstream traffic state propagation. This ensures a complete model including network-wide effects of spillbacks. For the sake of clarity, we avoid using the indices i, j as much as possible for the explanations below, and only use them when strictly necessary. We show that our methodology inherits the numerical exactness from the original VT for triangular FDs.

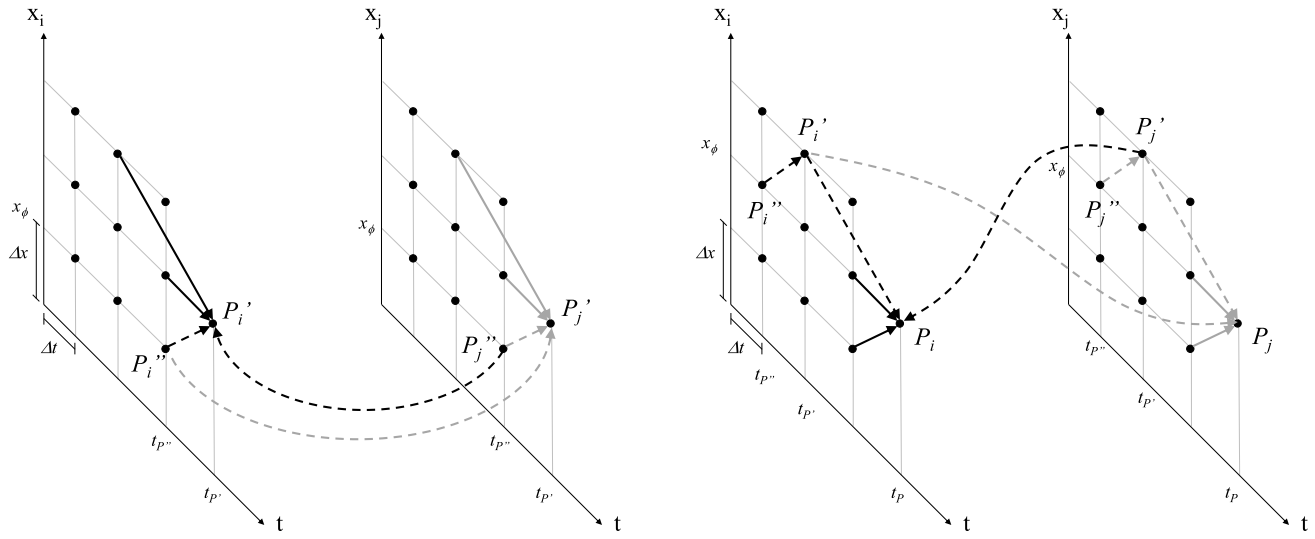
3.2.2. Multi-dimensional variational graph at inter-corridor connections

To further examine the effects of the discontinuous Moskowitz function, the analysis of corresponding excerpts of the variational graph G is convenient. We define the location of an inter-corridor connection I_{ij} in G as x_Φ. At this position, the turning ratio α_{ij} determines which portion of the flow is transferred from one corridor to another. Therefore, $x = x_\Phi + \Delta x$ is the first location in G for a specific corridor, where N is influenced by transfer flows. In other words, the discontinuity is located between x_Φ and $x_\Phi + \Delta x$. For transfer flows to be considered, they need to be fully included in G. In other words, the respective beginning and ending times need to be mapped in G.

We recall that Eq. (4) is valid as long as the Moskowitz function is continuous. It is then valid for all (x,t), where x is not in the vicinity of the inter-corridor connection, i.e. $x \neq x_\Phi$ and $x \neq x_\Phi + \Delta x$. Between those positions, a discontinuity exists and the original formulation has to be modified as per the following two criteria. First, recall that the graphical interpretation of Eq. (4) consists of three from-nodes and the corresponding edges. In order to account for transfer flows in the case of multiple corridors, Eq. (4) should also reflect the traffic states from the adjacent corridor and the turning ratios. Second, Eq. (4) cannot be evaluated at x_Φ and $x_\Phi + \Delta x$ because the edges in G related to the terms of the minimum operation cross the discontinuity which violates the underlying assumptions of VT. This is shown in Fig. 6. Thus, Eq. (4) must be modified for $x = x_\Phi$ and $x = x_\Phi + \Delta x$.

Fig. 6 depicts two excerpts of the multi-dimensional variational graph G for corridors C_i and C_j around the inter-corridor connection located at x_Φ. Edges which cross the discontinuity are shown as dashed lines. The terms in the minimum operation of Eq. (4) corresponding to these dashed edges need to be modified in order to account for the existence of the discontinuity. The black colour represents edges which are evaluated for corridor C_i, and the grey those being evaluated for corridor C_j. We label the points which have to be treated differently from the original formulation. These points are P'' at $(x_\Phi, t_{P''})$, P' at $(x_\Phi + \Delta x, t_{P'})$, and P at (x_Φ, t_P). Note that $t_{P''} = t_P - (1+\theta)\Delta t$, $t_{P'} = t_P - \theta\Delta t$, and $t_{P''} = t_{P'} - \Delta t$. The explanations in Sections 3.2.3 and 3.2.4 refer to these points.

Fig. 6a illustrates the propagation of traffic states from upstream nodes (i.e. free-flow traffic states). This propagation is affected by the discontinuity when the corresponding edges start at x_Φ and end at $x_\Phi + \Delta x$. Thus, the first term in Eq. (4) has to be modified when calculating $N(x,t)$ for $x = x_\Phi + \Delta x$. Fig. 6b shows the propagation of traffic states from downstream nodes (i.e. congested traffic states). It is affected by the discontinuity when the corresponding edges start at $x_\Phi + \Delta x$ and end at x_Φ. Consequently, the



a) Downstream propagation of traffic states.

b) Upstream propagation of traffic states.

Fig. 6. Excerpts of the multi-dimensional variational graph G at the inter-corridor connection located at x_Φ.

second term in Eq. (4) has to be modified when calculating $N(x,t)$ for $x = x_\Phi$. The following subsections describe this extension of Eq. (4) for the propagation of traffic states from upstream and downstream nodes in the vicinity of the inter-corridor connection, i.e. at $x = \{x_\Phi, x_\Phi + \Delta x\}$. Note that downstream and upstream propagation of traffic states do not necessarily refer to the same pair of corridors. The propagation depends on the specific intersection layout and turning possibilities. Consider the following example: At an inter-corridor connection of three corridors A, B, and C, flow can transfer from corridor A to corridor B, and from corridor B to corridor C. That implies that upstream traffic states propagate from corridor A to corridor B, and downstream traffic states propagate from corridor C to corridor B. Such examples are included in the case study shown in Section 6.

3.2.3. *Propagation of traffic states from upstream nodes at $x = x_\Phi + \Delta x$*

This section explains how to include source terms when traffic states propagate from upstream nodes across the discontinuity of the Moskowitz function. In this case, the first term in the minimum operator in Eq. (4) becomes decisive, as it refers to the free-flow states travelling from upstream. We modify this term at $x = x_\Phi + \Delta x$ to account for the upstream demand, and incorporate inflows and outflows based on turning ratios at inter-corridor connections.

Recall that each term in Eq. (4) consists of a known N and the costs c along a path. In order to modify the equation, we adapt these two elements for its first term.

First, we focus on the known N-value at the upstream node, which equals to $N_{P''}$ following the designations in Fig. 6a. We assume that N is known for all $t < t_{P'}$. This is feasible because we can solve for N moving from left to right in the variational graph. The turning ratio α_{ij} specifies the transfer flow from $x_{i,\Phi}$ to $x_{j,\Phi} + \Delta x$. Similarly, the turning ratio α_{ji} specifies the transfer flow from $x_{j,\Phi}$ to $x_{i,\Phi}+\Delta x$. The demand at P_i' can be calculated as the sum of the flow which stays on the corridor C_i, i.e. $N^i_{P''} \cdot (1-\alpha_{ij})$, and the inflow from corridor C_j, i.e. $N^j_{P''} \cdot \alpha_{ji}$. This is depicted by the black dashed lines in Fig. 6a for C_i.

Second, we examine the costs c associated to the path from $N_{P''}$ to $N_{P'}$. Note that $c = 0$ for the propagation of traffic states from upstream nodes, as such states correspond to free-flow conditions. This fact does not change by considering inflows, and therefore the costs for these edges remain equal to zero.

Eventually, we determine the term $N^i_{P'}$ by evaluating the following equation:

$$\begin{aligned} N^i_{P'}(x_\Phi + \Delta x, t_{P'}) = \min(&N^i_{P''}(x_\Phi, t_{P''}) \cdot (1-\alpha_{ij}) + N^j_{P''}(x_\Phi, t_{P''}) \cdot \alpha_{ji}, \\ &N^i(x_\Phi + 2\Delta x, t - \theta\Delta t) + \Delta x \kappa, \\ &N^i(x_\Phi + \Delta x, t - \Delta t) + \beta). \end{aligned} \tag{5}$$

The term $N^j_{P'}$ can be determined analogously. For both corridors, it can be seen that this equation only requires to consider $N_{P''}$ at the position of the inter-corridor connection x_Φ. All other N are known from the original VT formulation. Note that signal phases do not have to be considered explicitly as they are already reflected in the values of $N^i_{P''}$ and $N^j_{P''}$, i.e. they are implicitly taken into account.

3.2.4. *Propagation of traffic states from downstream nodes at $x = x_\Phi$*

In order to complete our modelling framework, we have to ensure the correct propagation of traffic states from downstream nodes across the inter-corridor connections where discontinuities of the Moskowitz function appear. Recall that only congested

traffic states can be propagated from downstream. Thus, we modify the second term in Eq. (4) for $x = x_{\Phi}$ which is decisive when congestion occurs. In other words, in this section we explain how to model spillbacks. Without the loss of generality, the formulation in the following refers to corridor C_i only. The formulation applies analogously for evaluating the traffic conditions on corridor C_j. For the sake of clarity, we avoid using the coordinates of the points P'', P', and P in the following formulations except for the Eq. (10). They are described in Section 3.2.3 and illustrated in Fig. 6.

Again, the second term of the minimum operation consists of two elements. The first element is the known Moskowitz function value $N_{P'}$ at the downstream node $P'(x_{\Phi} + \Delta x, t - \theta t)$. The second one refers to the costs c, i.e. the number of vehicles which can pass a moving observer travelling from the downstream node P' to the point $P(x_{\Phi}, t)$.

We first focus on the value $N_{P'}$. Note that any changes in $N_{P'}$, i.e. $\Delta N_{P'}$ (see Fig. 5b) should first be apportioned to each of the corridors upstream of I_{ij} according to the signal phases and turning ratios. In the absence of congestion the portion of $N_{P'}$ which corresponds to the Moskowitz surface related to N_P would simply be $N_{P''} \cdot (1 - \alpha_{ij})$ as per the first term in Eq. (5). The effect of congestion can be described as a reduction of N, denoted as ΔN in Fig. 5b. Since we know $N_{P'}$ and $N_{P''}$, we can calculate ΔN. Note that corresponding spillbacks can originate on both corridors C_i and C_j. Thus, we have to consider two different scenarios for each corridor. For corridor C_i we can write:

$$\Delta N^{ii}_{P'} = \left(N^{i}_{P''} \cdot (1 - \alpha_{ij}) + N^{j}_{P''} \cdot \alpha_{ji} \right) - N^{i}_{P'}, \tag{6a}$$

$$\Delta N^{ij}_{P'} = \left(N^{i}_{P''} \cdot \alpha_{ij} + N^{j}_{P''} \cdot (1 - \alpha_{ji}) \right) - N^{j}_{P'}. \tag{6b}$$

Recall that VT is built upon the concept of the moving observer. Keeping this notion in mind, one can interpret $\Delta N_{P'}$ as the difference in the number of vehicles a moving observer travelling from $N_{P''}$ to $N_{P'}$ passes due to congestion compared to free-flow conditions. The indices i and j correspond to the corridors where flows originate and propagate to, respectively. For example, $\Delta N^{ij}_{P'}$ relates to vehicles in congestion coming from corridor C_i with destination in corridor C_j. $\Delta N^{ii}_{P'}$ relates to vehicles in congestion coming from C_i with destination in corridor C_i. Note that the portion of $N_{P'}$ related to the upstream Moskowitz functions on each corridor is determined based on the respective turning ratios α, as shown for the free-flow case in Eq. (5). This has to be considered when the change in the vehicle number $\Delta N_{P'}$ is apportioned. We do so by dividing $\Delta N_{P'}$ by the turning ratio associated to each corridor. As an example assume a turning ratio of $\alpha_{ij} = 50\%$ and a $\Delta N^{ij}_{P'} = 5$ vehicles. That would mean that 5 vehicles coming from P''_i did not reach P'_j due to congestion. Moreover, this $\Delta N^{ij}_{P'} = 5$ would mean that 10 vehicles departed P''_i as only every second vehicle wanted to travel to P'_j. To find then the portion of $N_{P'}$ corresponding to the Moskowitz function related to N_P, denoted as $\widehat{N}_{P'}$ in Eq. (7), we subtract from $N_{P''}$ the total number of vehicles leaving P'' being blocked by the congestion at P'. This lets us formulate the effects of congestion at P' related to the Moskowitz function at x_{Φ}.

$$\widehat{N}^{ii}_{P'} = N^{i}_{P''} - \Delta N^{ii}_{P'} \cdot \frac{1}{1 - \alpha_{ij}}, \tag{7a}$$

$$\widehat{N}^{ij}_{P'} = N^{i}_{P''} - \Delta N^{ij}_{P'} \cdot \frac{1}{\alpha_{ij}}. \tag{7b}$$

To find the upper bound for N_P due to congestion, we further have to consider the costs along the path from P' to P which corresponds to the second element of the term related to congestion in Eq. (4). Again, we consider the case of congestion propagation only on corridor C_i. Note that the costs are $\Delta x\kappa$ in the original formulation. In our case, these costs only apply when the signal phase does not change during the interval $[t - (1 + \theta)\Delta t, t]$, i.e. in the time interval between $N_{P''}$ and N_P. However, when a signal phase change occurs during that interval, flows from both corridors might occur at different times within the interval. Thus, the costs c include vehicles originating from both corridors. We need to apportion c according to the vehicles' origin for a correct spillback propagation. In other words, we need to ensure that costs only reflect the vehicles related to the corridor where congestion is propagating to. Otherwise, we would overestimate the costs of the edge, and consequently the second term in Eq. (4). Ultimately, it could result in an overestimation of N_P and thus in an erroneous traffic state propagation from downstream. Assuming reasonable backward wave speeds and that there is only one signal phase change during the considered time interval, there are only two possibilities for such a phase change to occur:

- *Change from red to green:* In this case, a green phase is active at $t = t_P$. The correct maximum number of vehicles from C_i that a moving observer travelling from P'_i to P_i and from P'_j to P_i would count has thus to be reduced by the vehicles coming from C_j. Let the variable r denote the duration of the red phase for C_i starting at $t = t_{P''} = t_P - (1 + \theta)\Delta t$. This is equivalent to the green time for C_j for that same period. Then, the maximum number of vehicles discharging from C_j and going to C_i is $\Delta x\kappa \cdot \frac{r}{(1+\theta)\Delta t} \cdot \alpha_{ji}$. Consequently, the maximum number that can discharge from C_i and stay on that corridor is $\Delta x\kappa - \Delta x\kappa \cdot \frac{r}{(1+\theta)\Delta t} \cdot \alpha_{ji}$. As for $\Delta N_{P'}$, we consider that a part of the total inflow on corridor C_i stays on that corridor and a part turns into corridor C_j, so we divide by the appropriate turning ratios, i.e. $1 - \alpha_{ij}$ for flow staying in C_i and α_{ij} for flow from C_i to C_j. The costs c can then be formulated as:

$$c^{ii} = \frac{1}{1 - \alpha_{ij}} \cdot \left(\Delta x\kappa - \Delta x\kappa \cdot \frac{r}{(1 + \theta)\Delta t} \cdot \alpha_{ji} \right), \tag{8a}$$

$$c^{ij} = \frac{1}{\alpha_{ij}} \cdot \left(\Delta x \kappa - \Delta x \kappa \cdot \frac{r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right). \tag{8b}$$

- *Change from green to red:* In this case, a red phase is active at $t = t_P$. Let the variable g denote the duration of the green phase starting at $t = t_{P''} = t_P - (1+\theta)\Delta t$. Then, the maximum number of vehicles discharging from C_i and staying on this corridor is $\Delta x \kappa \cdot \frac{g}{(1+\theta)\Delta t} \cdot (1-\alpha_{ij})$. Analogously to the previous case, an formulation for the costs c can be found. However, for the case of an active red phase at $t = t_P$, the term related to capacity constraints in Eq. (4) becomes decisive. Therefore, and for the sake of clarity, we do not explicitly state the corresponding mathematical formulation in this paper.

The sum of $\widehat{N}_{P'}$ from Eq. (7) and the costs c from Eq. (8) is an important step for the final formulation of the spillback-induced upper bound for N_P. This sum can be written as:

$$\widehat{N}^{ii}_{P'} + c^{ii} = N^i_{P''} + \frac{1}{1-\alpha_{ij}} \cdot \left[\Delta x \kappa \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot \alpha_{ji} \right) - \Delta N^{ii}_{P'} \right],$$

$$\widehat{N}^{ij}_{P'} + c^{ij} = N^i_{P''} + \frac{1}{\alpha_{ij}} \cdot \left[\Delta x \kappa \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right) - \Delta N^{ij}_{P'} \right].$$

Note that the term in the squared brackets can become negative. This is for example the case when $\Delta N_{P'} = \Delta x \kappa$, i.e. the space between P'' and P' is fully congested, and both r and α are unequal to zero. Since the Moskowitz function is a monotonically increasing function, such values are impossible. We avoid them by applying a maximum function on the term in the squared brackets. Then, we can describe the influence of congestion from corridors C_i and C_j on the Moskowitz function N^i_P as:

$$N^i_P = N^i_{P''} + \frac{1}{1-\alpha_{ij}} \cdot \max\left(0, \Delta x \kappa \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot \alpha_{ji} \right) - \Delta N^{ii}_{P'} \right), \tag{9a}$$

$$N^i_P = N^i_{P''} + \frac{1}{\alpha_{ij}} \cdot \max\left(0, \Delta x \kappa \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right) - \Delta N^{ij}_{P'} \right). \tag{9b}$$

This equation formally corresponds to the intuition that we have to consider the different Moskowitz functions upstream and downstream of the discontinuity, as well as the current signal phases. Eq. (9a) becomes constraining when spillbacks come from corridor C_i, and Eq. (9b) becomes constraining for spillbacks from corridor C_j. The general formulation for $N_P(x_\Phi, t)$ includes then Eq. (9), as well as the free-flow and the capacity related terms. Note that this model conceptually corresponds to a FIFO diverge.

$$\begin{aligned} N^i_P(x_\Phi, t) = \min(& N^i(x_\phi - \Delta x, t - \Delta t), \\ & N^i_{P''}(x_\phi, t - (1+\theta)\Delta t) + \frac{1}{1-\alpha_{ij}} \cdot \max\left(0, \Delta x \kappa \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot \alpha_{ji} \right) - \Delta N^{ii}_{P'} \right), \\ & N^i_{P''}(x_\phi, t - (1+\theta)\Delta t) + \frac{1}{\alpha_{ij}} \cdot \max\left(0, \Delta x \kappa \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot (1-\alpha_{ji}) \right) - \Delta N^{ij}_{P'} \right), \\ & N^i(x_\phi, t - \Delta t) + \beta). \end{aligned} \tag{10}$$

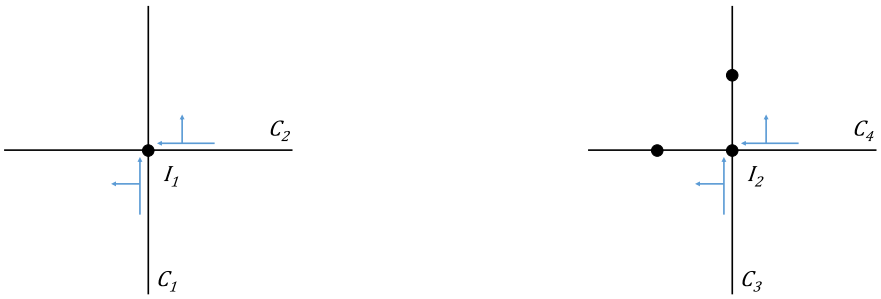
3.3. Implementation

The proposed equations enable us to numerically derive the KWT solution for a network with turning flows. The corresponding pseudo-algorithm is shown in algorithm 1. It describes the application of the proposed equations for a network $\mathcal{N}$. Based on this network, the set of corridors $\mathcal{C}$ is derived. It includes all turning ratios A, the control settings, and the inflow at origins. For a total simulation period T, one iterates through all t for each corridor C and evaluates Eqs. (4)–(10).

Algorithm 1 Application of nVT on a network $\mathcal{N}$

```
Step 0: Initialize network 𝒩 including turning ratios A, control settings, and inflows at origins
Step 1: Derive the set of corridors 𝒞 from 𝒩
Step 2: Apply the nVT framework on 𝒞:
for t in T do
  for C ∈ 𝒞 do
    Evaluate Eq. (4) for all x ≠ x_Φ and x ≠ x_Φ + Δx
    Evaluate Eq. (5) for all x = x_Φ + Δx
    Evaluate Eq. (10) for all x = x_Φ
  end for
end for
```

This nVT model allows to solve complex heterogeneous KWT problems related to networks with signalized control. Such problems involve source terms which are discrete in time and space. As VT allows the incorporation of any type of bottlenecks, nVT inherits this ability. The required inputs are the FD, turning ratios, origin flows, the network topology, and control settings. In the next sections, we will further evaluate our model, compare it to a microscopic KWT simulation in a case study, and discuss its limitations and capabilities.



a) Free-flow scenario. **b)** Congested scenario.

Fig. 7. Test networks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Proof of concept

In this section, we investigate the performance of our nVT model on a small toy network. This proof of concept includes the evaluation of the incorporation of source terms and the propagation of congestion across inter-corridor connections.

We test nVT for a 'free-flow scenario' and a 'congested scenario' on the networks depicted in Fig. 7. The network on the left consists of two corridors, C_1 and C_2, that intersect each other at the inter-corridor connection I_1. Recall that we defined inter-corridor connections as intersections with turning ratios $\alpha > 0$. While the right-hand network is similar, two additional intersections exist, one on each corridor. These intersections act as active bottlenecks downstream of the inter-corridor connection I_2 and therefore lead to congestion. The corridors are labelled as C_3 and C_4, respectively. The designation of the two scenarios refers to the occurring traffic states downstream of the inter-corridor connection. We assume possible turning movements for both inter-corridor connections I_1 and I_2, as illustrated by the blue arrows in the figure. We set a cycle length of 90 s, and a red and green phase of 45 s for all approaches at all intersections. All offsets are set to zero. We assume a triangular FD with jam density $\kappa = 150$ veh/km, a free-flow speed $u = 10$ m/s and a backward wave speed $w = -5$ m/s. The inflow demand is set to a volume-to-capacity ratio of $VOC = 1.0$, which refers to the intersection capacity, for both scenarios.

In the following, we first evaluate the traffic dynamics for the network displayed in Fig. 7a. We discuss our implementation of source flows for the case where no congestion is propagated from downstream of I_1 (Section 4.1). We then analyse the results for the congested case (see Fig. 7b), where the bottlenecks downstream of I_1 are active (Section 4.2). This will confirm that our proposed nVT model is able to capture traffic dynamics in its full range including spillbacks, which occur due to the active bottlenecks. Additionally, we verify our results by reproducing the traffic conditions on the same networks (given its topology and control settings) with the microscopic simulator SUMO (Lopez et al., 2018) and an implementation of Newell's car-following model. This verification is crucial as it allows us to show that our proposed nVT is indeed in line with classical KWT at the network level. As other studies (e.g. Daganzo and Lehe, 2016; Daganzo et al., 2011; Gayah and Daganzo, 2011), our implementation uses Newell's car-following model with unbounded acceleration/deceleration. Yellow intervals, dilemma zones, and reaction times are not modelled. Note that our validation does not verify whether KWT itself is an appropriate model for traffic, as this has been addressed by other studies (e.g. Polson and Sokolov, 2015).

4.1. Inclusion of source terms

The evaluation of nVT for the first network enables us to isolate the effects of source terms for the case where no downstream congestion affects the traffic flow close to the inter-corridor connection. This allows us to study the effects of different turning ratios. We assume turning ratios of $\alpha_{12} = 0.25$ and $\alpha_{21} = 0.5$. The total simulated time is $T = 500$ s, and the time-step length which determines the size of the multi-dimensional variational graph G is $\Delta t = 0.1$ s. Recall that this specifies Δx as well.

Fig. 8 shows the results for both corridors as contour plots. The x-axis displays time in seconds, and the y-axis space in metres. The colour bar represents the density ranging from bright yellow for low densities, to dark blue for high ones. These densities were obtained with the proposed nVT model. We also show the trajectories extracted from SUMO as black and white curves, on top of the contour plots. The black curves depict trajectories originating on corridor C_1, while the white ones correspond to those originating on C_2. The effects of signal phases at the intersection at x = 250 m are clearly identified. The dark blue areas correspond to the jam density and thus represent the evolution of queues upstream of the inter-corridor connection on both corridors. The discharge flows coloured in green can also be clearly observed in the figure. More important are the evident transfer flows between both corridors. This is illustrated by two facts. First, densities $k > 0$ exist downstream of I_1 also during red phases. Second, the discharge flows are split at I_1, as depicted by the different densities upstream and downstream of I_1 for each corridor even during the green phase. This is highlighted by the change in colours. The trajectories from SUMO represent a perfect match to the predicted traffic states by our nVT model. They show that our framework successfully incorporates source terms for the case when congestion does not propagate to the inter-corridor connection.

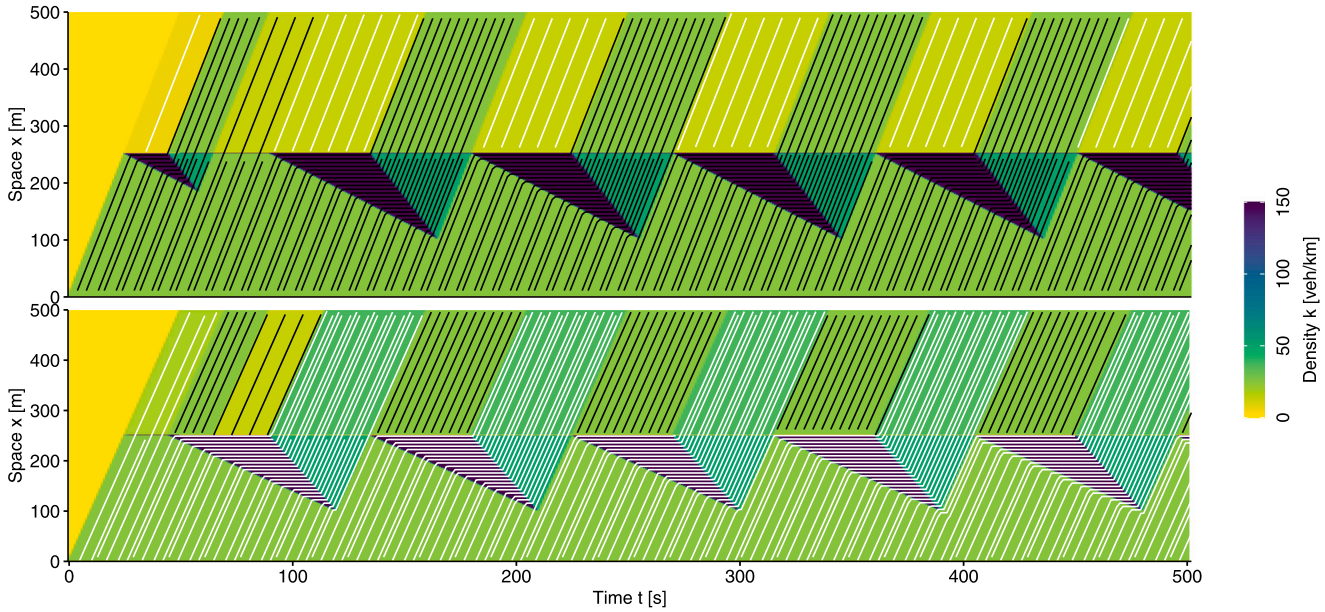


Fig. 8. Free-flow scenario: Time–space diagrams for corridor C_1 (top) and corridor C_2 (bottom) with a traffic signal at I_1 ($x = 250$ m). The colours represent the density which was obtained with the nVT model. The curves indicate trajectories obtained with the microscopic traffic simulator, and their colour indicates the origin (i.e black trajectories originate in C_1 and white ones originate in C_2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Propagation of spillbacks

In the second step, we discuss a more general case where congestion appears downstream of the inter-corridor connection. Our proposed framework should propagate spillbacks along and across corridors. That is, the back of the queue has to spillover to at least one of the corridors, depending on the signal phase.

Fig. 9 shows the simulation results. The axis and colours of the figure are the same as those of Fig. 8. The black curves depict trajectories originating on corridor C_3, while the white ones correspond to those originating on C_4. The graph clearly shows the effects of the additional signals at x = 400 m which behave as active bottlenecks. The congestion starting on corridor C_3 propagates to I_2, where it affects both corridors upstream of I_2. This can be seen by the queues which grow on both corridors with each cycle. Also, it can be seen that there are no flow transfers to a downstream link if the other downstream link is fully congested. That again highlights the FIFO nature of the diverge model. Again, the traffic states predicted by our nVT model fit the trajectories recorded from SUMO, illustrating that our framework is able to successfully propagate congestion across inter-corridor connections.

5. Inheritance of VT properties

Our proposed framework extends the concept of VT to the network level by incorporating source terms. This section demonstrates that nVT inherits (i) the exact numerical calculation of the Moskowitz function apart from potential sampling errors associated to VT, and (ii) the capability of modelling complex intra-link bottlenecks. The combination of these two properties make nVT advantageous compared to other methods such as CTM and LTM.

5.1. Numerical error

In order to solve a given KWT problem with the nVT model, we have to define a multi-dimensional variational graph G. The choice of the time-step size Δt in G is crucial for the computational cost. The original VT has the advantage of being exact in determining the Moskowitz function, independently of the grid size. The only existing error is the so-called 'sampling error' (Daganzo and Menendez, 2005) that originates from an inaccurate sampling of data along the boundary.

The incorporation of discrete source terms at inter-corridor connections does not alter the fact that the framework is based on VT. It also inherits the exact numerical calculation of the Moskowitz function, as well as any potential sampling errors associated to VT. Our framework is able to consider inflows and outflows as long as the corresponding terms are mapped in the variational graph. This is not different to existing VT solutions, where bottlenecks are only considered when they are incorporated in G. Thus, nothing changes compared to the current VT techniques with regard to numerical errors. In order to illustrate this, we evaluate the congested scenario from the previous section (see Fig. 7b) for the time-steps $\Delta t = \{0.1, 1, 5\}$ s. Note that the same analysis could be conducted for any other scenario. We plot N at I_2 across t for both corridors (see Fig. 10).

The y-axis displays the Moskowitz function N, the x-axis shows the time t in seconds. We plot the curves resulting from the different times-steps as solid, dashed and dotted lines. However, comparing the numerical values of all Moskowitz functions $N_{\Delta t=0.1}$,

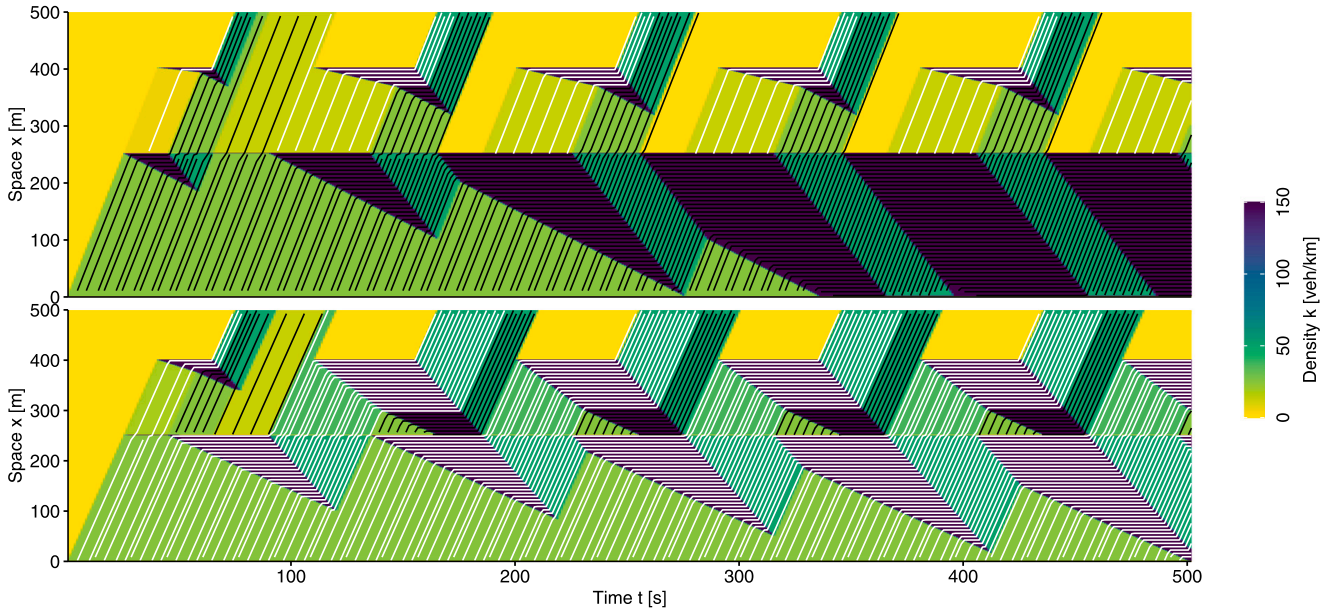
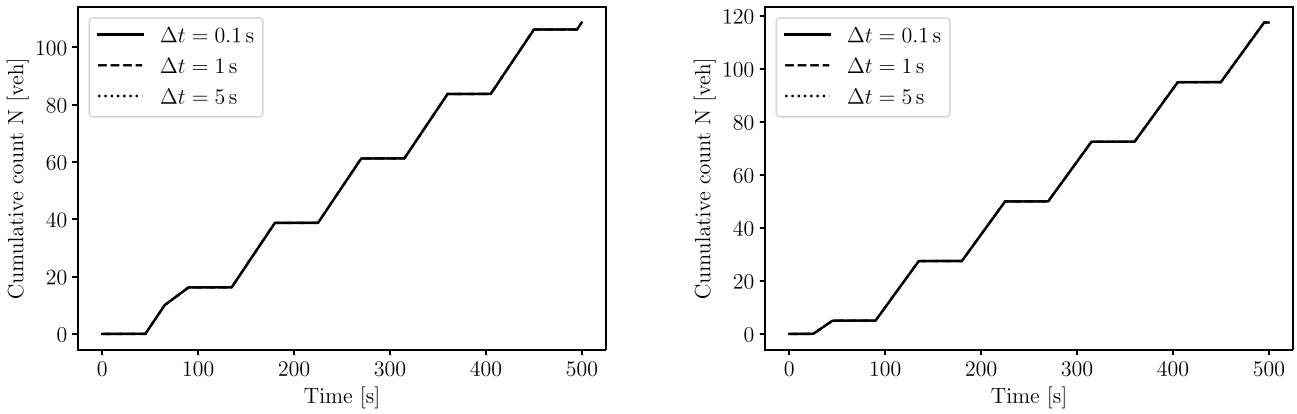


Fig. 9. Congested scenario: Time–space diagrams for corridor C_3 (top) and corridor C_4 (bottom) with a traffic signal at I_2 ($x = 250$ m). The colours represent the density which was obtained with the nVT model. The curves indicate trajectories obtained with the microscopic traffic simulator, and their colour indicates the origin (i.e black trajectories originate in C_3 and white ones originate in C_4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



a) Moskowitz functions at I_2 on corridor C_3.

b) Moskowitz functions at I_2 on corridor C_4.

Fig. 10. Impact of the time-step Δt on the results.

$N_{\Delta t=1}$, and $N_{\Delta t=5}$ reveals that there is no difference between any of them. The figure illustrates this by showing the exact same shape with a perfect overlap of all curves. In other words, the Moskowitz function $N(x,t)$ is found independently of the time-step Δt, as long as all bottlenecks and source terms are mapped in G. This confirms that our framework does not add any additional error to the original VT framework and thus is exact in determining N given a triangular FD and no sampling error at the boundary.

5.2. Capability of modelling complex intra-link bottlenecks

One main advantage of VT is the capability of modelling complex bottlenecks (see Daganzo and Menendez, 2005). This section demonstrates, by means of an example, that despite the extension to networks by considering source terms at intersections, our framework is capable of modelling such complex bottlenecks.

We again apply our model to the network used in the previous section (see Fig. 7b). Additionally, we insert random bottlenecks in both corridors. These could be interpreted as short blockings due to pedestrians crossing the road (Knoop and Daganzo, 2018). In order to model such bottlenecks with nVT, we simply need to set the capacity $\beta = 0$ (see Eq. (4)) where the blockings occur. The ability to represent any bottleneck by adapting the variational graph G allows for an easy and fast implementation of such.

Fig. 11 shows the resulting time–space diagram, where the axis and colours of the figure are the same as those of Fig. 8. Again, the black curves depict trajectories originating on corridor C_3, while the white ones correspond to those originating on C_4. These trajectory data are extracted from SUMO, in which we reproduced the scenario including the short bottlenecks to verify our nVT

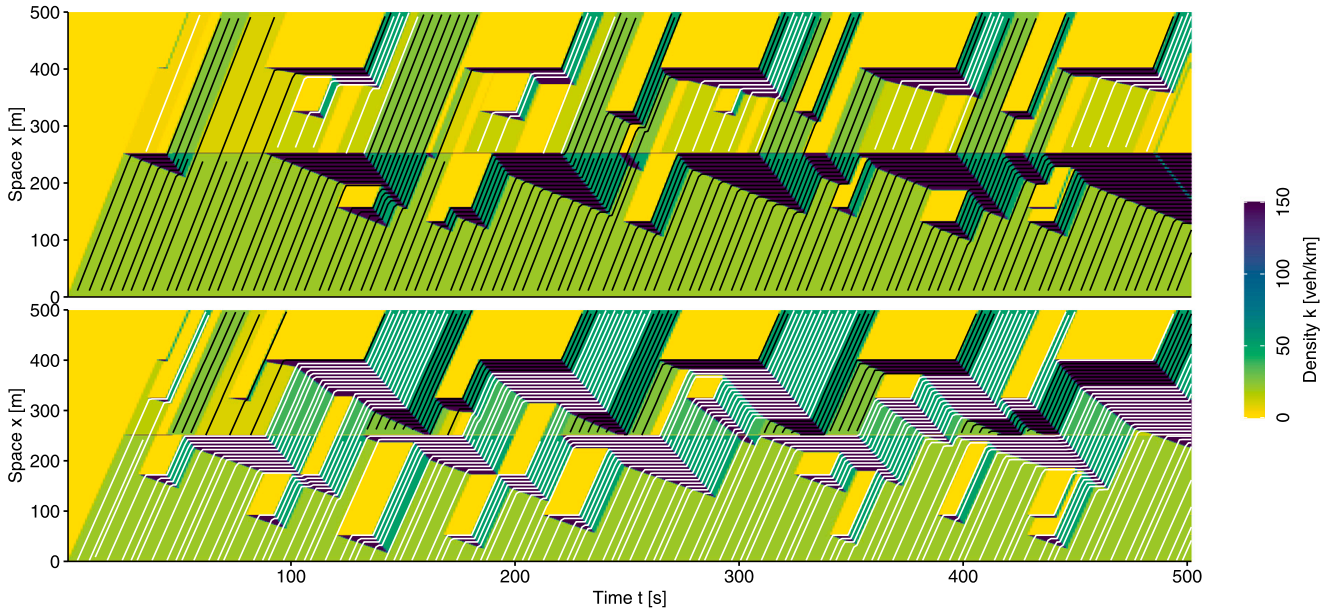


Fig. 11. Random bottlenecks scenario: Time–space diagrams for corridor C_3 (top) and corridor C_4 (bottom) with a traffic signal at I_2 ($x = 250\,\text{m}$), and random bottlenecks. The colours represent the density which was obtained with the nVT model. The curves indicate trajectories obtained with the microscopic traffic simulator, and their colour indicates the origin (i.e black trajectories originate in C_3 and white ones originate in C_4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

solution. Next to queues growing at intersections, the effects of short blockings are clearly visible, with many small queues appearing across the time–space region. In addition to those local effects of short queues, it can also be seen that the overall capacity of the corridor is impacted, and a larger queue in corridor C_4 downstream of intersection I_2 at $x = 250\,\text{m}$ appears, leading to a spillback. Such a spillback additionally affects the adjacent corridor C_3, and thus the entire network. This further confirms the importance of modelling such intra-link bottlenecks even on the network scale. Overall, the graph effectively demonstrates the effects of bottlenecks on traffic dynamics, and clearly shows that the nVT solution fits the SUMO results very well. This illustrates that our model captures the effects of bottlenecks as well as a microscopic KWT simulation, even if such bottlenecks are inserted irregularly in time and space.

6. Application of nVT to a realistic network

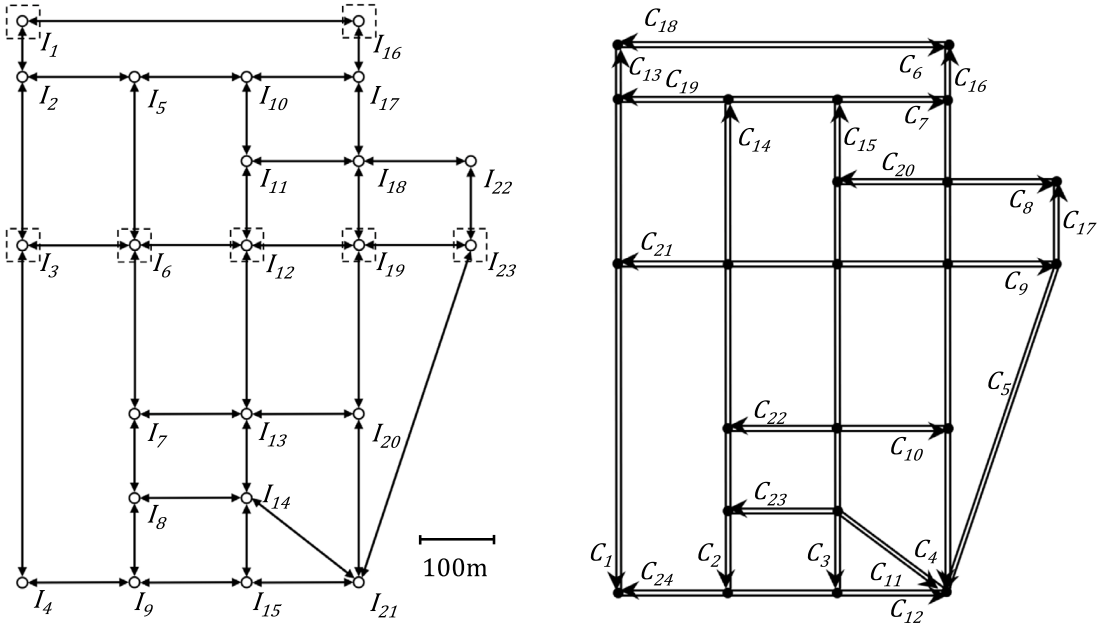
The previous section shows a very good fit of the trajectories from SUMO and the traffic states resulting from our proposed nVT model for scenarios with spillbacks and random intra-link bottlenecks. In this section, we showcase the consistency between our nVT model and the KWT on a larger and more realistic network. For this purpose, we compare the predicted traffic states from nVT to the ground truth. Similar to the examples described in Section 4, we implement Newell's car-following model in SUMO to represent KWT at the network level. Note that this implementation and the associated assumptions are not related to a limitation of nVT, but are only needed to make the KWT solutions of both models comparable. This allows us to derive an appropriate ground truth for our case study, and show the consistency of nVT with KWT at the network level.

We create a network based on the well-known Sioux Falls example.[1] The solution of the KWT is the Moskowitz function $N(x,t)$. It allows us to derive all sorts of indicators, such as average flows, densities, and speeds, but also travel times and delays. Thus, we present our results as Moskowitz functions in this case study. We compare the results derived from our nVT framework to those obtained with SUMO. First, we measure the average flows per cycle according to both methods. This is a suitable aggregation for minimizing the stochastic aspects of SUMO. Second, we show the Moskowitz functions for two locations in the network for the purpose of illustration. Third, we analyse the computational effort of both methods, as the associated cost for macroscopic models is usually low and thus advantageous for numerous applications.

6.1. Case study design

The nature of our framework and that of the microscopic simulation are fundamentally different. The former is macroscopic, and traffic can be deterministically assigned based on origin flows and turning ratios. Moreover, intersections are modelled as points in space. The latter includes a car-following model, as well as routing aspects next to origin flow and turning ratio definitions. In particular, SUMO allows specifying turning ratios at intersections which are then approximated by an internal route-swapping optimization method. This method aims to find a set of individual vehicle routes which approximates the specified turning ratios.

[1] Derived from: https://github.com/bstabler/TransportationNetworks.



a) Original network.

b) Decomposed network and corresponding set of corridors .

Fig. 12. Sioux Falls network for the case study.

However, deviations from the set turning ratios still occur. Also, the geometrical layout of intersections in SUMO deviates from the representation of intersections within the nVT framework, in which intersections are modelled as points in space with no physical dimension. Having these differences between the two models in mind, we apply several simplifications, described in the following, to be able to conduct a reasonable comparison.

Fig. 12a presents the network for which the case study is performed. For the sake of simplicity, we only consider intersections with 4 or fewer legs, and thus reduce the original Sioux Falls network slightly. In total, the network consists of 23 intersections connected by 36 bi-directional links, as highlighted by the arrows. All intersections are signalized with a cycle time of 90 s, and a green and red phase of 45 s. All offsets are set to zero.

In order to apply our nVT model, origin nodes and inflows, as well as turning ratios at inter-corridor connections are required. For our case study, we define the nodes with less than four legs as origins. At these nodes, virtual links are created from which the flows enter the network. Additionally, we define turning ratios for seven intersections, i.e. seven inter-corridor connections. Note that the number of inter-corridor connections is not limited within nVT in general. This number is chosen to further increase the interpretability of the observed phenomena in the case study as it limits the impact of vehicle routing in SUMO; thereby increasing the comparability of both methods. At the same time, these inter-corridor connections assure the occurrence of network traffic phenomena such as circular route dependencies in the presented case study. For the sake of simplicity, turning ratios are set to $\alpha = 0.5$ for all seven inter-corridor connections. Nevertheless, any other value between 0 and 1 is feasible. The inter-corridor connections are highlighted as dashed squares in Fig. 12. Note that by defining origin nodes and turning ratios, no destination nodes need to be set explicitly. We evaluate the case study for five different demand scenarios with volume-to-capacity ratios of $VOC \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ where the capacity again refers to that of intersections. The simulation period is one hour. We choose a time-step of $\Delta t = 0.1$ s, and an FD with the same parameters as described in the previous section.

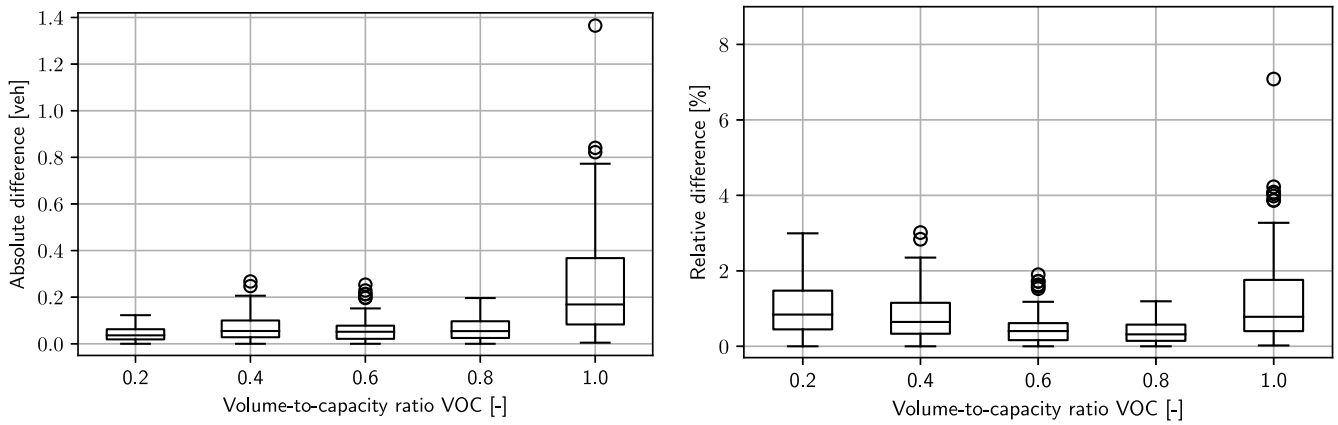
Furthermore, as described in step 1 in Section 3.1, we decompose the network into corridors based on the horizontal and vertical orientation of links. In total we define 24 corridors which are displayed in Fig. 12b. Note that this network specification includes circular route dependencies for the corridors C_1 - C_9 - C_{16} - C_{18}, as well as in the opposite direction C_{13} - C_6 - C_4 - C_{21}. Recall that the method of decomposition does not affect the final result of nVT.

6.2. Results and discussion

6.2.1. Moskowitz functions

To evaluate the results of our nVT model, we extract the Moskowitz function at the upstream end of each link and compare it to the Moskowitz function derived from the SUMO output. Subsequently, we measure the average flow per cycle and calculate the differences between the results from nVT and SUMO. This results in a total of 86 (one per lane, including virtual links) values for each demand scenario. Fig. 13 shows the box-and-whisker diagrams for these differences related to each demand scenario.

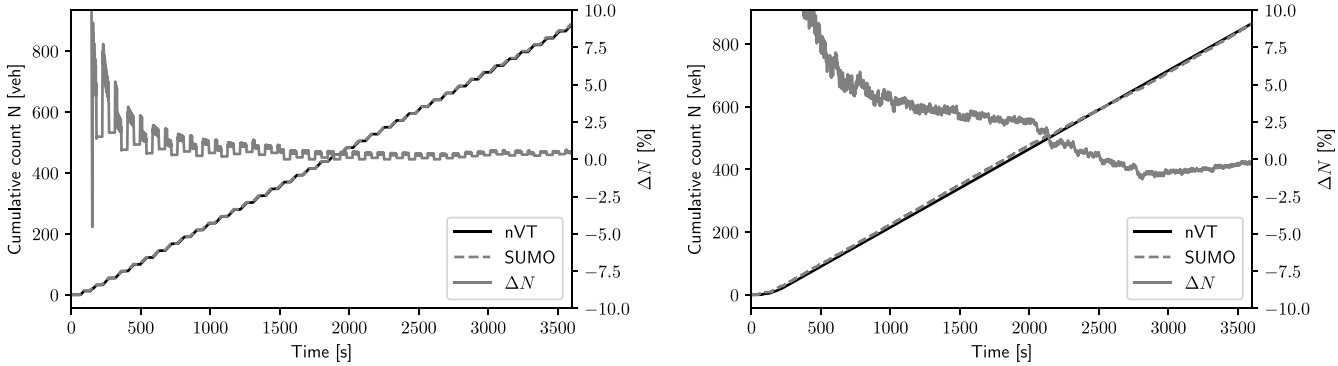
On the left (Fig. 13a) we show the absolute difference, and on the right (Fig. 13b) the relative difference. The latter corresponds to the absolute difference divided by the average number of discharging vehicles per cycle according to SUMO. The absolute and



a) Absolute differences.

b) Relative differences.

Fig. 13. Differences of cycle-based counts between nVT and SUMO for all demand scenarios.



a) Moskowitz functions at intersection I_3.

b) Moskowitz functions at inter-corridor connection I_{18}.

Fig. 14. Moskowitz functions at an intersection and an inter-corridor connections from nVT and SUMO.

relative errors are displayed on the y-axis, whereas the x-axis corresponds to the volume-to-capacity ratio VOC. For most cases, we observe very low differences. The relative difference for scenarios with a low VOC is marginally higher, simply because of the low flows per cycle. The exception is the last scenario that has the highest relative difference. However, these statistical outliers occur due to the high influence of stochasticity and routing complexity when the network is saturated. Therefore, this outcome is expected, as a larger number of vehicles leads to more stochasticity and a higher routing complexity in SUMO. This consequently increases the differences between the SUMO and the nVT results. Nonetheless, most of the observed relative differences are below 3%. Additionally, the median relative differences are all below 1%, which corresponds to median absolute differences below 0.2 vehicles counted per cycle for all scenarios. Given the numerical errors, the above mentioned stochasticity, the occurring circular route dependencies, and the different nature of the microscopic and macroscopic modelling approaches, we adjudge these differences as small.

In addition to these box-and-whisker diagrams, we present two Moskowitz functions in Fig. 14 for illustration purposes. We display the Moskowitz functions N from our nVT approach as solid black lines, and SUMO as dashed grey lines for the intersection I_3 (Fig. 14a), and the inter-corridor connection I_{18} (Fig. 14b). Additionally, we plot the relative difference of both cumulative curves as solid grey lines.

The left y-axis shows the Moskowitz functions N in vehicles downstream of the intersections, while the right y-axis shows the relative difference ΔN in per cent. The x-axis displays the simulation time in seconds. Fig. 14a corresponds to an intersection, where there are no turning flows, i.e. $\alpha = 0$. One can observe the characteristic step-wise increase due to the different signal phases. The relative difference is very high in the beginning, but soon converges to a value close to 0%. The high values for low t are due to the fact that the difference is related to the N from SUMO, and these values are very low at the beginning of the simulation. Moreover, stochastic influences appear to be large for short simulation durations. However, after a reasonable simulation period (i.e. warm-up period), the results from SUMO and nVT match each other very well. Fig. 14b is an example for an inter-corridor connection. Due to the turning ratios $\alpha = 0.5$, a constant flow occurs. Again, both curves match each other very well for longer simulation periods, and the error converges to 0% at the end of the simulation period. Note that the shape of the cumulative curves corresponding to nVT matches the expected step-wise and linear shape, respectively, perfectly. Thus, the main reasons for the existing errors presumably result from the aforementioned fundamental differences between SUMO and nVT. An in-depth sensitivity analysis (Ge et al., 2015)

Table 1
Computational effort comparison of SUMO and nVT.

VOC	Computation time [s]			
	SUMO	nVT ($\Delta t = 0.1$ s)	nVT ($\Delta t = 1$ s)	nVT ($\Delta t = 5$ s)
0.2	33.8	194.0	13.1	2.3
0.4	54.9			
0.6	88.5			
0.8	106.1			
1.0	188.6			

could be conducted to further explore these differences paying special attention to the relation between the different inputs and the outputs of the SUMO simulations, but that is considered out of scope for this paper.

6.2.2. Computational effort

Compared to microscopic simulations, one great advantage of our method is the computational efficiency. This results from the macroscopic nature of the model. We show a comparison of the computational effort of both methods for the presented Sioux Falls case study in Table 1. The methods were evaluated on a computer with an Intel(R) Xeon(R) W-2145 CPU with 3.70 GHz and 64 GB RAM.

The demand scenarios are specified in the left-most column as a function of the VOC. We see that the computational cost in seconds for SUMO scales with the increasing demand. Since each vehicle is simulated separately it is logical that a growing number of vehicles leads to an increasing computational effort. In contrast, this is not the case for nVT since the computational cost scales with the multi-dimensional variational graph G. The computational cost does not scale with demand, i.e. the actual number of vehicles simulated, but only with the number of points in G. The results in the table show the effect of an increased time-step Δt, which leads to a reduced number of points in G, and therefore to a drastic decrease of computational effort required. While $\Delta t = 0.1$ s leads to a high cost, the choice of $\Delta t = 5$ s results in a computational cost of nearly two orders of magnitude lower than the highest one of SUMO. In fact, the absolute difference of three minutes corresponds to roughly 600% increase of computation time from a relative perspective. Such differences become important for use cases where a large number of repetitions of the traffic model are required, e.g. simulation-based optimization or model calibration (e.g Osorio and Bierlaire, 2009; Tilg et al., 2018, 2020b; Ameli et al., 2020; Ge et al., 2014). At the same time, no numerical error is introduced by the larger time-step as indicated in Section 5.1. The appropriate choice of the time-step makes our proposed nVT model suitable for real-time applications or as part of a model-based optimization framework.

Overall, this case study clearly shows the applicability of our nVT model at the network level. The results match those of a microscopic KWT simulation for several demand scenarios very well. Moreover, the nVT framework can evaluate network-wide traffic dynamics at a low computational cost. We conclude that our nVT model represents a first step towards a fully VT-based KWT simulation applicable to large-scale networks.

7. Discussion

In the previous sections we developed the nVT framework, provided a proof of concept, demonstrated the inheritance of properties related to original VT, and showed its applicability for the case of the Sioux Falls network. This section aims at discussing the proposed methodology, highlighting its limitations, and capabilities.

The nVT framework is an extension of the original VT introduced by Daganzo (2005a). It extends the applicability from single corridors without any inflows or outflows to the network level. It incorporates real-valued source terms which are discrete in time and space into the VT framework. This is necessary as flow conservation might not apply anymore at the corridor level. Our nVT model accounts for the corresponding discontinuities in the Moskowitz function and is able to propagate free-flow and congested traffic states across intersections and throughout the entire network while accounting for the effects of complex intra-link bottlenecks. Nevertheless, the included assumptions (Section 3.3) lead to certain limitations which are discussed below:

1. *Fundamental diagram*: Our proposed model inherits the assumption of a concave and piecewise differentiable FD from the original VT. This is a reasonable and common assumption across many traffic models. As written before, triangular FDs with θ being integer valued facilitate the modelling of traffic in signalized networks. However, the general applicability of VT does not require θ to be integer-valued (Daganzo and Menendez, 2005).
2. *Variational graph*: The variational graph G needs to include both bottlenecks and source terms for their effects to be considered. This can be ensured by choosing a Δt small enough to accurately capture the beginning and end of bottlenecks, inflows and outflows. For bottlenecks independent of traffic states, such as fixed-time traffic lights, an accurate mapping is trivial, since the duration of signal phases is known a priori. Traffic-dependent bottlenecks, e.g. buses, are more difficult to consider. Possible methods to deal with them are based on iterative approaches (e.g. Dakic et al., 2020).
3. *Queueing discipline*: The assumed FIFO rule corresponds to links, where flows are not separated according to their destinations before intersections. Implementing diverges where queueing disciplines other than FIFO apply might not be trivial. However, such lanes could be modelled with the proposed model by incorporating diverges prior to the intersection.

4. *Network topology*: Our framework cannot model unprotected turns at intersections. A possible way to include such flows into nVT could be based on the merge model proposed by Daganzo (1995). The implementation of a merge model into nVT is non-trivial as it might violate FIFO rules at the network level. However, it is only necessary for special cases of unprotected turns. Thus, we adjudge this to be out of scope, and refrain from implementing respective models into our framework.
5. *Turning ratios*: nVT assumes constant turning ratios for the simulation period. These could be estimated based on floating car data. This assumption is made to simplify the model formulation, but could potentially be relaxed by including time interval-specific turning ratios in Eq. (10) in Section 3.2.4.

These limitations indicate that for some applications, other KWT solution methods such as the CTM, the LTM, or the time- and space-lagrangian VT formulation might be more appropriate. The choice of the solution method depends on the properties of the problem at hand. Section 5.1 showed that our model inherits the exact evaluation of $N(x,t)$ from the original VT. This makes it superior to the CTM whenever numerical precision is important. Moreover, in Section 5.2 we demonstrated the ability of the developed nVT to model complex intra-link heterogeneities. For example, so-called short blockings which could occur randomly due to high frequent pedestrian crossings on a given road segment (Knoop and Daganzo, 2018) are easily taken into account. Whenever such a blocking occurs, the capacity of the corresponding edge in G is set to zero. Consequently, the bottleneck is considered in the solution method. This is an advantage compared to the LTM, or to VT solutions in lagrangian coordinates, where the incorporation of intra-link heterogeneities is more cumbersome. Furthermore, existing extensions of eulerian VT account for stochastic bus arrivals (Dakic and Menendez, 2018), and continuous source terms within links which could represent starting and ending trips (Laval et al., 2016). Such an incorporation of source terms is not possible in the lagrangian VT formulation. Both, complex intra-link heterogeneities such as random short blockings, and other methodological VT extensions (see Section 2) can be conveniently combined with nVT. Our proposed framework then allows to evaluate the effects at the network level. Last but not least, we showed in Section 6 that the computational effort of our nVT only scales with time and space, similar to the original VT. This is beneficial compared to microscopic simulations which usually also scale with the number of vehicles and thus congestion occurrence.

8. Conclusion

This paper proposes an extension of the eulerian VT (Daganzo, 2005a,b; Daganzo and Menendez, 2005) for the application at the network level. First, the network of interest is decomposed into a set of non-overlapping corridors covering the entire network. Second, we introduce a mathematical framework to account for non-zero inflows and outflows at intersections within VT. By considering the discontinuities in the Moskowitz function which occur due to such inflows and outflows, we account for potential violations of the flow conservation on such corridors. Our nVT model is able to propagate free-flow and congested traffic states across intersections and therefore throughout the entire network. We successfully verify our framework for a simple network with the microscopic simulation SUMO. Additionally, we show that our extension inherits the numerical exact calculation of the Moskowitz function for triangular FDs from the original VT, as well as the capability to model complex intra-link heterogeneities. Finally, we apply our framework to a case study for the Sioux Falls network. We again compare the results to the solution derived from SUMO. The traffic states predicted by our method clearly show a good fit to the ones from SUMO, with median relative differences in cycle-based flows of less than 1% in all considered demand scenarios. Moreover, our results indicate the low computational cost which can be achieved with our VT extension.

We conclude that we successfully extended the eulerian VT for signalized urban networks. For given supply and demand, network-wide traffic conditions can be predicted according to KWT. This includes heterogeneous complex types of KWT problems, including time–space dependent FD, moving bottlenecks, and random short blockings. The literature review showed the wide range of applications of VT, some of which may profit from the possibility to apply VT at the network level as well. For example, it opens the door to a network-wide evaluation of multi-modal traffic management strategies as the design and operation of intermittent bus lanes. Such a design can be formulated as a model-based optimization where nVT can effectively be applied as traffic simulation due to its low computational cost. In summary, the developed model applies to use cases where a microscopic simulation is computationally too expensive, but the existence of intra-link bottlenecks such as buses or pedestrians hinders the efficient application of other macroscopic models such as CTM and LTM. Moreover, our proposed nVT model is an important step to provide the grounds for existing VT applications and other methodological VT extensions at the network level.

While our framework constitutes a clear contribution, many additions are relevant for future work. As we focused merely on signalized intersections, the implementation of merge schemes seems to be the next logical step. One possibility could be the implementation of Daganzo's merge model (Daganzo, 1995) into our methodology. This would additionally allow the modelling of partially conflicting traffic streams at intersections. Moreover, our framework is limited to constant turning ratios. While this still allows varying transfer flows, stochastic turning ratios could also be considered in the future in order to enhance the model. Another topic for future work is the integration of our framework in a traffic assignment. Perceiving such an assignment as an optimization, turning ratios could be the decision variables while the objective could be the network-wide average speeds. Moreover, our model is a first step in the direction of semi-analytical network MFD estimation. Current methodologies such as the method of cuts (Daganzo and Geroliminis, 2008) apply only to single corridors, as they are based on the original VT. The authors are currently working towards such an estimation method (Tilg et al., 2021) applicable to urban networks.

CRediT authorship contribution statement

Gabriel Tilg: Conceptualization, Methodology, Software, Validation, Formal analysis, investigation, Writing - original draft, Writing - review & editing. **Lukas Ambühl:** Conceptualization, Methodology, Software, Investigation, Writing - review & editing. **Sergio Batista:** Conceptualization, Methodology, Investigation, Writing - review & editing. **Monica Menendez:** Conceptualization, Supervision, Methodology, Writing - review & editing. **Fritz Busch:** Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

G. Tilg acknowledges the support by the research project MobiDig: Mobilität Digital Hochfranken funded by the German Federal Ministry of Transport and Digital Infrastructure. S. F. A. Batista and M. Menéndez acknowledge support by the NYUAD Center for Interacting Urban Networks (CITIES), UAE, funded by Tamkeen under the NYUAD Research Institute Award CG001 and by the Swiss Re Institute, UAE under the Quantum Cities™ initiative. L. Ambühl acknowledges the support by the ETH Research Grant ETH-27 16-1 under the project name SPEED.

References

Ambühl, L., Loder, A., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2018. A functional form with a physical meaning for the macroscopic fundamental diagram. Transp. Res. B http://dx.doi.org/10.1016/j.trb.2018.10.013, URL: http://www.sciencedirect.com/science/article/pii/S0191261517310123.

Ameli, M., Lebacque, J.-P., Leclercq, L., 2020. Simulation-based dynamic traffic assignment: Meta-heuristic solution methods with parallel computing. Comput.-Aided Civ. Infrastruct. Eng. 35 (10), 1047–1062. http://dx.doi.org/10.1111/mice.12577, URL: arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/mice.12577.

Boyaci, B., Geroliminis, N., 2011. Estimation of the network capacity for multimodal urban systems. Procedia - Soc. Behav. Sci. 16, 803–813. http://dx.doi.org/10.1016/j.sbspro.2011.04.499.

Chen, P., Wei, L., Meng, F., Zheng, N., 2020. Vehicle trajectory reconstruction for signalized intersections: A hybrid approach integrating Kalman filtering and variational theory. Transportmet. B Transp. Dyn. 1–20. http://dx.doi.org/10.1080/21680566.2020.1781707.

Chow, A.H., Li, S., Szeto, W., Wang, D.Z., 2015. Modelling urban traffic dynamics based upon the variational formulation of kinematic waves. Transportmetr. B Transp. Dyn. 3 (3), 169–191.

Claudel, C.G., Bayen, A.M., 2010a. Lax–Hopf based incorporation of internal boundary conditions into Hamilton–Jacobi equation. Part I: Theory. IEEE Trans. Automat. Control 55 (5), 1142–1157.

Claudel, C.G., Bayen, A.M., 2010b. Lax–Hopf based incorporation of internal boundary conditions into Hamilton-Jacobi equation. part ii: Computational methods. IEEE Trans. Automat. Control 55 (5), 1158–1174.

Costeseque, G., Lebacque, J.-P., 2014. A variational formulation for higher order macroscopic traffic flow models: Numerical investigation. Transp. Res. B 70, 112–133. http://dx.doi.org/10.1016/j.trb.2014.08.012, URL: http://www.sciencedirect.com/science/article/pii/S0191261514001477.

Daganzo, C., 1992. The cell transmission model. Part I: A simple dynamic representation of highway traffic.

Daganzo, C.F., 1995. The cell transmission model, part II: network traffic. Transp. Res. B 29 (2), 79–93.

Daganzo, C.F., 2005a. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transp. Res. B 39 (2), 187–196. http://dx.doi.org/10.1016/j.trb.2004.04.003.

Daganzo, C.F., 2005b. A variational formulation of kinematic waves: Solution methods. Transp. Res. B 39 (10), 934–950. http://dx.doi.org/10.1016/j.trb.2004.05.003.

Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transp. Res. B 45 (1), 278–288. http://dx.doi.org/10.1016/j.trb.2010.06.006.

Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transp. Res. B (42), 771–781. http://dx.doi.org/10.1016/j.trb.2008.06.008.

Daganzo, C.F., Knoop, V.L., 2016. Traffic flow on pedestrianized streets. Transp. Res. B 86, 211–222. http://dx.doi.org/10.1016/j.trb.2015.12.017, URL: http://www.sciencedirect.com/science/article/pii/S0191261516000023.

Daganzo, C.F., Lehe, L.J., 2016. Traffic flow on signalized streets. Transp. Res. B 90, 56–69. http://dx.doi.org/10.1016/j.trb.2016.03.010.

Daganzo, C.F., Menendez, M., 2005. A variational formulation of kinematic waves: Bottleneck properties and examples. In: Proceedings of the 16th International Symposium on Transportation and Traffic Theory.

Dakic, I., Ambühl, L., Schümperlin, O., Menendez, M., 2020. On the modeling of passenger mobility for stochastic bi-modal urban corridors. Transp. Res. C 113, 146–163. http://dx.doi.org/10.1016/j.trc.2019.05.018.

Dakic, I., Menendez, M., 2018. On the use of Lagrangian observations from public transport and probe vehicles to estimate car space-mean speeds in bi-modal urban networks. Transp. Res. C 91, 317–334. http://dx.doi.org/10.1016/j.trc.2018.04.004.

Duret, A., Yuan, Y., 2017. Traffic state estimation based on Eulerian and Lagrangian observations in a mesoscopic modeling framework. Transp. Res. B 101, 51–71.

Friesz, T.L., Han, K., Neto, P.A., Meimand, A., Yao, T., 2013. Dynamic user equilibrium based on a hydrodynamic model. Transp. Res. B 47, 102–126.

Gayah, V.V., Daganzo, C.F., 2011. Effects of turning maneuvers and route choice on a simple network. Transp. Res. Rec. (2249), 15–19. http://dx.doi.org/10.3141/2249-03.

Gayah, V.V., Ilgin Guler, S., Gu, W., 2016. On the impact of obstructions on the capacity of nearby signalised intersections. Transportmetr. B 4 (1), 48–67. http://dx.doi.org/10.1080/21680566.2015.1052111.

Ge, Q., Ciuffo, B., Menendez, M., 2014. An exploratory study of two efficient approaches for the sensitivity analysis of computationally expensive traffic simulation models. IEEE Trans. Intell. Transp. Syst. 15 (3), 1288–1297.

Ge, Q., Ciuffo, B., Menendez, M., 2015. Combining screening and metamodel-based methods: An efficient sequential approach for the sensitivity analysis of model outputs. Reliab. Eng. Syst. Saf. 134, 334–344.

Girault, J.-T., Gayah, V.V., Guler, I., Menendez, M., 2016. Exploratory analysis of signal coordination impacts on macroscopic fundamental diagram. Transpor. Res. Record: J. Transpor. Res. Board (2560), 36–46. http://dx.doi.org/10.3141/2560-05.

Guler, S.I., Gayah, V.V., Menendez, M., 2016. Bus priority at signalized intersections with single-lane approaches: A novel pre-signal strategy. Transp. Res. C 63, 51–70.

Han, K., Piccoli, B., Szeto, W., 2015. Continuous-time link-based kinematic wave model: formulation, solution existence, and well-posedness. Transportmetr. B Trans. Dyn. 4 (3), 187–222.

Hans, E., Chiabaut, N., Leclercq, L., 2015. Applying variational theory to travel time estimation on urban arterials. Transp. Res. B 78, 169–181. http://dx.doi.org/10.1016/j.trb.2015.04.004, URL: http://www.sciencedirect.com/science/article/pii/S019126151500079X.

Jin, W.-L., 2015. Continuous formulations and analytical properties of the link transmission model. Transp. Res. B 74, 88–103.

Kawai, K., Takenouchi, A., Ikawa, M., Kuwahara, M., 2019. Traffic state estimation using traffic measurement from the opposing lane—Error analysis based on fluctuation of input data. In: Intelligent Transport Systems for Everyone's Mobility. Springer, pp. 247–263.

Kawasaki, Y., Hara, Y., Kuwahara, M., 2017. Real-time monitoring of dynamic traffic states by state-space model. Transport. Res. Proc. 21, 42–55. http://dx.doi.org/10.1016/j.trpro.2017.03.076.

Knoop, V.L., Daganzo, C.F., 2018. The effect of crosswalks on traffic flow. Eur. J, Transp. Infrastr. Res. 18 (2), 145–157. http://dx.doi.org/10.18757/ejtir.2018.18.2.3227.

Laval, J.A., Castrillón, F., 2015. Stochastic approximations for the macroscopic fundamental diagram of urban networks. Transp. Res. B 81, 904–916. http://dx.doi.org/10.1016/j.trb.2015.09.002.

Laval, J.A., Costeseque, G., Chilukuri, B., 2016. The impact of source terms in the variational representation of traffic flow. Transp. Res. B 94, 204–216. http://dx.doi.org/10.1016/j.trb.2016.09.011.

Laval, J.A., Leclercq, L., 2013. The Hamilton-Jacobi partial differential equation and the three representations of traffic flow. Transp. Res. B 52, 17–30. http://dx.doi.org/10.1016/j.trb.2013.02.008.

Lebacque, J.P., Khoshyaran, M.M., 2013. A variational formulation for higher order macroscopic traffic flow models of the GSOM family. Transp. Res. B 57, 245–265. http://dx.doi.org/10.1016/j.trb.2013.07.005.

Leclercq, L., Becarie, C., 2012. Meso lighthill-whitham and richards model designed for network applications. In: 91th Annual Meeting of the Transportation Research Board (TRB 2012), Washington, DC.

Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. Transp. Res. B 57 (EPFL-ARTICLE-188334), 468–484. http://dx.doi.org/10.1016/j.trb.2013.05.005.

Leclercq, L., Laval, J., Chevallier, E., 2007. The Lagrangian coordinate system and what it means for first order traffic flow model. In: The proceedings of the 17th International Symposium on Transportation and Traffic Theory, New York, USA.

Leclercq, L., Paipuri, M., 2019. Macroscopic traffic dynamics under fast-varying demand. Transp. Sci. 53 (6), 1526–1545. http://dx.doi.org/10.1287/trsc.2019.0908.

Li, J., Zhang, H.M., 2015. Bounding tandem queuing system performance with variational theory. Transp. Res. B 81, 848–862. http://dx.doi.org/10.1016/j.trb.2015.07.013.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves II. A theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 229 (1178), 317–345.

Loder, A., Dakic, I., Bressan, L., Ambühl, L., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2019. Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram. Transp. Res. B 129, 1–19.

Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wießner, E., 2018. Microscopic traffic simulation using SUMO. In: The 21st IEEE International Conference on Intelligent Transportation Systems. IEEE,

Mahut, M., Florian, M., Tremblay, N., 2003. Space-time queues and dynamic traffic assignment: A model, algorithm and applications. In: Transportation Research Board 82nd Annual MeetingTransportation Research Board.

Mazaré, P.E., Dehwah, A.H., Claudel, C.G., Bayen, A.M., 2011. Analytical and grid-free solutions to the lighthill-whitham-richards traffic flow model. Transp. Res. B 45 (10), 1727–1748. http://dx.doi.org/10.1016/j.trb.2011.07.004.

Mehran, B., Kuwahara, M., 2013. Fusion of probe and fixed sensor data for short-term traffic prediction in urban signalized arterials. Int. J. Urban Sci. 17 (2), 163–183. http://dx.doi.org/10.1080/12265934.2013.776291.

Mehran, B., Kuwahara, M., Naznin, F., 2012. Implementing kinematic wave theory to reconstruct vehicle trajectories from fixed and probe sensor data. Transp. Res. C 20 (1), 144–163.

Moskowitz, K., 1965. Discussion of 'freeway level of service as influenced by volume and capacity characteristics' by D.R. Drewand C. J. Keese. Highway Res. Record 99, 43—44.

Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part I: General theory, part II: Queueing at freeway bottlenecks, part III: Multi-destination flows. Transp. Res. B 27 (4), 281–313.

Osorio, C., Bierlaire, M., 2009. An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. European J. Oper. Res. 196 (3), 996–1007.

Polson, N., Sokolov, V., 2015. Bayesian analysis of traffic flow on interstate I-55: The LWR model. Ann. Appl. Stat. 9 (4), 1864–1888. http://dx.doi.org/10.1214/15-AOAS853, arXiv:1409.6034, URL: https://arxiv.org/pdf/1409.6034.pdf.

Richards, P.I., 1956. Shock waves on the highway. Oper. Res. 4 (1), 42–51.

Saeednia, M., Menendez, M., 2016. A decision support system for real-time platooning of trucks. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1792–1797.

Seo, T., Bayen, A.M., Kusakabe, T., Asakura, Y., 2017. Traffic state estimation on highway: A comprehensive survey. Annu. Rev. Control 43, 128–151.

Sun, Z., Ban, X.J., 2013. Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors. Transp. Res. C 36, 268–283.

Takenouchi, A., Kawai, K., Kuwahara, M., 2019. Traffic state estimation and its sensitivity utilizing measurements from the opposite lane. Transp. Res. C 104 (April), 95–109. http://dx.doi.org/10.1016/j.trc.2019.04.016.

Tampère, C.M., Corthout, R., Cattrysse, D., Immers, L.H., 2011. A generic class of first order node models for dynamic macroscopic simulation of traffic flows. Transp. Res. B 45 (1), 289–309. http://dx.doi.org/10.1016/j.trb.2010.06.004.

Tilg, G., Ambühl, L., Batista, S., Menendez, M., Leclercq, L., Busch, F., 2021. Semi-analytical estimation of macroscopic fundamental diagrams: From corridors to networks. In: 100th Annual Meeting of the Transportation Research Board (TRB 2021). Washington, DC.

Tilg, G., Amini, S., Busch, F., 2020a. Evaluation of analytical approximation methods for the macroscopic fundamental diagram. Transp. Res. C 114, 1–19. http://dx.doi.org/10.1016/j.trc.2020.02.003, URL: http://www.sciencedirect.com/science/article/pii/S0968090X19307661.

Tilg, G., Ul Abedin, Z., Amini, S., Busch, F., 2020b. Simulation-based design of urban bi-modal transport systems. Front. Future Transpor. 1.

Tilg, G., Yang, K., Menendez, M., 2018. Evaluating the effects of automated vehicle technology on the capacity of freeway weaving sections. Transp. Res. C 96, 3–21.

Wada, K., Usui, K., Takigawa, T., Kuwahara, M., 2018. An optimization modeling of coordinated traffic signal control based on the variational theory and its stochastic extension. Transp. Res. B 117, 907–925. http://dx.doi.org/10.1016/j.trb.2017.08.031, URL: http://www.sciencedirect.com/science/article/pii/S019126151730749X, TRB:ISTTT-22.

Wu, K., Guler, S.I., 2018. Optimizing transit signal priority implementation along an arterial. Transp. Res. Rec. 2672 (20), 215–227. http://dx.doi.org/10.1177/0361198118790324.

Wu, K., Guler, S.I., Gayah, V.V., 2017. Estimating the impacts of bus stops and transit signal priority on intersection operations: Queuing and variational theory approach. Transp. Res. Rec. (ISSN: 21694052) 2622 (1), 70–83. http://dx.doi.org/10.3141/2622-07.

Xu, G., Yu, Z., Gayah, V.V., 2020. Analytical method to approximate the impact of turning on the macroscopic fundamental diagram. Transp. Res. Rec. 0361198120933274.

Yperman, I., 2007. The link transmission model for dynamic network loading (Ph.D. thesis). Katholieke Universiteit Leuven.

From corridor to network macroscopic fundamental diagrams: A semi-analytical estimation approach

Gabriel Tilg [a,], Lukas Ambühl [b], Sergio Batista [c], Monica Menendez [c], Ludovic Leclercq [d], Fritz Busch [a]

[a] *Chair of Traffic Engineering and Control, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Germany*
[b] *Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zurich, Switzerland*
[c] *Division of Engineering, New York University Abu Dhabi, United Arab Emirates*
[d] *Univ. Gustave Eiffel, Univ. Lyon, ENTPE, LICIT, F-69518, Lyon, France*

Abstract

The design of network-wide traffic management schemes or transport policies for urban areas requires computationally efficient traffic models. The macroscopic fundamental diagram (MFD) is a promising tool for such applications. Unfortunately, empirical MFDs are not always available, and semi-analytical estimation methods require a reduction of the network to a corridor which introduces substantial inaccuracies. We propose a semi-analytical methodology to estimate the MFD for realistic urban networks, without the information loss induced by the reduction of networks to corridors. The methodology is based on the method of cuts, but applies to networks with irregular topologies, accounts for different spatial demand patterns, and determines the upper bound of network flow. Thereby, we consider flow conservation and the effects of spillbacks, both at the network level. Our framework decomposes the network into a set of corridors, creates a hypernetwork including the impacts of source terms, and then treats the dependencies across corridors (e.g. due to turning flows and spillbacks). Based on this hypernetwork, we derive the free-flow and capacity branch of the MFD. The congested branch is estimated by considering gridlock characteristics and utilizing recent advancements in MFD research. We showcase the applicability of the proposed methodology in a case study with a realistic setting based on the Sioux Falls network. We then compare the results to the original method of cuts, and a ground truth derived from the cell transmission model. The results analysis reveals that our method is over five times more accurate than the state of the art in estimating the network-wide capacity and jam density. Moreover, they clearly indicate the MFD's dependency on spatial demand patterns. Compared to simulation-based MFD estimation approaches, the potential of the proposed framework lies in the modeling flexibility, explanatory value, and reduced computational cost.

Keywords: macroscopic fundamental diagram, method of cuts, network modeling, traffic flow theory, variational theory

1. Introduction

Population growth in large metropolitan areas leads to disruptions in the transportation system. Improving its efficiency essentially requires the design of appropriate monitoring and control schemes. Aggregated

E-mail addresses: gabriel.tilg@tum.de (Gabriel Tilg)., lukas.ambuehl@ivt.baug.ethz.ch (Lukas Ambühl)., sergio.batista@nyu.edu (Sergio Batista)., monica.menendez@nyu.edu (Monica Menendez)., ludovic.leclercq@univ-eiffel.fr (Ludovic Leclercq)., fritz.busch@tum.de (Fritz Busch).

traffic models based on the macroscopic fundamental diagram (MFD) (Godfrey, 1969; Herman and Prigogine, 1979; Mahmassani et al., 1987; Daganzo, 2007; Geroliminis and Daganzo, 2008) are powerful tools, consistent with traffic flow theory, for modeling system dynamics in large urban areas. These models require the definition of regions in the city network, where traffic conditions are approximately homogeneous, thus showing a clear relationship between travel production and the accumulation of vehicles. Such models have a wide range of applications, including the design of control strategies (Keyvan-Ekbatani et al., 2012; Aboudolas and Geroliminis, 2013; Yang et al., 2018; Geroliminis et al., 2012; Kouvelas et al., 2017; Sirmatel and Geroliminis, 2018, 2020; Ni and Cassidy, 2019; Ingole et al., 2020; Haddad and Mirkin, 2020; Haddad and Zheng, 2020; Guo and Ban, 2020; Batista et al., 2021; Sirmatel et al., 2021; Li et al., 2021), pricing schemes (Zheng et al., 2016; Yang et al., 2019), public transport priority (Chiabaut et al., 2018; Haitao et al., 2019) and traffic management (Yildirimoglu et al., 2018; Amini et al., 2018; Batista and Leclercq, 2019; Batista and Leclercq, 2020; Aghamohammadi and Laval, 2020; Dandl et al., 2020).

The respective traffic models require the estimation of the MFD to describe the evolution of aggregated traffic states. In the absence of real data, or for hypothetical scenario analyses, one solution is to use microscopic simulations. Unfortunately, these have high computational costs, and their results remain sensitive to the specific demand patterns. Alternatively, analytical solutions exist for estimating the MFD. For example, Ambühl et al. (2020) developed such an analytical model. However, limited or unavailable a-priori knowledge limits the simple estimations with the proposed functional form. Another possibility is to use semi-analytical approaches, such as the method of cuts (MC) (Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013), or the stochastic approximation (SA) (Laval and Castrillón, 2015); see Tilg et al. (2020a) for an analysis of both methods. So far, however, these methods only explicitly model single corridors, which are defined as an ordered sequence of links. Hence, their application is limited to small and regular synthetic urban networks, where the difference between the MFD of an abstract corridor and that of the whole network is minor. Hitherto, the literature lacks a comprehensive framework to semi-analytically estimate MFDs for realistic, complex urban networks. Below we highlight some major limitations of the corridor simplification.

The topological mapping of a network into a single corridor is not necessarily accurate. Existing approaches assume that control and network topology are relatively similar and regular throughout the network (i.e. similar offsets, green-to-cycle ratios, block length, etc.). Thus, the network average features can be mapped into a single corridor. The *idealized* MFD refers to a theoretical upper bound for traffic conditions for such a corridor (Daganzo, 2007). However, the topological features of the network are lost, leading to inaccuracies in the approximated MFD. For example, the capacity of a single corridor is typically constrained by the most restrictive intersection. This might not be true for general networks, since in many cases multiple routes connect the same origin-destination (OD) pair. In addition to these deficiencies from a supply perspective, the aggregation of a network to one corridor also simplifies demand-related aspects. For example, it implies that traffic dynamics evolve in a similar manner on all corridors in the network. This is a strong assumption, as it rarely occurs in urban networks with a large number of OD pairs which lead to very complex demand patterns. Furthermore, diverging flows at intersections inevitably lead to undersaturated conditions at downstream links. Neglecting this contradicts the network-wide flow conservation. Lastly, the effects of spillbacks may propagate on adjacent corridors. Thus, the need for an efficient approximation of the network MFD considering these aspects is obvious.

We should also note that in many cases, the observed macroscopic performance of an urban network is substantially lower than the theoretically derived idealized MFD. The reason for this lies in the observed traffic heterogeneity, non-stationary traffic states, and the path-flow distribution of the demand (Mahmassani et al., 2013; Knoop et al., 2015; Leclercq et al., 2015; Geroliminis and Sun, 2011; Mazloumian et al., 2010). These observed aggregated traffic states are herein denoted as the *realized* MFD, in contrast to the *idealized* MFD (Ambühl et al., 2021; Loder et al., 2019). Unfortunately, the two different MFD notions are currently used interchangeably. Both MFD types are very similar for a corridor, under steady-state conditions, and homogeneous demand. However, they might differ significantly for general networks. Simply put, the idealized MFD corresponds to stationary flows resulting from a spatial demand pattern that perfectly utilizes the given supply. Thus, the ambiguity of results stemming from models that simplify the network to a corridor, as required by existing semi-analytical approaches, and neglect the network-specific

influence of demand-related aspects, becomes apparent. First approaches to tackle these issues are reported in Geroliminis and Boyacı (2012) and Xu et al. (2020), which incorporate turning flows at intersections. The former study analyzes these effects on a simulation basis and focuses on incoming turns. The second one models the effects merely stochastically. Neither of them describes network-wide spillbacks nor network conservation explicitly, which is important to identify the underlying mechanisms for congestion dynamics. Only such understanding enables the proper design of measures to mitigate congestion and increase urban network capacities. The proposed methodology addresses this by considering the effects of turning ratios on the network MFD. The realized MFD is especially important because it replicates observed traffic states based on which further applications (e.g. traffic control, transport planning, analysis of local bottleneck effects at the network level) can be designed. Additionally, as we will discuss later, it could potentially be extended to also estimate the idealized MFD, although this is left for future research.

In this paper, we propose a semi-analytical framework to estimate the realized MFD at the network level. Thereby, we do not rely on extensive empirical traffic data sets or expensive microscopic simulations. In contrast, we develop a framework based on semi-analytical methods but do not reduce the network to a corridor. We focus on a single region and explicitly consider flow conservation as well as the effects of spillbacks at the network level. Thus, we can account for the effects of different demand patterns on the MFD's shape. The contributions are threefold:

1. We develop a general framework to semi-analytically estimate the MFD at the network level. So far, existing methods were not able to account for spatial demand patterns and the related violation of the conservation of flows. Our framework proposes a way to address this challenge. We decompose a given network into a set of corridors and construct a hypernetwork in the spirit of the original MC. The decomposition maintains the connection of different corridors and, thus, turning relations are actively considered. These inter-corridor connections enable us to model the inter-dependencies of traffic dynamics between corridors in the network, e.g. the network-wide effects of spillbacks. This is a major difference to current approaches, where multiple corridors are either reduced to a single one or modeled as being largely independent of each other, which may lead to inaccurate traffic state estimations.

2. Considering the effects of turning relations, we propose three different approaches to derive the maximum flows at intersections, and two different approaches to approximate the network-wide jam density. Hereby, the challenge is to model queue propagation while accounting for traffic state interdependencies. To our best knowledge, no method exists so far to approximate the network-wide jam density. With the maximum flows at intersections and the network-wide jam density, the MFD is then estimated following the philosophy of the original MC. The proposed approaches vary in terms of modeling complexity, computational cost, and estimation accuracy. Thus, they offer high flexibility and enable the choice of a case-specific MFD estimation. Moreover, they shed light on the effects of related assumptions and thereby provide additional explanatory value.

3. We evaluate the proposed framework using a realistic network. Within this case study, we compare the proposed approaches to the state of the art for such networks and the cell transmission model (CTM) (Daganzo, 1992, 1995) acting as a ground truth. Thereby, we provide proof of our concept and test the framework for a range of different input parameters.

The remainder of this paper is organized as follows. First, we provide a brief summary of the concept of variational theory (VT), the original MC, as well as existing attempts to extend it for realistic networks with source terms. Then, we present the framework to estimate the realized MFD at the network level. Subsequently, we conduct a case study for the realistic network of Sioux Falls. Finally, we draw conclusions and outline potential avenues for future work. We provide a nomenclature table, including all used symbols in appendix A. Appendix B and C provide additional equations and mathematical derivations.

2. Review of the variational theory and the method of cuts

This section briefly describes the basics of VT and MC as our framework builds upon these methods. Moreover, we highlight the deficiencies of current MC-based approaches to estimate the network MFD.

2.1. Variational theory

Daganzo (2005a,b) formulated the VT framework to solve complex and heterogeneous kinematic wave theory (KWT) problems. This framework determines the cumulative number of vehicles $N(x,t)$ that have passed the location x of a road by time t, for given boundary conditions. It requires the definition of fundamental diagram(s) (FD), i.e. $q(\kappa)$, network topology including signal control settings, bottlenecks (moving or stationary), as well as boundary data N_B along a curve B. The density is denoted by κ. The FD is characterized by the free-flow speed u, the backward wave speed w, and the link jam density κ_{max}. The capacity q_{max} is derived at the optimal density κ_{opt}, i.e. $q_{max} = q(\kappa_{opt})$.

To determine the cumulative count N at a point $P(x,t)$, we need to find the valid path $p \in \mathcal{P}$ starting at the boundary B, ending at P, and leading to the minimal increase of N over p. A path p is denoted as valid if its slope lies between the extremal speeds $v \in [w,u]$ where w has a negative value. Based on the slope of the path, costs z_p are defined. These costs refer to the maximum number of vehicles that can pass a moving observer traveling along the path p. Given the set of paths $\mathcal{P}$, as well as the associated costs z_p and boundary values $N_{B,p}$ associated with each path p, the cumulative count N_P can be found by a shortest path search as follows:

$$N_P = \min_{p \in \mathcal{P}} \{N_{B,p} + z_p\}. \tag{1}$$

By defining a numerical grid representing the solution space, we can find the cumulative count N based on eq.(1). Many possibilities for building problem-specific numerical grids exist (Daganzo and Menendez, 2005). One option for signalized corridors is to define the temporal and spatial distances as Δt and Δx, respectively. Then, the points on this grid are connected with horizontal and slanted edges which slopes correspond to $v \in \{w, 0, u\}$. The cumulative number $N(x,t)$ where (x,t) are points on the numerical grid, can then be found as follows:

$$\begin{aligned} N(x,t) = \min(&N(x-\Delta x, t-\Delta t), \\ &N(x+\Delta x, t-\theta\Delta t) + \Delta x \kappa_{max}, \\ &N(x, t-\Delta t) + \beta). \end{aligned} \tag{2}$$

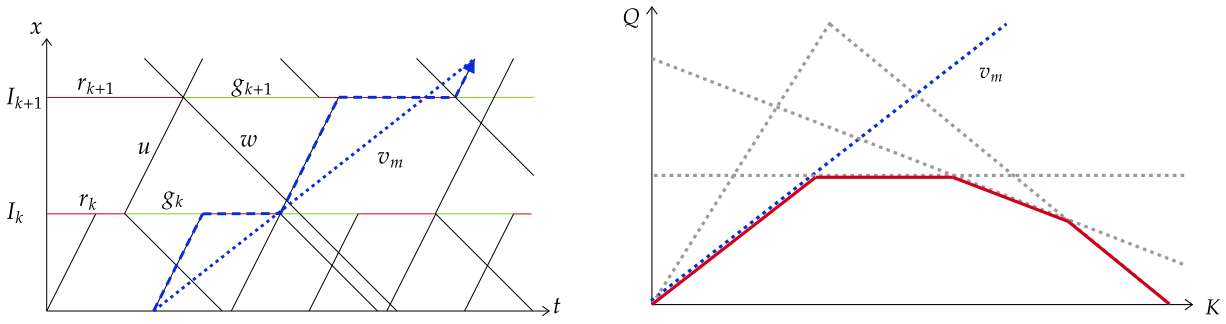
where $\theta = |\frac{u}{w}|$ is the absolute value of the ratio of the free-flow and the backward wave speed. If θ is an integer, the equation represents a lopsided numerical grid which facilitates the application of eq.(2), as nodes in the graph align horizontally. In cases where θ is not an integer, a shear transformation (e.g. Laval and Chilukuri, 2016) can be applied to transform a triangular FD such that θ becomes an integer. Otherwise, although possible, the construction of the network becomes more cumbersome. The parameter β accounts for any capacity constraints such as traffic signals. More specifically, if $\beta = 0$ the capacity equals zero, which can for example, represent the red phase of a traffic signal. Green phases can be modelled by setting $\beta = q_{max}$. For more details, please refer to Daganzo and Menendez (2005); Leclercq and Paipuri (2019); Tilg et al. (2021).

2.2. Original method of cuts

Based on the concept of VT, Daganzo and Geroliminis (2008) introduced the MC to analytically approximate the MFD for regular corridors with uniform link FD, block lengths, and signal control characteristics across all intersections. Leclercq and Geroliminis (2013) further extended the methodology to consider heterogeneous topologies, i.e. corridors with varying block lengths and signal settings. Laval and Castrillón (2015) integrated the MC into a stochastic context to be able to apply it to long corridors with varying

parameters. Daganzo and Lehe (2016) formulated a linear program based on the MC to further simplify the estimation of the MFD for corridors. Dakic et al. (2020) proposed a stochastic shortest path algorithm to model bi-modal interactions and their stochasticity along the corridor. A fundamental assumption of the MC and its extensions is the existence of stationary traffic states. In the following, we provide a brief summary of the current state of the art, and how to apply the MC for a simple example. We refer interested readers to the original papers for further details.

MC's required parameters are the number of links in the corridor, the number of lanes, the length of each link, the signal parameters, and the FD. Similar to VT, a numerical grid is defined based on these parameters, i.e. the so-called global variational graph or *hyperlink*. However, the graph is not discretized with Δt and Δx, but consists of horizontal edges only at the location of intersections I_k referring to both green and red phases, g and r. The index k describes a link and its downstream intersection along a specific corridor. The horizontal edges are connected with slanted edges, which slopes are either equal to the free-flow speed u, or the backward wave speed w. The slanted edges start at the bottleneck termini and end where they reach the next horizontal edge representing a red phase. As in VT, costs are associated with each edge and relate to the maximum number of vehicles that would pass a moving observer traveling along such an edge. Figure 1a shows an example for a hyperlink corresponding to a corridor with two intersections I_k and I_{k+1}. The horizontal edges are defined by green and red phases g_k, g_{k+1}, r_k and r_{k+1}.



a) Example of the variational graph including the path of observer m.

b) MFD defined by a family of cuts, including the one corresponding to observer m.

Figure 1: Variational graph including a moving observer's path and its translation to the MFD.

Based on the hyperlink, the MFD for the given corridor can be estimated as follows. The flow q_P at a certain point $P(x, t_0)$ approaches a location-independent limit for the case of a steady state given a certain initial density. Daganzo and Geroliminis (2008) described this flow as $q = N_P/t, t = t_0 \to \infty$. The upper bound for the flow in the corridor can be derived as follows:

$$Q(K) = \inf_{v_m} \{K v_m + R(v_m)\}, \tag{3}$$

where $Q(K)$ is the MFD, K is the average density in the corridor, and v_m is the average speed of a moving observer m. $R(v_m) = Z_m/t, t = t_0 \to \infty$ is the maximum flow that passes, on average, the moving observer m. The parameter Z_m is the sum of all costs along the path of moving observer m. This equation corresponds to a straight curve in the (Q, K) - plane and is referred to as a *cut*.

Figure 1 illustrates the procedure for a single cut as an example. The blue dashed line corresponds to the path of moving observer m. The costs Z_m are zero since only edges representing red phases and slanted edges with $v = u$ are part of the moving observer's route. The corresponding costs z of these edges are zero, as no vehicles can pass the moving observer traveling along them. The average speed v_m is displayed as a dotted blue curve. In Figure 1b the corresponding cut is shown as a blue dotted line again. Repeating this procedure, and thus evaluating eq.(3) for different K and moving observers with different average speeds

v_m leads to a family of cuts, shown as the grey dotted lines in Figure 1b. These cuts constrain the average flow Q along with the evaluated range of K. The lower envelope of all cuts is the corridor MFD, $Q(K)$.

2.3. Extension attempts to the network level

The resulting MFD from the MC is independent of the corridor demand since a single corridor without any turning flows only involves one OD pair. Therefore, the demand in this context only refers to the loading, and the MFD covers all stationary loading levels. From the network perspective, however, demand distribution in the form of path flows can lead to heterogeneous link flows which have a strong influence on the shape of the realized MFD. Applying MC to networks requires a reduction of the network to a single corridor. This inevitably introduces a bias with respect to the actual demand distribution and supply conditions.

Daganzo and Geroliminis (2008) originally proposed to identify a *representative* corridor of the network and apply the MC to find the idealized network MFD. This represents the already mentioned reduction of the network's complexity to a single corridor. In Girault et al. (2016), the authors reduced a bidirectional grid network to four corridors representing each cardinal direction. They applied the MC to each corridor and subsequently averaged the resulting corridor MFDs. This approach neglects some of the inter-dependencies such as spillback propagation between corridors. Moreover, the network reduction implicitly assumes that the most constraining intersection is saturated when the capacity is reached. However, this is not necessarily the case when turning ratios are considered, as some vehicles might turn before such an intersection, and thus reducing the maximum throughput.

Some initial attempts have been made to incorporate the effects of turning flows into MC. Geroliminis and Boyacı (2012) modeled the effects of inflows for the estimation of a route-specific MFD, although they did not attempt to estimate network MFDs. They modeled inflows as bottlenecks and did not thoroughly incorporate them. Therefore, their resulting MFD describes traffic dynamics for a certain route and not for the full corridor. Recall that we defined a corridor as an ordered sequence of links. In distinction to that, we define a route as an ordered sequence of links a vehicle travels from its origin to its destination. Thus, certain routes might correspond to a corridor, others might not. Xu et al. (2020) extended the stochastic model of Laval and Castrillón (2015) to account for turning flows. The authors proposed to estimate the effects of turning flows as a Markov process based on a two-ring model. They modeled turning flows with a global ratio which was kept constant across intersections in the network. Therefore, they again simplified the network to a single corridor. Aghamohammadi and Laval (2021) proposed a maximum likelihood approach for the method of Laval and Castrillón (2015) to improve the estimation of the MFD for networks. While their results are promising, the methodology still builds upon a single corridor. Other analytical approaches (Gan et al., 2017; Jin et al., 2013b) are also based on the two-ring model and are therefore limited to very regular and simplified network topologies.

3. General methodology for the extended method of cuts

In an attempt to overcome the mentioned drawbacks of existing methods, and to address this research gap, we extend the MC in order to apply it to realistic networks with varying turning flows at each intersection. Thereby, we include network-wide traffic dynamics, do not reduce the entire network to a single corridor, and thus consider its original topology. This results in a general framework to estimate the realized MFD for realistic urban networks. Hereafter, we refer to this framework as 'nMC'.

Figure 2 describes the overall framework which consists of three steps which are explained in the following:

- *Step 0* initializes the problem by defining supply and required demand characteristics. This consists of the definition of the network topology and the signal control settings. Moreover, we define the turning ratios at each intersection. Note that this implies an indirect definition of routes, and thus reflects the observed spatial demand pattern for a specific case. These parameters describe the physical road network $\mathcal{N}$ composed by a set of directed links and intersections, as well as the demand that exists for that network.

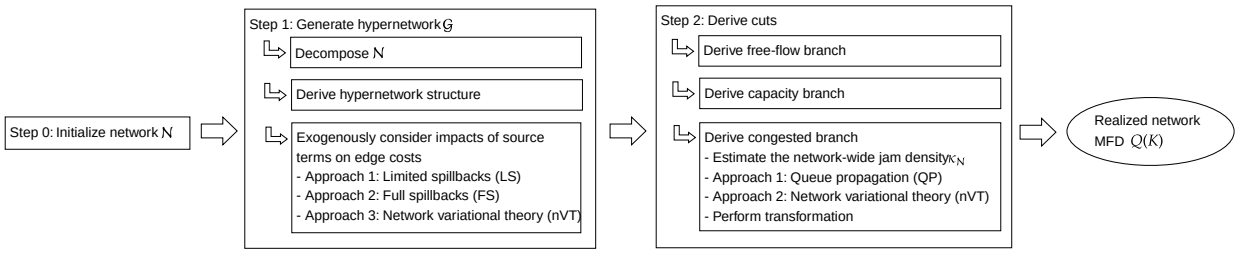


Figure 2: Flowchart of the proposed framework, the nMC.

- *Step 1* transforms $\mathcal{N}$ into a hypernetwork $\mathcal{G}$ which is the basis for deriving the MFD. Recall that a hyperlink represents a physical corridor in the original MC (see Figure 1a). Analogously, a hypernetwork represents the physical road network in time and space and can be seen as a multi-dimensional variational graph. To generate $\mathcal{G}$, we decompose $\mathcal{N}$ into a set of corridors, define the structure of the hypernetwork $\mathcal{G}$, and exogenously derive the costs for horizontal edges in $\mathcal{G}$. For this, we propose three different approaches with varying modeling complexity, computational cost, and estimation accuracy. These approaches are labeled as 'limited spillbacks' (LS), 'full spillbacks' (FS), and 'network variational theory' (nVT) which is based on an extension of VT to networks (Tilg et al., 2021), and further explained in Section 4.

- *Step 2* derives the cuts based on which the free-flow branch, the capacity, as well as the congested branch of the MFD are estimated. In the spirit of the original MC (see Figure 1b), we send moving observers through $\mathcal{G}$ in order to estimate stationary traffic states. This step includes the approximation of the network-wide jam density and the utilization of symmetries in the propagation of free-flow and congested traffic states across the network. Again, two different approaches are proposed, namely the queue propagation (QP) and the network variational theory (nVT) approach, which are explained in Section 5.3.

The nMC enables one to estimate the realized MFD for realistic urban road networks. It accounts for turning flows and the corresponding interdependencies of traffic states across the physical road network. Compared to simulation-based approaches, this procedure describes the observed macroscopic traffic behavior in urban networks in a fast and efficient manner. We propose several approaches to (i) allow one to flexibly trade-off modeling complexity and computational burden against estimation accuracy depending on the use case at hand, and (ii) shed light on the impact of some of the modeling assumptions. Below we describe steps 1 and 2 mentioned above in more detail. Step 0 is trivial as it refers to the problem initialization and requires no further explanation.

4. Generation of a hypernetwork

The correct representation of $\mathcal{N}$ as a hypernetwork $\mathcal{G}$ is key to our framework. It allows us to derive the cuts similar to the original MC (see Section 2.2) while accounting for source terms and network-wide spillback propagation. These cuts are the basis to find the realized network MFD. In the following, we list the main assumptions of our framework:

- The focus lies on unimodal networks. That is, we do not consider any modes apart from private vehicles. Nevertheless, there are multimodal extensions for the original MC which can potentially be integrated into our framework. We consider this, however, out of scope and leave it for future work.

- We only consider signalized intersections without modeling conflicting streams explicitly. However, conflicting traffic streams can be approximated by reducing the average capacity of corresponding intersection approaches accordingly, e.g. based on headway distributions. Related to that, other

intersection types such as 4-way stops can be included if capacity estimates for the corresponding intersection approaches exist.

- We assume that vehicles follow a FIFO discipline on all links and at diverges (Newell, 1993). Turning lanes can be introduced by splitting a link, and defining the turning ratios at the downstream end of each lane accordingly. For example, the turning ratio at the downstream end of a right-turning lane would be 1, while it would be 0 at the downstream end of the straight-going lane.
- For the sake of simplicity, our framework only applies to cases where vehicles either remain on the main corridor or change to a single adjacent corridor. More complicated topologies, e.g. intersections with multiple turning options such as left-, right-, and U-turns are left for future work. While the general framework is able to account for multiple turning options since the basic functionality requires no substantial modification, it would at least necessitate the specification of turning ratios for each option and the coupling of multiple corridors at intersections.
- While we allow turning ratios to vary between intersections, we assume them to be constant across time.
- We assume that stationary states exist for any spatial demand pattern if the temporal change is slow. Note that this is an assumption inherited from the original MC and further applies to the general notion of the MFD.

We generate a hypernetwork $\mathcal{G}$ representing a physical road network $\mathcal{N}$ including signal settings and turning flows. First, $\mathcal{N}$ is decomposed into a set of corridors $\mathcal{C}$. Second, the structure of a hypernetwork $\mathcal{G}$ is defined. As Laval et al. (2017) showed, VT cannot endogenously account for source terms. Consequently, they are considered as an exogenous input as the overall framework is based on the concept of VT. Note that only the turning ratio is an exogenous input. The resulting costs of horizontal edges in the hypernetwork, which represent maximum flows, are endogenously derived depending on the method applied, i.e. LS, FS, or nVT. This concludes the generation of the hypernetwork $\mathcal{G}$. Note that the resulting maximum flows in $\mathcal{G}$ represent the network traffic at the capacity level.

4.1. Network decomposition

The decomposition of the physical road network $\mathcal{N}$ into a set of non-overlapping corridors facilitates the generation of $\mathcal{G}$. Furthermore, this set of corridors is an input to the MFD estimation (see Section 5.1). Tilg et al. (2021) describe such a decomposition in the context of VT. Following their methodology is advantageous for our framework, as their proposed network VT solution method, the 'nVT', can be utilized to derive the hypernetwork (see Section 4.3.3). We provide a brief summary of the decomposition method hereafter.

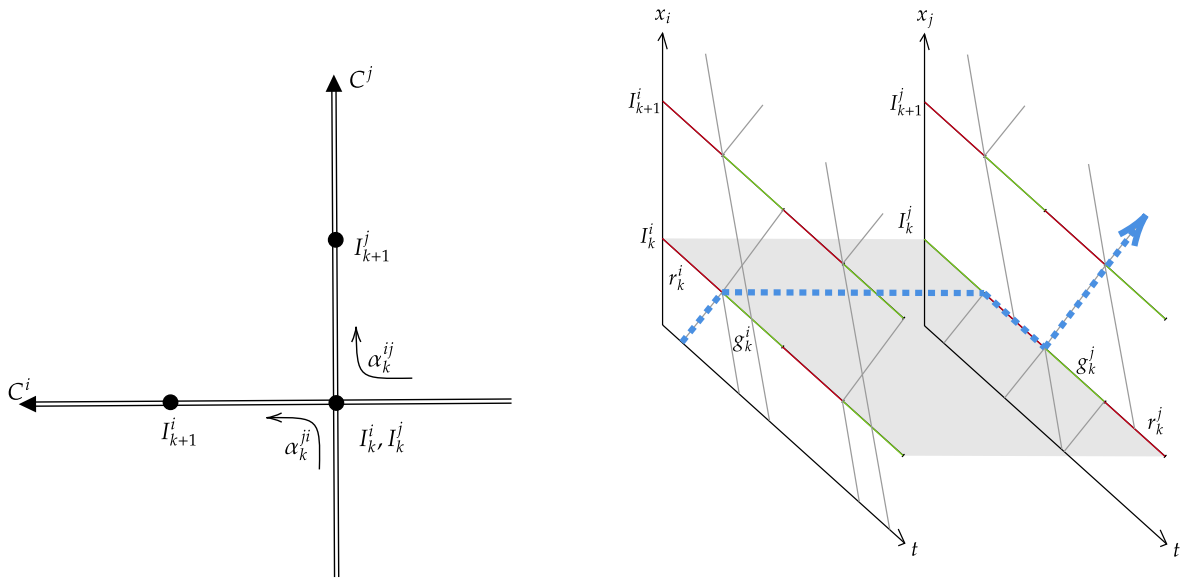
The initialization of the problem (Step 0) defines a physical road network $\mathcal{N}$ consisting of intersections $I \in \mathcal{I}$ and links $L \in \mathcal{L}$. Moreover, it includes the specification of signal control settings, i.e. red r and green times g, cycle lengths c and offsets o, as well as turning ratios $\alpha \in A$ at each intersection I. All information from $\mathcal{N}$ including topological features, turning ratios α, and control settings for $I \in \mathcal{I}$ have to be retained when the network is decomposed (Step 1). Note that the turning ratios are an exogenous input. The two main requirements for decomposing the network $\mathcal{N}$ are:

1. The set $\mathcal{C}$ includes all links $L \in \mathcal{L}$.
2. Each link L exists only once in $\mathcal{C}$, i.e. the corridors do not have any overlapping segments.

For small toy networks, we can determine the set $\mathcal{C}$ manually from $\mathcal{N}$. More general, realistic networks can be decomposed according to the actual layout of the roads. This includes arterials, avenues, and streets. Each road can be represented as a corridor C and as such be incorporated in $\mathcal{G}$. This will always satisfy both conditions mentioned above. Requirement 2 does not preclude the potential existence of circular routes in the network. Note that the set $\mathcal{C}$ might affect the MFD to a certain extent. This aspect will be treated in Section 5.1 and analyzed in Section 6.2.3.

4.2. Structure of the hypernetwork

The set of corridors $\mathcal{C}$ allows us to generate a hypernetwork $\mathcal{G}$ representing the maximum traffic flows in the entire network $\mathcal{N}$. This includes the generation of hyperlinks for each corridor along with intersections and corresponding turning ratios α according to the physical network topology. Hence, in contrast to the original MC where only single corridors could be modeled, $\mathcal{G}$ represents networks where multiple edges downstream and upstream of an intersection might exist. This expansion of the concept of the corridor-specific hyperlink to a network-wide hypernetwork $\mathcal{G}$ represents one of the major cornerstones of our framework.



a) Physical road network $\mathcal{N}$. **b)** Hypernetwork $\mathcal{G}$.

Figure 3: Example of a network with two corridors, and a corresponding hypernetwork excerpt.

Figure 3a illustrates an example of a decomposed network based on which we further introduce our notation. It shows two corridors C connected at one intersection I. The indices denoted as superscripts i, j refer to a specific corridor. The index k as subscript refers to an intersection and the link directly upstream of it. The subscript is unique in combination with the corridor-related index. For example, the intersection I^i_k is the k^{th} intersection on corridor C^i. Similar to that, subscripts of maximum flows $\tilde{q}$ indicate intersections, and superscripts the corridor to which the flow refers. For example, the variable $\tilde{q}^i_k$ describes the maximum cycle-based average flow on corridor C^i upstream of intersection I^i_k. By maximum flow, we do not refer to the capacity, but the maximum flow that can be sustained on average during one cycle as a function of both the demand and the supply. The indices of turning ratios α denote the origin and destination corridor of the corresponding flows at a given intersection. For example, the turning ratio α^{ij}_k describes the ratio of the flow aiming to change from C^i to C^j at intersection $I^i_k = I^j_k$.

The network $\mathcal{N}$ can be translated into a hypernetwork $\mathcal{G}$, see Figure 3b. For a single corridor, the graph $\mathcal{G}$ consists of horizontal and slanted edges, the latter with slopes equal to u or w. The duration of green and red phases is denoted with g and r, respectively. Subscripts denote the intersection to which these phases refer, while superscripts refer to the corresponding corridors. For example, g^i_k is the duration of the green phase at intersection I^i_k on corridor C^i. The horizontal edges have associated costs z equal to the maximum flow $\tilde{q}$ passing a moving observer traveling along the respective edge. In the original MC, this flow is zero during red times, i.e. $\tilde{q} = 0$, and equal to the link capacity during green times, i.e. $\tilde{q} = q_{max}$ (see Section 2.2). However, the occurrence of inflows and outflows at intersections might affect the observed

maximum flow, and consequently the related costs z. For example, the maximum flow $\tilde{q}_k^i$ during a green phase at an intersection I_k^i might be reduced when a net outflow occurs at the upstream intersection I_{k-1}^i. Also, if the inflow at an intersection I_k^i exceeds the capacity of the downstream intersection I_{k+1}^i, spillbacks occur after a while, potentially affecting $\tilde{q}_k^i$.

These examples highlight the importance of the inclusion of effects of turning flows and indicate that such impacts can be represented via $\tilde{q}$. Thus, we aim at modifying the costs z of horizontal edges to represent the effects of source terms. As explained before, these effects need to be considered exogenously. Therefore, we propose three different methods to exogenously determine the effects of source terms on the maximum flows $\tilde{q}$ at intersections, and adapt the costs of related edges in $\mathcal{G}$.

4.3. Exogenous consideration of source terms

To complete the generation of the hypernetwork $\mathcal{G}$, we exogenously account for the effects of source terms at intersections which represent inflows and outflows. In particular, we modify the costs of horizontal edges in $\mathcal{G}$ representing green phases. Thereby, we include both demand- and supply-related aspects, such as undersaturated intersection approaches and the occurrence of spillbacks coming from the downstream intersection. Note that while the edge costs in $\mathcal{G}$ shall represent maximum flows, similar to the original MC, the hypernetwork allows deriving an upper bound for flows during the loading and unloading phase of the network as well. For this purpose, we present three different approaches of increasing modeling complexity and computational cost in exchange for higher accuracy:

1. *Approach 1 - Limited spillbacks (LS)*: We do not consider spillback propagation across corridors and focus on its effects at the link level. Its field of application is for use cases where no substantial network-wide spillbacks occur.

2. *Approach 2 - Full spillbacks (FS)*: This approach builds on the first one, but it also includes spillback propagation throughout the network. Therefore, it also applies to use cases with substantial network-wide spillbacks.

3. *Approach 3 - Network variational theory (nVT)*: Based on the nVT proposed by Tilg et al. (2021), this approach allows to accurately consider spillback propagation according to the KWT. While being more exact and offering a larger field of applications, this approach requires significantly more modeling and computational efforts than the other two.

In the following subsections, we first introduce the mathematical framework for the approximate approaches FS and LS. Then, we describe the utilization of nVT to complete the hypernetwork and highlight the synergies between nVT and nMC.

4.3.1. Approach 1 - Limited spillbacks (LS)

This approach applies to use cases where a fast and simple MFD estimation is essential, and spillbacks are limited to individual links. For such use cases, we propose a solution method that provides a reasonably accurate estimation of spillback effects at the link level to derive the costs z for edges representing green phases in $\mathcal{G}$. The approximation of such effects is based on analytical calculations, making it fast and efficient.

This section formulates an optimization problem with non-linear constraints considering the interdependencies of maximum flows $\tilde{q}$ between different intersections $I \in \mathcal{I}$ in the network $\mathcal{N}$. First, we consider demand-related aspects and the network-wide conservation of flows. Second, we account for supply reductions due to spillbacks.

Assumptions. For the sake of clarity, we list the most important assumptions for this approach below:

- The link FD is triangular.
- Traffic signals at all intersections I have a common cycle length.

- We only consider the effects of congestion at the link level and no spillback propagation across intersections.

Demand. The demand-induced reductions of maximum flows $\tilde{q}_{k,d}$ due to source terms need to be represented in $\mathcal{G}$. This is of particular importance for the estimation of the capacity branch of the MFD. The index d represents that this maximum flow is constrained due to demand. Note that such effects were implicitly included in the original MC, e.g. undersaturated intersections due to upstream bottlenecks. However, source terms eliminate the possibility for such an implicit consideration. The following example lets us derive the key formulation to represent demand-related effects in our model.

Consider a case where net outflows at the intersection I^i_{k-1} lead to flows at the downstream link L^i_k and intersection I^i_k that are lower than the link capacity q_{max}. This is illustrated in an excerpt of the time-space diagram corresponding to C^i in Figure 4.

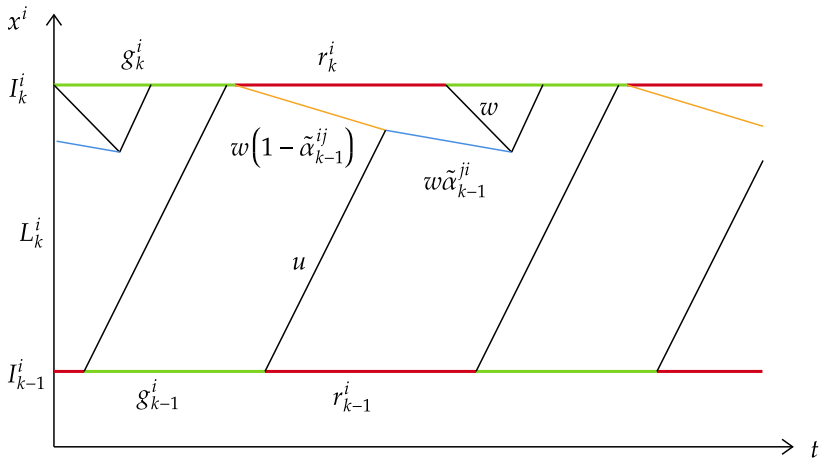


Figure 4: The influence of undersaturated links.

The figure shows a link L^i_k between two intersections, I^i_{k-1} and I^i_k, on corridor C^i. The red and green lines represent the signal phases at both intersections. Shock waves with the slopes u and w are shown as black lines. We assume turning ratios such that the total flow during g^i_k is undersaturated, i.e. $\tilde{q}^i_k < q_{max}(g^i_k/c^i_k)$ where c^i_k denotes the cycle length at I^i_k. The shock waves illustrating the propagation of the queue from I_k are shown as orange and blue lines. The orange shock wave results from flows staying on C^i, i.e. retaining flows, which join the queue. Thus, the queue grows with speed $w(1-\tilde{\alpha}^{ij}_{k-1})$. The blue shock wave results from turning flows from C^j which join the queue. Thus, the queue grows with $w\tilde{\alpha}^{ji}_{k-1}$. Assuming a triangular FD, the parameter $\tilde{\alpha}$ can be derived as shown in appendix C and written as:

$$\tilde{\alpha} = \alpha \frac{u}{u + w(1-\alpha)}. \tag{4}$$

This example shows that the maximum flow $\tilde{q}^i_k$ can be undersaturated since the flow which discharges from the queue does not last for the entire green phase g^i_k. Accounting for intersection-specific signal settings, we can describe the demand $\tilde{q}^i_{k,d}$ for an entire cycle as:

$$\tilde{q}^i_{k,d} = \tilde{q}^i_{k-1}(1-\alpha^{ij}_{k-1})\frac{g^i_{k-1}}{g^i_k} + \tilde{q}^j_{k-1}\alpha^{ji}_{k-1}\frac{r^i_{k-1}}{g^i_k}. \tag{5}$$

This equation describes the maximum flow $\tilde{q}^i_{k,d}$ which can occur considering demand only. Note that if a net inflow exists, i.e. $\tilde{q}^i_{k,d} > q_{max}$, the demand exceeds the link capacity q_{max}. To address such cases, the required supply-related constraints are introduced in the following.

Supply. Next, we model the effects of spillbacks on the maximum flows, i.e. $\tilde{q}_{k,s}$, denoted by the additional subscript s. Note that such effects are implicitly considered in the original MC as flow conservation is kept. However, the occurrence of inflows and outflows at intersections might lead to a violation of flow conservation at the corridor level even though it is still satisfied at the network level. Therefore, in order to consider the effects of spillbacks, we propose to find the most constraining spillback duration σ_k exogenously as explained in the following. Note that the assumption of a FIFO diverge implies that a spillback occurring on a single link downstream of I_k affects all other outgoing links of I_k, too, if there is turning demand. Below, we derive the spillback duration and incorporate that into $\tilde{q}_{k,s}$ by means of an example.

Recall that flow conservation applies at the link level despite the existence of source terms as so far we have assumed those source terms are located at intersections. Thus, we can derive the duration σ_k during which a spillback in a link affects the upstream intersection I_k by applying traditional concepts of traffic flow theory.

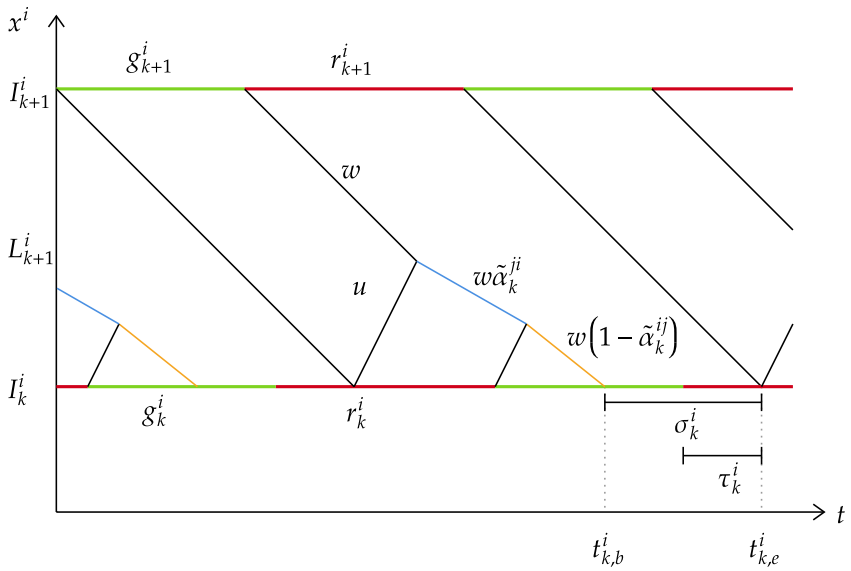


Figure 5: Spillback occurrence in a time-space diagram.

Figure 5 displays an example of such a spillback occurrence for the stationary case. It shows an excerpt of a time-space diagram for a link L^i_{k+1} between two intersections, I^i_k and I^i_{k+1}. The red and green lines represent the signal phases at both intersections. Shock waves with the slopes u and w are shown as black lines. For this case, we assume turning ratios such that the total flow during g^i_{k+1} is oversaturated, i.e. the demand exceeds the intersection capacity $\tilde{q}^i_{k+1,d} > q_{max}(g^i_{k+1}/c^i_{k+1})$. The queue grows with speed w while saturation flows reach the queue. Once vehicles from upstream of I^i_k join the queue, its propagation speed changes. This is highlighted by the blue and orange curves in the figure. The blue curve represents a queue growth due to inflows from the adjacent corridor (i.e. turning vehicles coming from corridor C^j to C^i), similar to Figure 4. The orange one represents a queue growth due to the straight flows (i.e. vehicles staying on C^i). Once the orange shock wave reaches I^i_k the flow of the corresponding upstream link is blocked, and the queue spills over to this link. This marks the beginning of the spillback duration, $t^i_{k,b}$. The spillback ends once the queue at I^i_k dissolves again at the time instant $t^i_{k,e}$. The total duration of the spillback impact is denoted as σ^i_k. This blockage affects the capacity of upstream links and the demand for downstream links on all adjacent corridors. Note that $t^i_{k,e}$ occurs during a red phase in the shown example, and the following considerations refer to that case. Nevertheless, the concept is valid for $t^i_{k,e}$ occurring during a green phase as well. The equations can be derived analogously.

First, we calculate the spillback impact ending time $t^i_{k,e}$ which can easily be found from the time-space

diagram:

$$t^i_{k,e} = r^i_{k+1} + \frac{l^i_k}{w}. \tag{6}$$

Then, we derive the parameter τ^i_k, which denotes the time between the beginning of the active phase and $t^i_{k,e}$. Considering the offset o^i_k, we derive the parameter as follows:

$$\tau^i_k = mod(t^i_{k,e}, c^i_k) - o^i_k. \tag{7}$$

where *mod* is the modulo operator which calculates the remainder of the division of $t^i_{k,e}$ and c^i_k.

To facilitate the estimation of σ^i_k, we shift our perspective from the time-space diagram to a cumulative plot, as illustrated in Figure 6. We then resort to the illustrative and well-established technique of comparing cumulative counts N at different locations introduced as the 3-detector problem by Newell (1993). As described in Tilg et al. (2021), a discontinuity in the Moskowitz function N occurs at intersections with source terms. Thus, it is important to note that the following explanations always refer to a link segment where the flow conservation applies and no discontinuity exists. Correspondingly, we derive the cumulative count N^i_{k+1} at the downstream intersection I^i_{k+1}, and $N^i_{k+\delta x}$ at an infinitesimally small distance downstream of I^i_k, denoted as $I^i_{k+\delta x}$. Furthermore, we assume the spillback duration at I^i_k and $I^i_{k+\delta x}$ are the same, $\sigma^i_k = \sigma^i_{k+\delta x}$. The count N^i_{k+1} represents the cumulative curve corresponding to the capacity, and $N^i_{k+\delta x}$ the demand. Since we are interested in deriving the spillback impact duration at $I^i_{k+\delta x}$, we shift N^i_{k+1} by l^i_k/w in the temporal dimension, and $l^i_k\,\kappa_{max}$ in the N-dimension, where κ_{max} is the link jam density. Figure 6

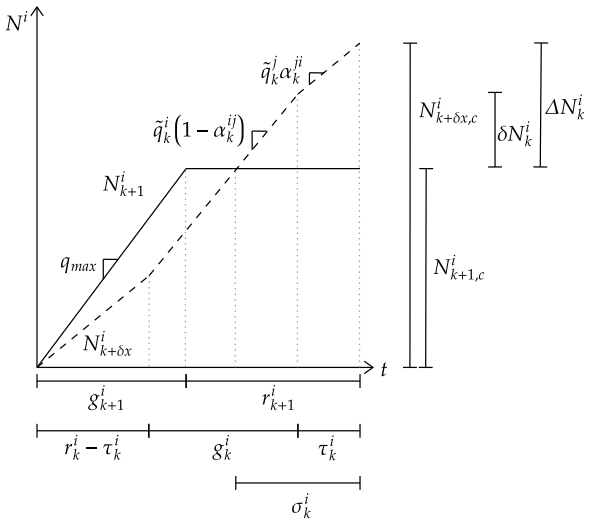


Figure 6: Spillback occurrence in a cumulative plot.

illustrates the shifted cumulative curves, where N^i_{k+1} and $N^i_{k+\delta x}$ are depicted as a solid and as a dashed curve, respectively. Both N-curves refer to a single cycle length c. This is sufficient, as we assume a common cycle length for all intersections, and analyze stationary states which occur periodically with such cycle length. The maximum cumulative count during the respective cycle is denoted with the subscript c. Based on both cumulative curves, we can derive the spillback impact duration σ^i_k as follows.

The cumulative count $N^i_{k+1,c}$ is equal to the product of link capacity q_{max} and the green phase at the downstream intersection I^i_{k+1}:

$$N^i_{k+1,c} = g^i_{k+1} q_{max}. \tag{8}$$

Note that this represents the essential assumption of the LS approach, i.e. queues blocking I^i_{k+1} are not considered in deriving σ^i_k. This assumption is relaxed in the FS approach, see Section 4.3.2.

The cumulative count related to the demand $N^i_{k+\delta x,c}$ equals the sum of the inflows for one cycle. Building upon eq.(5), we can write:

$$N^i_{k+\delta x,c} = \tilde{q}^i_k(1-\alpha^{ij}_k)g^i_k + \tilde{q}^j_k\alpha^{ji}_k r^i_k. \tag{9}$$

The comparison of both N-curves lets us determine the spillback duration σ^i_k at intersection I^i_k. More specifically, we can determine the actual cumulative plot at I^i_k as the lower envelope of both N-curves. The spillback begins once the flow is constrained by the cumulative curve at the downstream intersection, I^i_{k+1}. During that period, the flow is zero. We determine the difference between the demand and the capacity accounting for the spillback:

$$\Delta N^i_k = N^i_{k+\delta x,c} - N^i_{k+1,c} = \tilde{q}^i(1-\alpha^{ij})g^i_k + \tilde{q}^j\alpha^{ji}r^i_k - g^i_{k+1}q_{max}. \tag{10}$$

Next, we determine the part of ΔN^i_k which remains after subtracting the blocked vehicles during τ^i_k, denoted as δN^i_k. This simplifies the notation to derive σ^i_k.

$$\delta N^i_k = \Delta N^i_k - \tau^i_k\tilde{q}^j_k\alpha^{ji}_k. \tag{11}$$

Note that a positive δN^i_k corresponds to a case where $\sigma^i_k > \tau^i_k$, i.e. the spillback spans both phases. This case is illustrated in Figure 6. On the other hand, a negative δN^i_k corresponds to the case where $\sigma^i_k < \tau^i_k$, i.e. where the spillback only occurs during the red phase.

To derive the total spillback impact duration σ^i_k, we distinguish between the two cases referring to the sign of δN^i_k:

$$\sigma^i_k = \begin{cases} \tau^i_k + \frac{\delta N^i_k}{\tilde{q}^i_k(1-\alpha^{ij})} & \text{, if } \delta N^i_k > 0 \\ \tau^i_k + \frac{\delta N^i_k}{\tilde{q}^j_k\alpha^{ji}} & \text{, otherwise.} \end{cases} \tag{12}$$

With this equation, one can calculate the spillback duration due to a queue on C^i, which can be relevant for both corridors C^i and C^j at the intersection I^i_k. To approximately account for a FIFO diverge behavior, we propose to calculate ΔN_k for all outgoing links, and subsequently consider the maximum ΔN_k to compute σ_k. During a stationary state, the inflow and outflow in a link are balanced by the spillback duration. Thus, the maximum ΔN_k is required to ensure all flows are stationary. Note that this is an approximation, as we neglect the fact that spillbacks can also affect inflows on adjacent links. However, this approach does not aim at an extremely precise estimation of σ_k, but at estimating the spillback duration sufficiently precise for an accurate MFD estimation at the network level.

The total spillback duration σ^i_k is further divided into the corridor-specific blockage times, which are denoted as $\sigma^{i\rightarrow i}_k$ and $\sigma^{i\rightarrow j}_k$. These depend on the length of σ^i_k, i.e. whether the spillback lasts only during one signal phase, or spans two phases. Note that σ^i_k can occur during two phases at a maximum, as we assume a common cycle length for all intersections and therefore the theoretically maximum blockage duration equals to the red phase at the downstream intersection which cannot exceed a cycle length. As we assume that $t^i_{k,e}$ occurs during a red phase, $\sigma^{i\rightarrow j}_k$ is either τ^i_k or σ^i_k, if $\sigma^i_k < \tau^i_k$. Similar to that, $\sigma^{i\rightarrow i}_k$ is the difference of the total spillback duration and τ^i_k, or equal to zero in $\sigma^i_k < \tau^i_k$. To cover both cases, we write:

$$\begin{aligned} \sigma^{i\rightarrow i}_k &= \max(0, \sigma^i_k - \tau^i_k), \\ \sigma^{i\rightarrow j}_k &= \min(\tau^i_k, \sigma^i_k). \end{aligned} \tag{13}$$

It is left to the interested reader to verify that the framework can be derived analogously for the other case, where $t^i_{k,e}$ occurs during an active green phase.

Eventually, the corridor-specific spillback impacts allow us to derive the respective capacity constraints. Therefore, we can formulate the reduction of maximum flows on corridor C^i and corridor C^j due to spillbacks on corridor C^i:

$$\begin{aligned} \tilde{q}^{i\rightarrow i}_{k,s} &= q_{max}(1 - \frac{\sigma^{i\rightarrow i}_k}{g^i_k}), \\ \tilde{q}^{i\rightarrow j}_{k,s} &= q_{max}(1 - \frac{\sigma^{i\rightarrow j}_k}{r^i_k}). \end{aligned} \tag{14}$$

Maximum flows. The maximum flow on each link is then the minimum of the demand (eq.(5)) and the supply (eq.(14)):

$$\tilde{q}_k^i = \min\left(\tilde{q}_{k,d}^i, \tilde{q}_{k,s}^{i\rightarrow i}, \tilde{q}_{k,s}^{j\rightarrow i}\right). \tag{15}$$

This equation is formulated for each link and results in a non-linear system of equations. Its structure assembles widely applied concepts such as the CTM or VT. For network entry and exit links, the demand and supply parts have to be treated separately. For example, there are no demand-related restrictions for links on which flows enter the network. Similarly, there are no capacity constraints for links on which flows exit the network. These links represent origin and destination nodes within the network.

The maximum network-wide average flow does not necessarily result from maximum inflows at origin links. This is because internal bottlenecks might become active at lower inflows and the corresponding congestion can limit the network-wide maximum flows. The maximum average flow is reached when the inflows are as high as possible, but the effects of activated internal bottlenecks remain low. To treat this problem adequately, we formulate an optimization problem based on eq.(15). The capacity of a network is defined by the maximum stationary flows possible. We define the inflows at origin links as decision variables. Therefore, the equivalent of eq.(15) for origin links, i.e. the link index $k = 1$, can be written as follows:

$$\tilde{q}_1^i = \min\left(\lambda^i q_{max}, \tilde{q}_{1,s}^{i\rightarrow i}, \tilde{q}_{1,s}^{j\rightarrow i}\right). \tag{16}$$

where the parameter λ determines the share of q_{max} which is the demand at origin links. The overall optimization problem for all links in the network can be formulated as:

$$\max_{\lambda^i} \quad \sum_{i=1}^{n} \tilde{q}_k^i \tag{17a}$$

$$\text{s.t.} \quad 0 \leq \lambda^i \leq 1 \tag{17b}$$

since all constraints are effectively included in eq.(15).

In summary, this subsection proposes an approximate approach to derive the maximum flows $\tilde{q}_k^i$ at the intersection level. These are necessary to define the costs of horizontal edges in $\mathcal{G}$ representing green phases. The approach assumes a triangular fundamental diagram and common cycle lengths. The effects of spillbacks are only considered at the intersection level and not across entire corridors. Thus, we label it 'Limited spillbacks'. For use cases, where a more exact determination of the maximum flows $\tilde{q}_k^i$ is desired, we propose an alternative in the following subsection.

4.3.2. Approach 2 - Full spillbacks (FS)

The second approach is in large parts similar to the first one. The difference lies in the consideration of spillbacks at downstream intersections. For this purpose, we only need to adapt eq.(8) as explained in the following.

We modify the way to estimate the cumulative count $N_{k+1,c}$ representing the maximum number of vehicles passing the downstream intersection I_{k+1} during a cycle. In this approach, $N_{k+1,c}$ is equal to the product of the maximum flow $\tilde{q}_{k+1,s}^i$ at the downstream intersection I_{k+1} and the corresponding green phase:

$$N_{k+1,c}^i = g_{k+1}^i \tilde{q}_{k+1,s}^i. \tag{18}$$

Recall that the index s denotes that this flow can itself be impacted by spillbacks at the respective intersection. This dependency shows the recursive nature of this approach, and it allows us to model the propagation of spillbacks downstream of I_k^i. Nevertheless, the effects of any constraint are uniformly spread across the green time. This constitutes the main assumption of the FS approach.

The modification of the method to derive $N_{k+1,c}$ further affects the computation of ΔN_k^i. Based on eq.(10) and eq.(18) above, we write:

$$\Delta N_k^i = N_{k+\delta x,c}^i - N_{k+1,c}^i = \tilde{q}^i(1-\alpha^{ij})g_k^i + \tilde{q}^j\alpha^{ji}r_k^i - g_{k+1}^i\tilde{q}_{k+1,s}^i. \tag{19}$$

The difference in the demand and the capacity ΔN_k^i can then be put into eq.(11), and thus enables one to follow the procedure described for the previous approach to eventually calculate the corridor-specific spillback impacts. The FS approach enables one to consider congestion propagation throughout the network. Thus, we label this approach 'Full spillbacks'. However, this comes at the price of higher modeling complexity. More specifically, $\tilde{q}_{k+1,s}^i$ in eq.(18) is a variable in contrast to q_{max} in eq.(8) being a constant. Thus, the overall system of equations becomes more complex. However, the traffic dynamics are still simplified. To further increase the accuracy of modeling such dynamics, an appropriate KWT solution scheme shall be used. The application of such a scheme within the context of the proposed framework is presented in the following section.

4.3.3. Approach 3 - Network variational theory (nVT)

The 'nVT' as proposed by Tilg et al. (2021) enables one to describe the evolution of traffic states in a signalized network with precision and efficiency. Its advantage compared to other macroscopic traffic models is the ability to model complex intra-link bottlenecks. Its advantage compared to microscopic simulation models is the numerical efficiency, as demonstrated in Tilg et al. (2021). Moreover, nVT requires the same inputs as nMC, and builds upon a numerical grid similar to a hypernetwork. This leads to further synergies, as the hypernetwork $\mathcal{G}$ can be utilized for such a numerical grid. Nevertheless, compared to the other two approaches it requires the highest modeling and computational efforts to derive the costs for horizontal edges in $\mathcal{G}$.

Cumulative count. The numerical grid is defined by the time-step length Δt, the spatial step length $\Delta x = u\Delta t$, the length of each corridor $C \in \mathcal{C}$, and the length of the analysis period (see Section 2.1). The latter has to be chosen such that stationary traffic states are reached on each link $L \in \mathcal{L}$, since the MFD is based on such stationary traffic states. In other words, the total computation time of the numerical grid needs to be sufficiently long such that stationary states can be reached within this period. Lastly, turning ratios $\alpha \in A$ are input parameters. Note that this grid has some overlap with the hypernetwork $\mathcal{G}$ (e.g. at intersections) but is generally more fine-meshed as grid nodes also cover space within links.

The nVT finds the cumulative count N at each node $P(x,t)$ in the numerical grid. For a given node $P(x,t)$ within links, the classical VT formulation applies (see Section 2.1). Furthermore, Tilg et al. (2021) proposed to exogenously account for the effects of source terms at the points P which are exactly at, and right downstream of an intersection. The corresponding formulations are shown in the appendix B. These equations ensure the correct propagation of free-flow and congested traffic states across intersections despite the existing discontinuities in the Moskowitz surface at such locations. The result is the precise cumulative count N at each point in the numerical grid. For further details on the methodology, please refer to Tilg et al. (2021).

Edge costs. The cumulative count $N(x,t)$ found by nVT lets us now derive the costs for horizontal edges z_k^i in the hypernetwork $\mathcal{G}$. Recall that these costs correspond to the flow which would pass a moving observer traveling along such edges. Such moving observers always travel until the termination of a red phase in the original MC, which includes the entire green phase. Hence, it is sufficient to derive the cycle-specific average flows from the nVT solution to specify the costs of horizontal edges in $\mathcal{G}$. We measure the difference in N which occurs during the beginning and the end of green phases at all intersections to calculate the costs z. These costs consider both, the fact that intersections can be undersaturated due to net outflow at the upstream intersection, and spillbacks which block certain parts of the green phase. In both cases, the costs of the corresponding edge are effectively reduced. More formally, we can determine the costs z_k^i related to a green phase at intersection I_k^i as:

$$z_k^i = \frac{N_{k,s}^i - N_{k,e}^i}{g_k^i}. \tag{20}$$

where $N_{k,s}^i$ and $N_{k,e}^i$ denote the cumulative count at intersection I_k^i at the start and end of a green phase, respectively. Note that the numerator is always positive, as N is a monotonic increasing function in t.

Moreover, it will be necessary to derive the spillback duration from the Moskowitz surface as well. Again, such durations can be extracted from the nVT solution by analyzing the Moskowitz surface.

Optimization. Similar to the approximate approaches, we find the capacity state of the network with eq.(17). However, the maximum flows $\tilde{q}$ are obtained based on eq.(20), since the cost of horizontal edges in $\mathcal{G}$ corresponds to these flows.

5. Derivation of cuts

The previous section describes three approaches with different levels of modelling complexity to define the costs of horizontal edges in $\mathcal{G}$ representing green phases, which include the effects of source terms on network-wide traffic state evolution for the capacity state. Building upon $\mathcal{G}$, in this section, we propose a method to estimate the realized network MFD. Thereby, we apply methods from the original MC and utilize recent findings in MFD-related research. First, we estimate the free-flow branch by deriving cuts from $\mathcal{G}$. Similar to Leclercq and Geroliminis (2013), moving observers are free to choose any path in $\mathcal{G}$, and are not restricted to one direction as in Daganzo and Geroliminis (2008). Second, we approximate the capacity branch also based on $\mathcal{G}$. Finally, we estimate the network-wide jam density, and exploit the symmetry of the link FD indicated in Laval and Castrillón (2015)and Daganzo and Knoop (2016) to derive the congested branch.

5.1. Free-flow branch

Similar to the original MC, we rely on the concept of moving observers which travel through the network. We aim at estimating the free-flow branch of each route MFD to be able to aggregate them to the free-flow branch of the network MFD. However, flow conservation within a route might be violated when source terms exist, and thus MC cannot be directly applied (see Section 2.2). To address this, we propose an approximate approach. It is based on the commonly known notion of path flows, which we will refer to as *route flows* to avoid any confusion with the term *path* as used in Section 2.

Route flows consist of vehicles which share a common route from an origin to a destination, and flow conservation applies. Thus, we can apply the original MC by deriving hyperlinks for these routes from the hypernetwork $\mathcal{G}$. Spillback effects are accounted for as the costs are derived from $\mathcal{G}$, and demand-related effects are not considered as flow conservation applies. The general cost of an edge representing a green phase relates to the share of the route flow to the total flow at the specific intersection approach. By deriving all route MFDs, and subsequently taking the route length-weighted average, one can approximate the network MFD. Note that this corresponds to the aggregation of traffic states according to Edie's definitions (Edie, 1963).

Unfortunately, the number of routes connecting each OD pair can become intractable for realistic networks. Hence, to reduce the number of evaluated routes, we conjecture that the network MFD is estimated sufficiently well when the network variability is covered by the evaluated hyperlinks. The corresponding set of routes is non-overlapping and includes each intersection approach. Note that such a set is given by the set of corridors $\mathcal{C}$ defined by the network decomposition in Section 4.1. We derive the cuts corresponding to forward-moving observers as described in the original MC for each hyperlink. Eventually, we average the route-specific free-flow branches to derive the one corresponding to the network MFD. We conduct a sensitivity study in Section 6.2.3 to support this procedure and the involved assumptions.

5.2. Capacity branch

To derive the capacity branch of the network MFD, we fully utilize the hypernetwork $\mathcal{G}$. It already includes estimates for the maximum flows on each link, $\tilde{q}_k$. The computation of these flows considered both decreased demand due to a net outflow, and spillback induced capacity reductions. We take the link-weighted average to derive the network-wide capacity, as proposed in Geroliminis and Daganzo (2008). Note

that phenomena such as the short-blocks problem (e.g. Daganzo and Lehe, 2016) are implicitly taken into account in our methods to define the hypernetwork as we consider spillback effects explicitly.

5.3. Congested branch

To conclude the estimation of the MFD we approximate its congested branch as follows. First, we estimate the network jam density K_{max}. Then, we exploit a symmetry in the evolvement of free-flow and congested traffic states considering the estimation for K_{max}. This symmetry was shown by Laval and Castrillón (2015); Daganzo and Knoop (2016) concerning the FD which exists when a canonical density transformation is applied. This aspect reflects the symmetric propagation of free-flow and congested traffic states throughout a link. They utilized this symmetry also for the MFD. Our main assumption is that the symmetry still holds when the network jam density is reduced. We discuss this assumption below.

5.3.1. Network jam density

To our best knowledge, no explicit attempts to estimate the network-wide jam density in the context of MFD approximation exist. The network jam density differs from the link jam density because the network is not necessarily fully jammed in case of gridlock (Daganzo et al., 2011a; Mahmassani et al., 2013; Mazloumian et al., 2010). In the case study presented in Section 6 the significant improvement of the MFD estimation considering the network jam density becomes apparent.

In simple words, we let the traffic in the physical network reach the capacity state and then simultaneously set the capacity of all outgoing links at the boundary to zero. This results in queues originating on those links which propagate through the network until a gridlock state is reached. By approximating the queue growth throughout the network, we can derive the network jam density K_{max}.

When we restrict the outflow at the network boundaries, the density K is influenced by the remaining outflow and inflow. The strongest outflow restriction is represented by a simultaneous closing of all outgoing links at the boundary. If some links were not closed, vehicles could still leave the network. The inflow is determined by the maximum of the demand and the capacity at incoming links. While the demand is independent of the outflow restriction, the capacity at incoming links depends on whether queues have already reached these links. Due to traffic interdependencies in the network, queues can grow even though some outflow links might not be restricted. In other words, a queue growing from one outgoing link can finally block multiple incoming links at the network boundary. Thus, for the case of a non-simultaneous outflow restriction, we assume that net outflows into the system are larger than in the case of a simultaneous outflow restriction. Therefore, we conjecture that the maximum number of vehicles in the network, and thus K_{max} considering spatial demand patterns, can be achieved in the latter case. In the following, we propose an approximate approach and one based on nVT for this purpose.

Approach 1 - Queue propagation (QP). We assume the capacity of outgoing links at the boundary of $\mathcal{N}$ to be zero and trace the propagation of queues throughout the network. Thus, we label this approach as 'queue propagation' (QP). Once a stationary gridlock state is reached, we derive the number of vehicles per link, and subsequently, take the link-weighted average to derive K_{max}.

For each link L_k^i we derive the time instant $t_{k,dq}^i$ when the queue starts to grow at the link's downstream end, and approximate the time when it reaches the link's upstream end, $t_{k,uq}^i$. Note that the latter time instant equals to $t_{k-1,dq}^i$, i.e. the time when the queue reaches the upstream link L_{k-1}^i, if no earlier blockage from an adjacent link occurs. To derive $t_{k,uq}^i$, we distinguish between two cases.

In the first case, we focus on links where no spillback occurred when the network was at capacity, i.e. $\tilde{q}_k^i < q_{k,s}^i$. Such a case is displayed in Figure 7. It shows a time-space diagram for the link L_k^i between two intersections I_{k-1}^i and I_k^i.

The left part of the figure reflects the traffic conditions for the network capacity state (see Section 4.3.1). At time $t_{k,dq}^i$ a queue reaches the intersection I_k^i from downstream. Note that $t_{k,dq}^i$ can result from downstream links from both corridors, C^i and C^j. However, the earlier queue is decisive for the further propagation, as we assume FIFO diverging behavior. If the earliest queue arrives from C^j, it occurs during a red

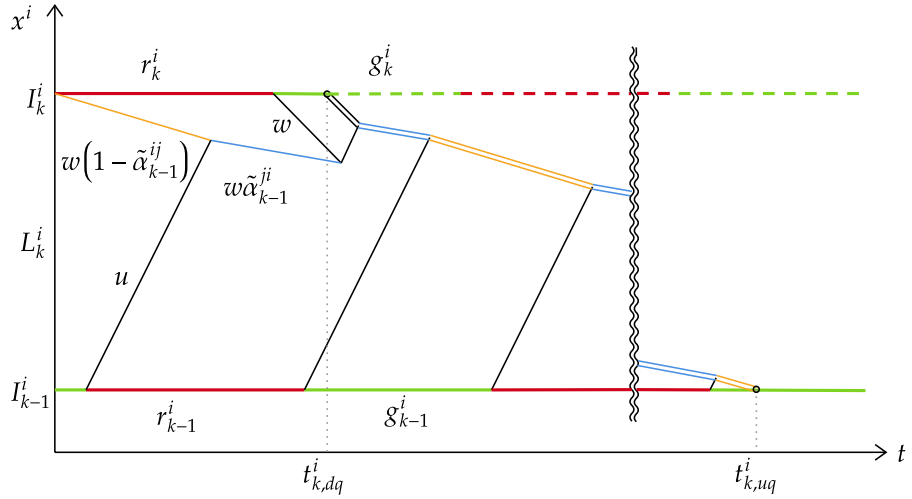


Figure 7: Queue growth for undersaturated links.

phase on C^i. The queue on link L_k^i starts to grow with the start of the respective red phase, as vehicles stopping due to a red light will not be able to continue their travel once the green phase starts due to the existing queue downstream. Formally, the time when the queue starts to grow at the downstream end of link L_k^i can be written as:

$$t_{k,dq}^i = \min\left(\lfloor\frac{t_{k+1,uq}^j}{c_k^i}\rfloor c_k^i, t_{k+1,uq}^i\right), \tag{21}$$

where the $\lfloor . \rfloor$ is the *floor* function, which ensures $t_{k,dq}^i$ starts with a cycle and therefore with the red phase if the decisive queue comes from C^j.

From this time onward, the capacity at I_k^i is reduced to zero and the queue starts to grow on link L_k^i. During the occurrence of discharging flows, the queue grows with speed w as shown in the figure. In case that flows from the upstream intersection I_{k-1}^i reach the queue, it grows with speeds $w(1-\tilde{\alpha}_{k-1}^{ij})$ and $w\tilde{\alpha}_{k-1}^{ji}$. The growth of the queue is displayed as a double line in the figure. For the sake of simplicity, we neglect the impact of discharging flows which are supposed to be small in the considered case, and approximate the growth as the cycle-based average of the blue and orange shock waves:

$$t_{k,uq}^i = t_{k,dq}^i + \frac{l_k^i}{w}\frac{c_k^i}{r_k^i\tilde{\alpha}_{k-1}^{ji} + g_k^i(1-\tilde{\alpha}_{k-1}^{ij})}. \tag{22}$$

In the second case, we focus on an intersection I_{k+1}^i that is oversaturated when the network is at capacity state, and therefore a spillback occurs. Thus, the queue grows faster as in the first case since incoming flows are typically close to saturation.

This is shown in Figure 8 as a time-space diagram for link L_{k+1}^i between the intersections I_k^i and I_{k+1}^i. Note that a spillback occurs at I_k^i denoted as σ_k^i. In the figure, a queue reaches I_{k+1}^i at $t_{k+1,dq}^i$ from downstream. The queue grows with speed w as long as saturation flow reaches it. Closer to I_k^i, the queue grows depending on the inflows with speed $w(1-\tilde{\alpha}_{k-1}^{ij})$ and $w\tilde{\alpha}_{k-1}^{ji}$, and reaches I_k^i at $t_{k+1,uq}^i$. If $t_{k+1,dq}^i$ equals the beginning of the red phase r_{k+1}^i, we can calculate $t_{k+1,uq}^i$ based on eq.(12) in Section 4.3.1, i.e. by deriving the spillback time σ_k^i. Contrarily, if $t_{k+1,dq}^i$ is close to the beginning of the green phase g_{k+1}^i, the queue grows with speed w nearly all the time. However, if $t_{k+1,dq}^i$ is exactly equal to the beginning of the green time, the queue starts growing at the beginning of the previous red phase. To describe this formally, we introduce the parameter $\gamma = [0,1]$, as shown in the figure, which is 1 if $t_{k,dq}^i = r_k^i$, and close to 0 if $t_{k,dq}^i$

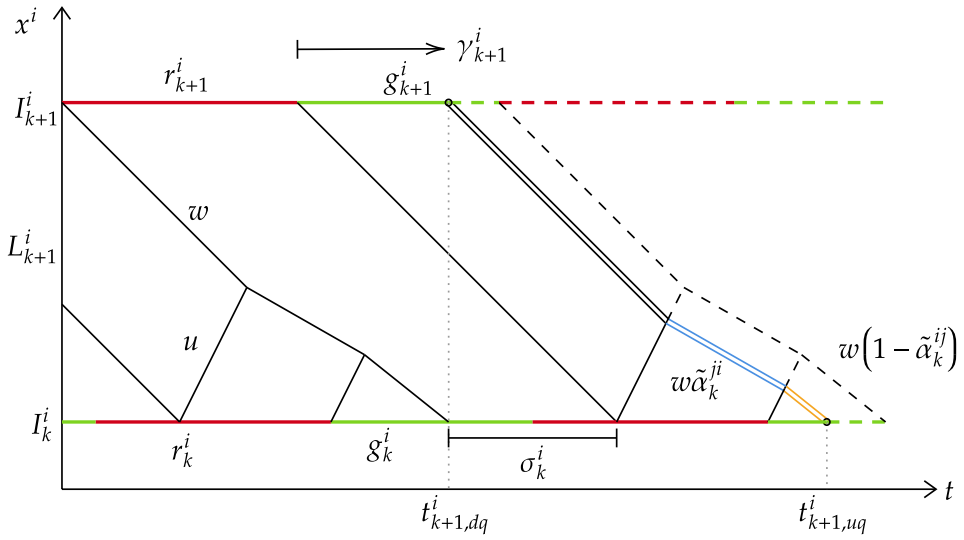


Figure 8: Queue growth for oversaturated links.

is close to g^i_k. To approximate all cases in between, we linearly interpolate between $\gamma = (0, 1]$. The case of $\gamma = 0$ is treated separately.

$$t^i_{k+1,uq} = \begin{cases} t^i_{k+1,dq} + \frac{l^i_{k+1}}{w} + \gamma^i_{k+1}(r^i_{k+1} - \sigma^i_k) & \text{, if } \gamma^i_{k+1} > 0 \\ t^i_{k+1,dq} + \frac{l^i_{k+1}}{w} - \sigma^i_k & \text{, otherwise.} \end{cases} \tag{23}$$

With equations (21)-(23) we approximate congestion propagation throughout the network. However, a blocked intersection due to a queue on C^i also blocks flows to C^j due to the FIFO diverging behavior. This leads to reduced inflows to the respective link. In the extremal case, the entire link on C_j might become empty due to such a blockage. This is essentially the reason for K_{max} being smaller than the link jam density κ_{max}. To account for such cases, we trace the last vehicle which enters and exits a link. A corresponding example is illustrated in Figure 9. In this example, a blockage occurs at intersection I^i_{k-1} at $t^i_{k-1,dq}$. This is due to a queue reaching the intersection from the adjacent corridor C^j, which is not shown in the figure. The remaining vehicles on link L^i_k join the existing queue at I^i_k, and leave the link at $t^i_{k,dq}$ during the respective green phase. In this example, no queue from downstream of I^i_k occurs at $t^i_{k,dq}$ and therefore the link L^i_k becomes empty. In contrast to the cases described above, the flows at I^i_k are reduced to zero due to a lack of demand and not due to a queue-induced blockage.

To describe this formally, let $c^i_{k,f}$ denote the beginning of the cycle in which the final vehicle reaches the existing queue at the downstream end of the link.

$$c^i_{k,f} = c^i_k \lfloor \frac{t^i_{k-1,dq} + \frac{l^i_k}{u}}{c^i_k} \rfloor, \tag{24}$$

The remaining flow relevant to $c^i_{k,f}$ is denoted as $q^i_{k,f}$. It can be derived based on inflows at the upstream intersection. The ratio of the remaining flow to the saturation flow q_{max} corresponds to the part of the green phase where flows larger than zero occur at I^i_k. Thus, the time $t^i_{k,dq}$ when $q = 0$ is equal to the red phase r^i_k of the cycle $c^i_{k,f}$, and the part of the green time until the intersection is cleared.

$$t^i_{k,f} = c^i_{k,f} + r^i_k + g^i_k \frac{q^i_{k,f}}{q_{max}}, \tag{25}$$

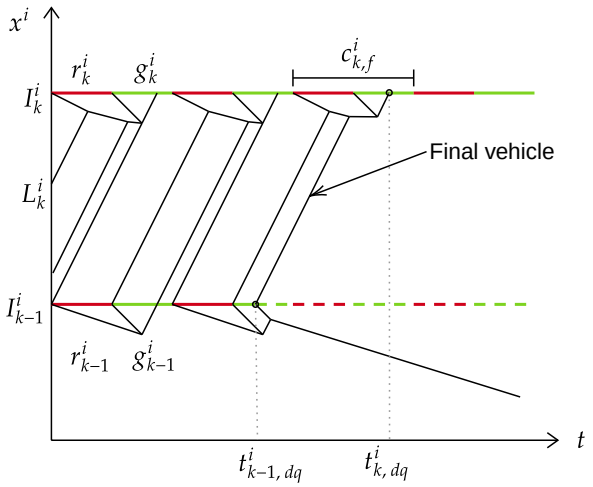


Figure 9: Downstream effects of spillbacks.

To account for this, we rewrite eq.(21):

$$t_{k,dq}^i = \min\left(\lfloor\frac{t_{k+1,uq}^j}{c_k^i}\rfloor c_k^i, t_{k+1,uq}^i, t_{k,f}^i\right). \tag{26}$$

This completes the set of equations needed to track congestion propagation throughout the network for the gridlock case. The equations (22), (23), and (26) are applied for each link L_k^i on a corridor C^i of the set $\mathcal{C}$. However, the queues from an adjacent corridor might reach an intersection earlier than the current one being evaluated with the equations. Thus, an iterative approach needs to be employed until the resulting times $t_{k,uq}^i$ and $t_{k,dq}^i$ converge.

Subsequently, we estimate the maximum density per link K_k^i during the gridlock state. Thereby, we consider that a link might not fully load due to a blockage from an adjacent corridor at the upstream intersection. The time span needed to fully load the link is $t_{k,uq}^i - (t_{k,dq}^i - l_k^i/u)$, which is based on the equations above. However, the actual loading duration for a link is $\max(t_{k-1,dq}^i - t_{k,dq}^i + l_k^i/u, 0)$, since $t_{k-1,dq}^i$ also accounts for blockages from adjacent corridors. The maximum operator is applied to exclude infeasible values. We approximate the actual maximum density per link as follows:

$$K_k^i = \kappa_{max}\frac{\max(t_{k-1,dq}^i - t_{k,dq}^i + l_k^i/u, 0)}{t_{k,uq}^i - t_{l,dq}^i + l_k^i/u}, \tag{27}$$

Finally, the network-wide jam density is calculated as:

$$K_{max} = \frac{\sum_{i,k} K_k^i}{\sum_{i,k} l_k^i}, \tag{28}$$

Approach 2 - Network variational theory (nVT). Similar to estimating the cost of horizontal edges in $\mathcal{G}$, we propose a second approach based on nVT to estimate the network-wide jam density K_{max}. Compared to the QP approach, it has the advantage of being more exact as fewer assumptions are involved. On the other hand, this approach has a higher modeling complexity and computational cost. Alternatively, any other numerical solution method to solve network KWT problems applies. However, there exist synergies between nVT and the nMC framework proposed in this paper which we intend to utilize (see Section 4.3.3).

To estimate K_{max}, we let the network reach the capacity state as in the estimation of costs for $\mathcal{G}$, and then block the outflow links. This leads to queues being propagated throughout the entire network. Once

the system reaches a stationary state, the link density can be derived by evaluating the difference of the Moskowitz function in the spatial dimension for each link, and subsequently dividing it by the link length:

$$K_k^i = \frac{N_{k,x=0}^i - N_{k,x=l_k^i}^i}{l_k^i}. \tag{29}$$

Based on K_k^i for each link, we eventually apply eq.(28) to derive K_{max}.

5.3.2. Density transformation

In order to finish the estimation of the congested branch of the network MFD, we utilize the symmetry between the free-flow and congested branch, as suggested by Laval and Castrillón (2015); Daganzo and Knoop (2016).

More specifically, we transform densities of the free-flow and capacity MFD branch:

$$\kappa' = \kappa - \frac{1}{2}\left(1 - \left(\frac{1}{-w} - \frac{1}{u}\right) q\right). \tag{30}$$

where κ' denotes the transformed density. Note that this equation is based on a normalized FD with $q_{max} = 1$ and $\kappa_{max} = 1$. Based on the estimation of the free-flow and the capacity branch in Sections 5.1 and 5.2, we know the corresponding (κ, q) values. As noted in Laval and Castrillón (2015), the respective transformed densities range between $[-0.5, 0]$.

To approximate the congested branch, we mirror those traffic states by changing the sign of κ' for each traffic state (κ', q). Re-transforming the original as well as the newly generated traffic states based on eq.(30), and considering K_{max} instead of the link jam density κ_{max} leads to a fully defined MFD including the congested branch. Hereby, the main assumption is that the symmetry still holds even though we utilize the reduced network jam density instead of the link jam density. This assumption is equivalent to assuming an increase of the average speed with which congestion spreads at the corridor level. Thus, also the average backward wave speed can be decreased (i.e. the value becomes more negative). This seems to be reasonable given the following example. Consider a corridor with three links in a network. Let us assume the most downstream link becomes fully congested, but then a queue from an adjacent corridor blocks the upstream intersection such that the second link remains empty, and the queue continues to grow on the most upstream link. Eventually, only the most downstream and most upstream links are fully congested, while the middle link remains empty. Congestion spreads with the corresponding shock wave speed on the most downstream link, skips the middle link, and then spreads on the most upstream link. The total time to reach a stationary state divided by the corridor length can be interpreted as the speed with which congestion spreads for this specific corridor. This speed is higher than for the case without any network effects. This corresponds to the implicit assumption that the average speed with which congestion spreads is increased, when we apply the shear transformation using the network jam density K_{max}. Thus, we conjecture that assuming the symmetry to hold for K_{max} is reasonable.

5.4. Summary

The sections above described the nMC framework we propose to estimate the realized MFD for realistic networks. This includes the effects of source terms such as undersaturated links and spillbacks. For the sake of clarity, we hereafter provide a brief summary.

The main required input is the network $\mathcal{N}$. This includes the network topology such as the number of intersections, respective control settings, number of links, the link FDs, as well as the turning ratios at intersections.

The next step is the generation of the hypernetwork (Section 4). This requires the decomposition of the network $\mathcal{N}$ into a set of corridors $\mathcal{C}$ (Section 4.1). Based on these corridors, the hypernetwork is created to reflect supply conditions induced by signal control and network topology (Section 4.2). Subsequently, we propose to exogenously consider the effects of source terms. We do so with two approximate approaches LS and

FS (Section 4.3.1 and 4.3.2), as well as the application of nVT (Section 4.3.3) as substantial methodological synergies exist with the herein proposed nMC.

With the hypernetwork $\mathcal{G}$ we can then estimate the free-flow branch (Section 5.1) and the capacity branch (Section 5.2). To estimate the congested branch, we propose a method which estimates the network-wide jam density and utilizes symmetries in the FD (Section 5.3).

6. Case study

In this section, we conduct a case study to test our framework for the realistic network of Sioux Falls. Thereby, we estimate the network-wide realized MFD by applying the nMC to derive the hypernetwork and the network jam density, as well as the current state of the art, and the CTM to derive a ground truth. Then, we report on the computational efficiency of the nMC. In addition, we investigate the role of the route selection for the MFD estimation as described in Section 5.1. Then, we investigate the role of turning ratios for both the MFD in general as well as for our MFD estimation. This shows that the MFD indeed is impacted by spatial demand patterns and that our methodology delivers satisfying results independently of $\mathcal{A}$. Lastly, we discuss the fundamental assumption of the existence of stationary traffic states.

6.1. Case study design

6.1.1. Network initialization

The case study is conducted based on the bi-directional Sioux Falls network as shown in Figure 10. For the sake of simplicity, we alter the original Sioux Falls network slightly by only considering intersections with four or fewer legs, and we disallow left-turns at intersections. In total, the network consists of 23 intersections, shown as filled and empty circles, connected by 36 bi-directional links, as highlighted by the arrows. All intersections are controlled by a fixed signal control scheme with a cycle time of 90 s, and a green and red phase each of 45 s. All offsets are set to zero. We assume an identical FD for all links, characterized by $q_{max} = 1800\,\text{veh/h}$, $u = 10\,\text{m/s}$, and $w = -5\,\text{m/s}$.

We define origin and destination nodes as those with less than four legs, e.g. nodes at the fringe of the network. They are shown as filled circles in the figure. At these nodes, we add virtual links which serve as entry and exit to the network. Additionally, we randomly select reasonable but different turning ratios for each intersection between $0.25 \leq \alpha \leq 0.75$. These limits are chosen to exclude extremal turning ratios which lead to special and rather unrealistic spillbacks occurrences in the network, as numerical investigations have shown. Still, the chosen turning ratios lead to an inter-dependency of traffic dynamics across the entire network. This includes the propagation of spillbacks across corridors but also the fact that not all links are necessarily saturated when the network operates at capacity.

6.1.2. Ground truth

In order to further evaluate our MFD estimation results, we compare it to a ground truth derived from the CTM which is a macroscopic, time- and space-discrete traffic model. We implement the same network topology and the total simulation time to 2 h. We set the simulation time step to $\Delta t = 0.25\,\text{s}$, and the spatial step size to $\Delta x = u\Delta t$ to satisfy the Courant-Friedrichs-Lewy condition (Daganzo, 1992). Setting a lower time-step was not feasible due to the high computational cost.

We define several different demand loading scenarios which are evaluated sequentially. In each scenario, the temporal demand profile is kept constant to reach stationary states after some time. To derive gridlock states, we reduce the capacity at destination links to create congestion which spreads throughout the network. Note that such an exogenous supply reduction is required to reach full gridlock. This reduction is increased in a step-wise manner to produce multiple congested traffic states. To plot the related MFD, we aggregate flow and density values for one cycle length once stationary states are reached at the end of the simulation. We verify the existence of a network-wide stationary traffic state by ensuring that no differences exist between the average densities of the last two cycles in each simulation run.

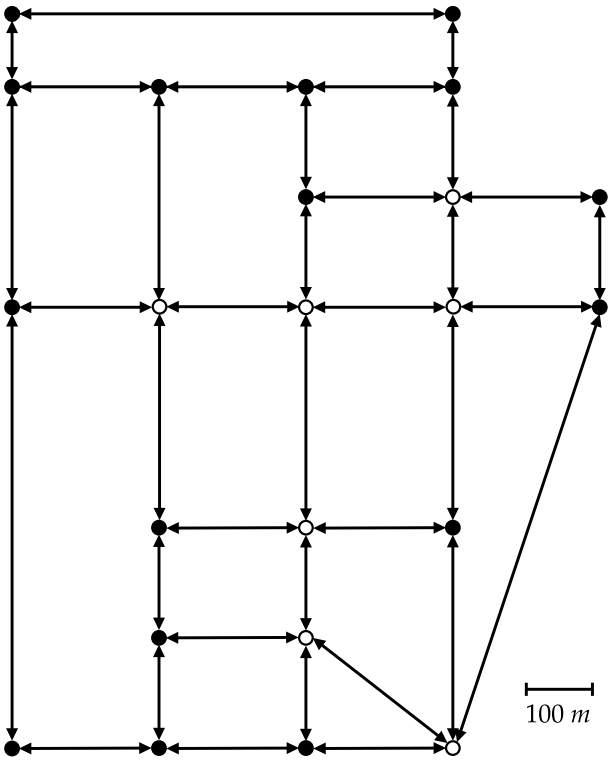


Figure 10: Sioux Falls network for the case study.

Note that the CTM includes a certain error in the calculated density and flow values due to the numerical diffusion. This error converges to zero with a decreasing time step Δt. The chosen time-step represents a reasonable trade-off between accuracy and computational cost. Nevertheless, the ground truth results should be treated accordingly when being interpreted.

6.1.3. Settings for the network variational theory approach

The nVT is applied to derive the hypernetwork and to estimate the network-wide jam density. For both applications, a numerical grid is defined based on the decomposed network (see Section 4.1). The total computation period is set to 2 h to reach stationary traffic states, and the time-step is set to $\Delta t = 0.1$ s. This small time-step highlights the lower computational cost and consequently increased accuracy of nVT compared to the CTM.

In order to estimate K_{max}, the capacity at outgoing links at the network boundary is set to zero after 1.5 h. This ensures that the stationary traffic states related to the network capacity are reached before the gridlock state occurs.

6.1.4. State of the art

We choose the MC by Daganzo and Geroliminis (2008), by Leclercq and Geroliminis (2013), and the SA by Laval and Castrillón (2015) as state of the art for this case study. Hereafter, we abbreviate them with 'Dag', 'Lec', and 'Lav'. The first method is chosen as it represents the original version of the MC applied to estimate the MFD. The authors find a representative corridor of the network, consider average supply parameters such as block lengths and signal settings, and estimate the MFD. The second method was the first to explicitly account for irregularities in corridors, i.e. differing block lengths and signal settings. The method can be applied to each corridor, and the resulting MFDs be averaged to approximate the network MFD. The last method accounts for stochasticity in the network. In the evaluated case, the level

of stochasticity remains small and only refers to the varying block lengths. Nevertheless, we include the method for comprehensiveness. Note that none of these methods explicitly account for the effects of source terms, i.e. the exchange of flows across corridors. To our best knowledge, methods that estimate the MFD while accounting for turning flows in realistic networks do not exist, as the reported studies all apply to simplistic networks such as regular grids or two-ring networks.

6.2. Results and discussion

In the following, we present several MFDs. The x-axis always displays the network-wide average density K in veh/km, and the y-axis the network-wide average flow Q in veh/h.

6.2.1. Impact of hypernetwork generation method

We compare the MFDs estimated by our proposed framework with the one derived by the 'Dag', 'Lec', and 'Lav' methods, as well as with the CTM ground truth. The results are displayed in Figure 11. The CTM results are shown as grey diamonds. The MFDs from the proposed framework are based on the nVT approach, as well as on the LS and FS approaches to derive $\mathcal{G}$ combined with the QP approximation for the network jam density. They are displayed as a solid black curve, a dashed black curve, and a dotted black curve, respectively. The MFD resulting from the state-of-the-art methods are shown as solid, dashed, and dotted grey curves.

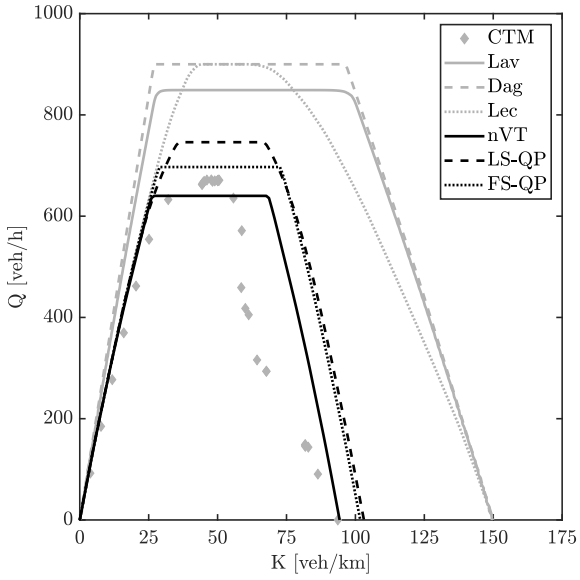


Figure 11: Resulting MFDs from the proposed framework based on the nVT, the LS-QP, and the FS-QP approach, the state-of-the-art methods represented by the original method of cuts by Daganzo and Geroliminis (2008) ('Dag') and by Leclercq and Geroliminis (2013) ('Lec'), the stochastic approximation by Laval and Castrillón (2015) ('Lav'), and the CTM ground truth.

The ground truth MFD reaches a capacity of ca. 670 veh/h, and a network-wide jam density of ca. 94 veh/km.

The nVT approach to derive the hypernetwork and estimate the network-wide jam density results in the most accurate MFD estimation as indicated in the figure. The capacity is ca. 25 veh/h lower than the one of the CTM, while the network-wide jam density is the same. The reason for this is that nVT itself solves the KWT problem at the network level accurately, and its integration into the hypernetwork generation, as well as into the estimation of K_{max}, let the overall framework profit from such accuracy. As expected, the estimated MFDs resulting from the hypernetwork built based on the approximate approaches are less

accurate. The first approach (abbreviated as 'LS-QP') which does not consider network-wide spillback propagation overestimates the capacity by ca. 75 veh/h and the jam density by ca. 10 veh/km. The second one, abbreviated as 'FS-QP', which includes network-wide spillback propagation, estimates Q_{max} as accurately as the nVT approach, with an error of 25 veh/h. However, here the difference results from an overestimation rather than an underestimation. The error in the jam density estimation is ca. 8 veh/h and thus similar to the LS-QP approach. In reality, an accurate capacity branch estimation is more important as the congested branch is rarely observed (Loder et al., 2019). Thus, we adjudge the increased modeling complexity of the FS approach as valuable for the presented case.

The differences in the network-wide capacity and the jam density estimation by the approximate approaches and the nVT one are due to the assumptions made within the approximations. In particular, averaging the downstream spillback impact uniformly on the downstream capacity neglects the temporal component of the spillback occurrence. However, such temporal aspects impact the further propagation of the spillback, i.e. to what share it is propagated to corridors C^i and C^j. Similar to that, we assume the inflow into a link as the equally weighted mean of both retaining and turning flows to derive K_{max}. The shorter a link is the more likely these assumptions are violated.

The results from the state-of-the-art methods substantially differ from the ground truth. Both the capacity and the jam density are significantly overestimated. However, this is expected, as these methods cannot account for the effects of source terms on the network MFD. In other words, they cannot account for the effects of demand patterns, i.e. turning ratios, which is important to derive an estimate for the realized MFD as our results clearly show.

Given the error included in the ground truth and the magnitude of improvement compared to the results of the state-of-the-art methods, the differences in Q_{max} and K_{max} appear to be small for all MFDs resulting from our proposed framework. Moreover, these first results indicate that indeed the overall estimation accuracy is the highest for the nVT approach and the consideration of spillback propagation throughout the network in the FS approach is valuable.

6.2.2. Computational costs

The modeling complexity of the nVT approach is clearly the highest compared to the other two semi-analytical approaches. This is not surprising, as it numerically solves the underlying KWT problem at the network level. For the case study settings, the nVT takes 4 h 50 min to derive the MFD. The FS-QP approach consists of more decision variables in the optimization problem than the LS-QP one and thus has the highest complexity of the two. Both approaches take 1.2 min for the case study settings and are thus substantially cheaper than the nVT from a computational perspective. The computational cost of the nVT strongly scales with the chosen time-step (Tilg et al., 2021). For example, if the time-step is set to $\Delta t = 1\,\text{s}$, the computation time is only 14.3 minutes for the case study. The other two methods scale primarily with the number of intersection legs in the network.

The computation times for these three approaches are of the magnitude of some minutes, while the CTM requires several days to derive the MFD for this case study.[1] This high computational time for the ground truth results from the small time-step (even though it is still 2.5 times larger than that used in the nVT approach), the resulting large number of cells, and the necessity to reach stationary states. Similar or even larger computational requirements could be equally expected from other simulations. Furthermore, note that the optimization procedure integrated in the nMC framework ensures that highest network-wide flows are found, while the CTM-related MFD only represents a point-wise evaluation of traffic states which can miss the true capacity.

6.2.3. Impact of route selection

As described in Section 5.1, an assumption made in order to reduce modeling complexity and computational burden while approximating the network MFD refers to the set of routes based on which the free-flow branch is estimated. In Section 5.1, we propose to estimate the free-flow branch for each corridor $C \in \mathcal{C}$

[1] The methods were evaluated on a computer with an Intel(R) Xeon(R) W-2145 CPU with 3.70 GHz and 64 GB RAM.

being aware that this is a simplification. In order to study the implication of this simplification, we estimate the MFD based on the nVT approach for two different sets of routes in the network. First, we evaluate the MFD based on $\mathcal{C}$ (as proposed in Section 5.1), which results in 24 different routes. Second, we extract the shortest paths for all OD pairs in the Sioux Falls network and define each path as a route for which the MFD is estimated. In this case, the number of routes increases to a total of 576. Note that the capacity branch is not affected by this assumption and the congested branch equals the transformed free-flow one. Therefore, we only evaluate the free-flow branch of the MFD for the two sets of routes and compare them below.

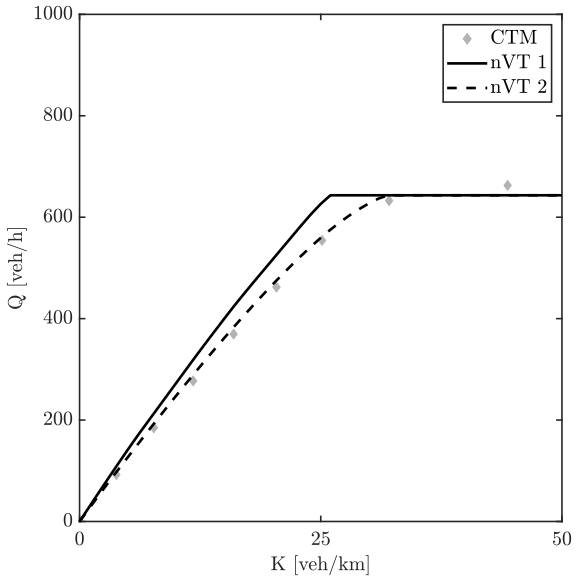


Figure 12: Comparison of the impact of route selection.

Figure 12 shows the free-flow branch of both MFDs. The solid black curve labeled as 'nVT 1' represents the derived free-flow branch based on $\mathcal{C}$. The dashed black curve 'nVT 2' illustrates the one which results from considering the second set of routes. The figure indicates that indeed the estimation's accuracy is further increased by the latter approach as the curve is closer to the ground truth MFD. However, the number of paths to be evaluated is 24 times higher for our case study. This leads to a significant increase in computational cost. In the end, the selection of routes for the MFD estimation is use-case dependent and the choice represents a trade-off between computational burden and estimation accuracy.

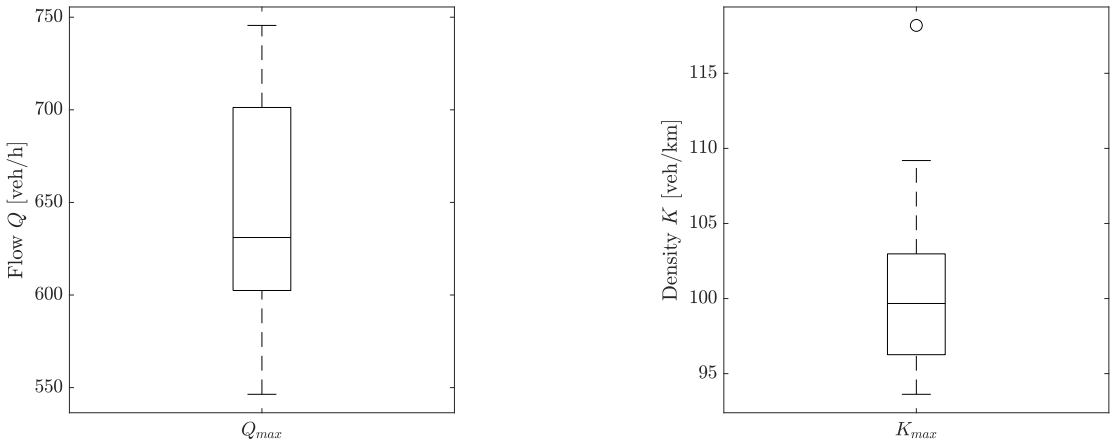
6.2.4. Impact of turning ratios

Next, we aim at evaluating our framework for the presented network with varying turning ratio sets $\mathcal{A}$. More specifically, we first analyze the variance of the resulting MFDs from the ground truth to validate that spatial demand patterns indeed have an impact. Second, we compare the results from our proposed framework to the state of the art and the ground truth to assess the quality of our approach. Last, we pick a specific set $\mathcal{A}$, introduce small deviations, and investigate the effects on the resulting MFDs based on our framework.This indicates the robustness of our MFD estimation method against different spatial demand patterns.

For the first analysis, we randomly select ten different reasonable sets $\mathcal{A}$ with turning ratios ranging again between $\alpha = [0.25, 0.75]$ with a mean at 0.5. We then evaluate the CTM ground truth, the LS, the FS, and the nVT approaches to derive the hypernetwork, and the QP and nVT approaches to approximate the network jam density K_{max}. Moreover, the state-of-the-art methods are applied. We estimate the free-flow

branch only for the set $\mathcal{C}$, as we have shown in the previous section that the impact of this assumption is minor.

As the proposed methodology primarily focuses on improving the estimation of the maximum flows $\tilde{q}$ and thus the network-wide capacity Q_{max}, as well as the jam density K_{max}, we merely compare the corresponding results. This allows us to drastically reduce the scenarios to be run with the CTM. The results from the CTM simulation lead to varying Q_{max} and K_{max} for each scenario as illustrated in Figure 13. A range of about 200 veh/h for Q_{max} can be observed. The values of K_{max} cover a range of nearly 25 veh/km. The significant variance in these results shows that the MFD indeed depends on the spatial demand pattern in the network. This further indicates that a framework to estimate the MFD for specific spatial demand patterns is beneficial. Despite the assumption regarding the existence of stationary states, which might be violated in reality, the approximated MFDs can be useful.



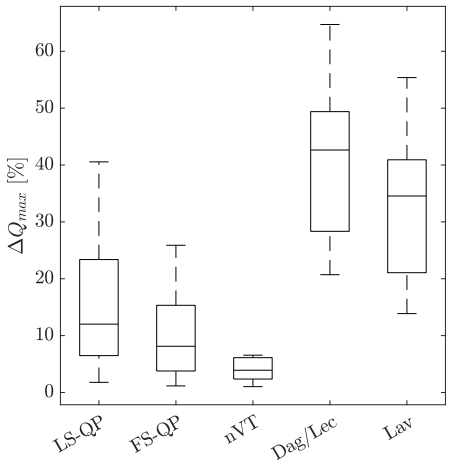
a) Q_{max} from CTM simulations with varying turning ratios.

b) K_{max} from CTM simulations with varying turning ratios.

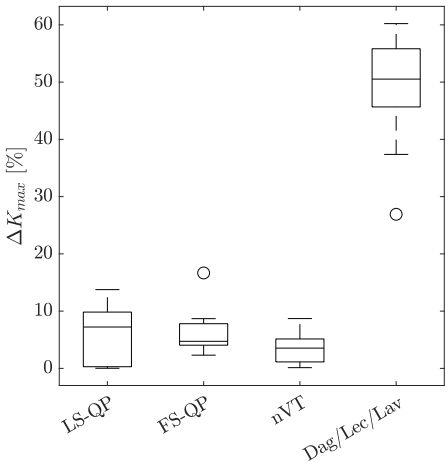
Figure 13: Boxplots showing the variance of Q_{max} and K_{max} across the evaluated scenarios.

To further assess the nMC, we calculate the absolute relative differences ΔK_{max} and ΔQ_{max} for the MFD estimations from both the proposed framework and the state of the art, compared to the ground truth. For example, the absolute relative difference ΔQ_{max} between the nVT approach and the CTM can be calculated as $\Delta Q_{max} = |\frac{Q_{max,nVT} - Q_{max,CTM}}{Q_{max,CTM}}|$. Analogous calculations are performed to derive the other absolute relative differences. Note that Q_{max} and K_{max} for all state-of-the-art methods remain invariant across the scenarios since they do not account for turning ratios.

Figure 14 summarizes the results. Figure 14a shows the difference in the capacity estimation from the proposed LS-QP, the FS-QP, the nVT approach, and the state of the art compared to the ground truth values. The differences of the methods 'Dag' and 'Lec' do not vary since their estimated capacity is equal. The figure clearly shows the substantial improvement in the estimation accuracy regarding the network-wide capacity of all three proposed methods. Moreover, it is apparent that the FS-QP approach, which models network-wide spillback propagation in more detail, is more exact than the LS-QP approach, which does not account for such effects. Furthermore, the nVT approach is even more precise in estimating the network-wide capacity. The average capacity estimation of the proposed approaches is more than five times as accurate as of the estimate by the state of the art. Additionally, we investigate the estimation of the network-wide maximum density K_{max} (see Figure 14b). While the average estimation of K_{max} by the nVT approach seems to be more accurate than the one by both approximate approaches, the differences are less obvious. Again, the proposed methods substantially improve the estimation of K_{max} compared to the state of the art, which assumes the jam density equal to the one of the link FD, i.e. $K_{max} = \kappa_{max}$. On average, the



a) Absolute value of the relative difference ΔQ_{max}.



b) Absolute value of the relative difference ΔK_{max}.

Figure 14: Analysis of the estimated network-wide capacity Q_{max} and jam density K_{max}. The evaluated approaches are the proposed framework, i.e. LS-QP, FS-QP, and nVT, as well as the state of the art, i.e. Dag, Lec, and Lav.

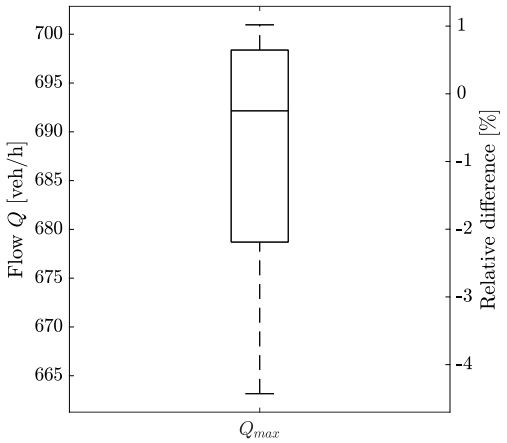
estimated K_{max} is more than five times closer to the ground truth value than the estimate from the state of the art.

The analysis above showed that our proposed framework has reasonable accuracy. This lets us conduct analyses that are not possible with other methods, such as the CTM, due to their high computational cost. To showcase this, we further examine the robustness of the estimated MFD with regard to small deviations of the turning ratios. The results can indicate how often the MFD has to be estimated during the day. To analyze the effect of small deviations of turning ratios, we first define a reference set $\mathcal{A}$ and then introduce random variations with a maximum of $\pm 5\,\%$ relative turning flows. We sample 100 sets with different variations and evaluate the MFD for each of them based on the FS-QP approach. The results are displayed in Figure 15 as boxplots for Q_{max} and K_{max}, including the relative differences to values referring to the reference set $\mathcal{A}$. The values of the capacity Q_{max} lie in a range of about 40 veh/h. Compared to the results displayed in Figure 13 which correspond to a strong variation of turning ratios, this is clearly a small range of variation. This is further highlighted by the relative differences which are below 4 %. However, it is apparent that the capacity of the estimated MFD is sensitive to turning ratios. Fortunately, due to the low computational cost of the proposed approach, one can increase the robustness of the results by simply introducing small deviations in the turning ratios and then taking the average capacity Q_{max}. Regarding the network-wide jam density, the absolute variation lies in a range of 5 veh/km. Considering that the relative differences are also below 4 %, except for two outliers, the sensitivity of the network jam density is similar to the capacity. Nevertheless, the robustness can also be increased by taking the average from an array of estimated values. This example further confirms that spatial demand patterns indeed impact the MFD and highlights the value the proposed framework provides for the analysis of urban traffic.

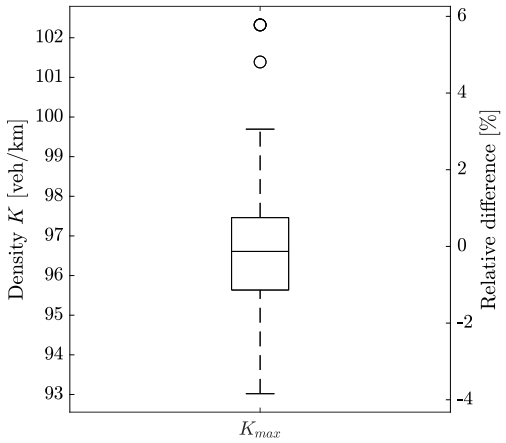
6.2.5. Stationary traffic states

Lastly, we want to discuss one of the main assumptions of our framework inherited from the general concept of the MFD and the original MC - the ability to reach stationary traffic states.

For a given spatial demand pattern, which is represented by turning ratios at intersections in our framework, we conjecture that stationary states exist for any such pattern if the temporal demand profile is constant. Both traffic states and shock waves travel with finite speeds within a network with finite length. Thus, shock waves reach the boundaries of the network within finite time. Once all shock waves arrive at the boundary, a stationary state is reached at the network level. The CTM, employed within our case study,



a) Q_{max} from FS-QP-based MFD estimation with varying turning ratios.



b) K_{max} from FS-QP-based MFD estimation with varying turning ratios.

Figure 15: The variance of Q_{max} and K_{max} for the evaluated scenarios with small deviations of turning ratios.

shows that such stationary states can be reached. Moreover, the existence of stationary states is a common assumption in related literature (e.g. Leblanc, 1975). While this indicates that the existence of stationary states is a reasonable assumption, their uniqueness might not necessarily be given.

On the contrary, the results from theoretical studies for simple network topologies (e.g. two-ring networks) indicate that congested traffic states might not be unique even when being stationary, but rather depend on the initial distribution of vehicles (Jin et al., 2013a). However, note that we only consider two different initial states to derive the MFD. The first regards an empty network, and the second to the network at capacity. Thus, the initial states are constant. This and the fact that our node models are deterministic indicate that the resulting MFDs from our framework are unique. Moreover, phenomena such as bifurcations are rarely observed in empirical data (e.g. Daganzo et al., 2011b). This further indicates that the assumption of a unique MFD for a fixed demand and supply is reasonable. Nevertheless, such considerations are only valid for a slow-varying demand profile. More specifically, if the demand changes faster than the system's response can be (e.g. due to a larger network), then no stationary states will occur (e.g. Leclercq and Paipuri, 2019; Daganzo, 2007).

Therefore, when applying the MFD derived from semi-analytical estimation methods, independently of whether the original MC or our nMC framework is utilized, one must keep in mind that stationary states might not be observed in real life. Due to fast-varying temporal demand profiles and possible hysteresis patterns, the empirically observed MFD might deviate from the semi-analytical estimated one. However, even in this extreme case, deriving the stationary MFD provides valuable information. In particular, the network capacity reflects the maximum outflow from the transportation system, which is a key element to characterize its loading and unloading during peak hours (e.g. Mariotte et al., 2017).

Furthermore, the semi-analytical MFD estimation can be useful as a first approximation for network design and planning purposes, e.g. to compare and evaluate several network-related measures, or as a tool to understand the network-wide impact of local bottlenecks with respect to a specific spatial demand pattern. For example, Hu et al. (2020) proposed a framework that utilizes the MFD to solve the network design problem. Furthermore, Tilg et al. (2020b) employed the concept of the MFD in a simulation-based optimization framework to design urban bi-modal transportation systems. These examples highlight that the MFD can indeed be used for network design and planning.

7. Conclusion

In this paper, we propose a general framework, the nMC, to estimate the network-wide realized MFD without relying on extensive empirical data or microscopic simulations. Its flexibility allows to trade-off modeling complexity and computational cost against estimation accuracy. We base our framework on the well-established original MC (Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013) and follow its philosophy while exogenously accounting for the effects of source terms at intersections. The framework consists of two main steps after the problem initialization. First, we build a hypernetwork that reflects the network flows in the capacity state. For this purpose, we propose three different methods, two approximate and a precise one based on the nVT (Tilg et al., 2021). The hypernetwork is the basis to derive the network MFD. Similar to the original MC, we derive cuts to approximate the free-flow branch of the MFD. Moreover, the hypernetwork allows us to efficiently estimate the network capacity. Last, we estimate the congested branch of the network MFD by approximating the maximum density in the network during gridlock, and utilizing symmetries of the underlying link FD. We thoroughly investigate the proposed framework in a case study. This includes the comparison to a ground truth provided by the CTM, as well as to the original MC (Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013) and the SA (Laval and Castrillón, 2015) as state-of-the-art for realistic networks. Furthermore, we analyze our framework's robustness to different spatial demand patterns. The results demonstrate that the proposed framework enables one to estimate the realized network MFD sufficiently well for a realistic network. Moreover, they evidently show the improvement of the estimation accuracy of the network-wide capacity and jam density by our proposed framework compared to existing methods.

Our methodology allows estimating the realized MFD for realistic urban networks without the information loss induced by the reduction of networks to a single corridor as implicitly done by existing methods. Therefore, our framework is able to account for different spatial demand patterns and determine the upper bound of the network-wide average flow. The potential of this methodology lies in its flexibility, and reduced computational cost compared to simulation-based studies. Moreover, it sheds light on the impact of specific assumptions on the MFD estimation. Note that existing approximation methods for realistic networks rather focus on the idealized MFD, and therefore neglect the effects of demand patterns. Following the empirical study of Ambühl et al. (2021) and the theoretical work by Leclercq and Paipuri (2019) which establish the demand-dependency of the MFD from a temporal perspective, our framework is methodological proof that the effects of the spatial aspects of demand on the realized MFD cannot be neglected as well. One field of application is the calibration of aggregate traffic models for large metropolitan regions (Mariotte et al., 2020).

Future work includes the derivation of the idealized MFD based on our proposed framework. This can be achieved by integrating the nMC into an optimization framework with turning ratios as decision variables and the maximum network-wide flow as the objective value. This would lead to the maximum MFD while considering the effects of spatial demand patterns on network-wide traffic states as well spatial dependencies of traffic dynamics. Moreover, multi-modal aspects can be incorporated by modeling the effects of buses on traffic flow and integrating a passenger model. The incorporation of the nVT facilitates the modeling of such aspects as moving or stationary intra-link bottlenecks. This is crucial for considering the effects of buses, which can be modeled as such a type of bottleneck, and therefore can be interesting for estimating the three-dimensional MFD (Geroliminis et al., 2014). Similar extensions of the MC exist already for the corridor level (Dakic et al., 2020; Chiabaut, 2015), and are therefore promising starting points for this research direction.

Declaration of competing interests

None.

Acknowledgments

G. Tilg acknowledges support from the German Federal Ministry of Transport and Digital Infrastructure (BMVI) for the funding of the project LSS (Capacity increase of urban networks). S. F. A. Batista and M. Menéndez acknowledge the support by the NYUAD Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001 and by the Swiss Re Institute under the Quantum Cities™ initiative. L. Ambühl acknowledges the support by the ETH Research Grant ETH-27 16-1 under the project name SPEED. L. Leclercq acknowledges funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 646592 - MAGnUM project).

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. Transportation Research Part B: Methodological 55, 265–281. doi:10.1016/j.trb.2013.07.003.

Aghamohammadi, R., Laval, J., 2021. Macroscopic fundamental diagram parameter estimation: A maximum likelihood approach, in: Proceedings of the 100^{th} Transportation Research Board Annual Meeting (CD-ROM), 12-16 January, Washington, DC, USA.

Aghamohammadi, R., Laval, J.A., 2020. Dynamic traffic assignment using the macroscopic fundamental diagram: A review of vehicular and pedestrian flow models. Transportation Research Part B: Methodological 137, 99–118. doi:10.1016/j.trb.2018.10.017. advances in Network Macroscopic Fundamental Diagram (NMFD) Research.

Ambühl, L., Loder, A., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2020. A functional form with a physical meaning for the macroscopic fundamental diagram. Transportation Research Part B: Methodological 137, 119–132. doi:10.1016/j.trb.2018.10.013.

Ambühl, L., Loder, A., Leclercq, L., Menendez, M., 2021. Disentangling the city traffic rhythms: A longitudinal analysis of mfd patterns over a year. Transportation Research Part C: Emerging Technologies 126, 103065. doi:10.1016/j.trc.2021.103065.

Amini, S., Tilg, G., Busch, F., 2018. Evaluating the impact of real-time traffic control measures on the resilience of urban road networks, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 519–524. doi:10.1109/ITSC.2018.8569678.

Batista, S., Leclercq, L., 2020. Regional dynamic traffic assignment with bounded rational drivers as a tool for assessing the emissions in large metropolitan areas. Transportation Research Interdisciplinary Perspectives 8, 100248. doi:10.1016/j.trip.2020.100248.

Batista, S.F.A., Ingole, D., Leclercq, L., Menéndez, M., 2021. The role of trip lengths calibration in model-based perimeter control strategies. IEEE Transactions on Intelligent Transportation Systems , 1–11doi:10.1109/TITS.2021.3049679.

Batista, S.F.A., Leclercq, L., 2019. Regional dynamic traffic assignment framework for macroscopic fundamental diagram multi-regions models. Transportation Science 53, 1563–1590. doi:10.1287/trsc.2019.0921.

Chiabaut, N., 2015. Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram. Transportation Research Part B: Methodological 81, 410–420. doi:10.1016/j.trb.2015.02.005.

Chiabaut, N., Küng, M., Menendez, M., Leclercq, L., 2018. Perimeter control as an alternative to dedicated bus lanes: A case study. Transportation Research Record 2672, 110–120. doi:10.1177/0361198118786607.

Daganzo, C., 1992. The cell transmission model. part i: A simple dynamic representation of highway traffic .

Daganzo, C.F., 1995. The cell transmission model, part ii: network traffic. Transportation Research Part B: Methodological 29, 79–93. doi:10.1016/0191-2615(94)00022-R.

Daganzo, C.F., 2005a. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transportation Research Part B: Methodological 39, 187–196. doi:10.1016/j.trb.2004.04.003.

Daganzo, C.F., 2005b. A variational formulation of kinematic waves: Solution methods. Transportation Research Part B: Methodological 39, 934–950. doi:10.1016/j.trb.2004.05.003.

Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B: Methodological 41, 49–62. doi:10.1016/j.trb.2006.03.001.

Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011a. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transportation Research Part B: Methodological 45, 278–288. doi:10.1016/j.trb.2010.06.006.

Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011b. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transportation Research Part B: Methodological 45, 278 – 288. doi:10.1016/j.trb.2010.06.006.

Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B: Methodological , 771–781doi:10.1016/j.trb.2008.06.008.

Daganzo, C.F., Knoop, V.L., 2016. Traffic flow on pedestrianized streets. Transportation Research Part B: Methodological 86, 211–222.

Daganzo, C.F., Lehe, L.J., 2016. Traffic flow on signalized streets. Transportation Research Part B: Methodological 90, 56–69. doi:10.1016/j.trb.2016.03.010.

Daganzo, C.F., Menendez, M., 2005. A variational formulation of kinematic waves: Bottleneck properties and examples, in: Proceedings of the 16th International Symposium on Transportation and Traffic Theory, Elsevier, Maryland. pp. 345–364.

Dakic, I., Ambühl, L., Schümperlin, O., Menendez, M., 2020. On the modeling of passenger mobility for stochastic bi-modal urban corridors. Transportation Research Part C: Emerging Technologies 113, 146–163. doi:10.1016/j.trc.2019.05.018.
Dandl, F., Tilg, G., Rostami-Shahrbabaki, M., Bogenberger, K., 2020. Network fundamental diagram based routing of vehicle fleets in dynamic traffic simulations, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–8. doi:10.1109/ITSC45102.2020.9294204.
Edie, L., 1963. Discussion of traffic stream measurements and definitions, in: Almond, J. (Ed.), Proceedings of the 2nd International Symposium on the Theory of Traffic Flow, OECD, Paris, France. pp. 139–154.
Gan, Q.J., Jin, W.L., Gayah, V.V., 2017. Analysis of traffic statics and dynamics in signalized networks: a poincaré map approach. Transportation science 51, 1009–1029. doi:10.1287/trsc.2017.0740.
Geroliminis, N., Boyacı, B., 2012. The effect of variability of urban systems characteristics in the network capacity. Transportation Research Part B: Methodological 46, 1607–1623. doi:10.1016/j.trb.2012.08.001.
Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42, 759–770. doi:10.1016/j.trb.2008.02.002.
Geroliminis, N., Haddad, J., 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, 348–359. doi:10.1109/TITS.2012.2216877.
Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transportation Research Part B: Methodological 45, 605–617. doi:10.1016/j.trb.2010.11.004.
Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transportation Research Part C: Emerging Technologies 42, 168–181. doi:10.1016/j.trc.2014.03.004.
Girault, J.T., Gayah, V.V., Guler, I., Menendez, M., 2016. Exploratory analysis of signal coordination impacts on macroscopic fundamental diagram. Transportation Research Record: Journal of the Transportation Research Board 2560, 36–46. doi:10.3141/2560-05.
Godfrey, J., 1969. The mechanism of a road network. Traffic Engineering & Control 11, 323–327.
Guo, Q., Ban, X., 2020. Macroscopic fundamental diagram based perimeter control considering dynamic user equilibrium. Transportation Research Part B: Methodological 136, 87–109. doi:10.1016/j.trb.2020.03.004.
Haddad, J., Mirkin, B., 2020. Resilient perimeter control of macroscopic fundamental diagram networks under cyberattacks. Transportation Research Part B: Methodological 132, 44–59. doi:10.1016/j.trb.2019.01.020. 23rd International Symposium on Transportation and Traffic Theory (ISTTT 23).
Haddad, J., Zheng, Z., 2020. Adaptive perimeter control for multi-region accumulation-based models with state delays. Transportation Research Part B: Methodological 137, 133–153. doi:10.1016/j.trb.2018.05.019. advances in Network Macroscopic Fundamental Diagram (NMFD) Research.
Haitao, H., Yang, K., Liang, H., Menendez, M., Guler, S.I., 2019. Providing public transport priority in the perimeter of urban networks: A bimodal strategy. Transportation Research Part C: Emerging Technologies 107, 171–192. doi:10.1016/j.trc.2019.08.004.
Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204, 148–151. doi:10.1126/science.204.4389.148.
Hu, G., Lu, W., Wang, F., Whalin, R.W., 2020. Macroscopic fundamental diagram based discrete transportation network design. Journal of Advanced Transportation 2020.
Ingole, D., Mariotte, G., Leclercq, L., 2020. Perimeter gating control and citywide dynamic user equilibrium: A macroscopic modeling framework. Transportation Research Part C: Emerging Technologies 111, 22 – 49. doi:10.1016/j.trc.2019.11.016.
Jin, W.L., Gan, Q.J., Gayah, V.V., 2013a. A kinematic wave approach to traffic statics and dynamics in a double-ring network. Transportation Research Part B: Methodological 57, 114–131. URL: http://dx.doi.org/10.1016/j.trb.2013.09.004, doi:10.1016/j.trb.2013.09.004.
Jin, W.L., Gan, Q.J., Gayah, V.V., 2013b. A kinematic wave approach to traffic statics and dynamics in a double-ring network. Transportation Research Part B: Methodological 57, 114–131. doi:10.1016/j.trb.2013.09.004.
Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., Papageorgiou, M., 2012. Exploiting the fundamental diagram of urban networks for feedback-based gating. Transportation Research Part B: Methodological 46, 1393–1403. doi:10.1016/j.trb.2012.06.008.
Knoop, V.L., Van Lint, H., Hoogendoorn, S.P., 2015. Traffic dynamics: Its impact on the macroscopic fundamental diagram. Physica A: Statistical Mechanics and its Applications 438, 236–250. doi:10.1016/j.physa.2015.06.016.
Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. Transportation Research Part B: Methodological 96, 26 – 45. doi:10.1016/j.trb.2016.10.011.
Laval, J.A., Castrillón, F., 2015. Stochastic approximations for the macroscopic fundamental diagram of urban networks. Transportation Research Part B: Methodological 81, 904–916. doi:10.1016/j.trb.2015.09.002.
Laval, J.A., Chilukuri, B.R., 2016. Symmetries in the kinematic wave model and a parameter-free representation of traffic flow. Transportation Research Part B: Methodological 89, 168–177.
Laval, J.A., Leclercq, L., Chiabaut, N., 2017. Minimal parameter formulations of the dynamic user equilibrium using macroscopic urban models: Freeway vs city streets revisited. Transportation Research Procedia 23, 517–530. doi:10.1016/j.trpro.2017.05.029.
Leblanc, L.J., 1975. An algorithm for the discrete network design problem. Transportation Science 9, 183–199.
Leclercq, L., Geroliminis, N., 2013. Estimating mfds in simple networks with route choice. Transportation Research Part B: Methodological 57, 468–484. doi:10.1016/j.trb.2013.05.005.
Leclercq, L., Paipuri, M., 2019. Macroscopic Traffic Dynamics Under Fast-Varying Demand. Transportation Science 53, 1526–1545. doi:10.1287/trsc.2019.0908.
Leclercq, L., Parzani, C., Knoop, V.L., Amourette, J., Hoogendoorn, S.P., 2015. Macroscopic traffic dynamics with heteroge-

neous route patterns. Transportation Research Part C: Emerging Technologies 59, 292–307. doi:10.1016/j.trc.2015.05.006.
Li, Y., Yildirimoglu, M., Ramezani, M., 2021. Robust perimeter control with cordon queues and heterogeneous transfer flows. Transportation Research Part C: Emerging Technologies 126, 103043. doi:10.1016/j.trc.2021.103043.
Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2019. Understanding traffic capacity of urban networks. Scientific Reports 9, 16283. doi:10.1038/s41598-019-51539-5.
Mahmassani, H., Williams, J.C., Herman, R., 1987. Performance of urban traffic networks, in: Proceedings of the 10th International Symposium on Transportation and Traffic Theory, pp. 1–20.
Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. Transportation Research Part C: Emerging Technologies 36, 480–497. doi:10.1016/j.trc.2013.07.002.
Mariotte, G., Leclercq, L., Batista, S., Krug, J., Paipuri, M., 2020. Calibration and validation of multi-reservoir mfd models: A case study in lyon. Transportation Research Part B: Methodological 136, 62–86. doi:10.1016/j.trb.2020.03.006.
Mariotte, G., Leclercq, L., Laval, J.A., 2017. Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. Transportation Research Part B: Methodological 101, 245–267. doi:10.1016/j.trb.2017.04.002.
Mazloumian, A., Geroliminis, N., 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, 4627–4647. doi:10.1098/rsta.2010.0099.
Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part i: General theory, part ii: Queueing at freeway bottlenecks, part iii: Multi-destination flows. Transportation Research Part B: Methodological 27, 281–313.
Ni, W., Cassidy, M., 2019. City-wide traffic control: modeling impacts of cordon queues. Transportation Research Part C: Emerging Technologies doi:10.1016/j.trc.2019.04.024.
Sirmatel, I.I., Geroliminis, N., 2018. Mixed logical dynamical modeling and hybrid model predictive control of public transport operations. Transportation Research Part B: Methodological 114, 325 – 345. doi:10.1016/j.trb.2018.06.009.
Sirmatel, I.I., Geroliminis, N., 2020. Nonlinear moving horizon estimation for large-scale urban road networks. IEEE Transactions on Intelligent Transportation Systems 21, 4983–4994. doi:10.1109/TITS.2019.2946324.
Sirmatel, I.I., Tsitsokas, D., Kouvelas, A., Geroliminis, N., 2021. Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics. Transportation Research Part C: Emerging Technologies 128, 103157. doi:10.1016/j.trc.2021.103157.
Tilg, G., Ambühl, L., Batista, S.F., Menendez, M., Busch, F., 2021. On the application of variational theory to urban networks. Transportation Research Part B: Methodological 150, 435–456. doi:10.1016/j.trb.2021.06.019.
Tilg, G., Amini, S., Busch, F., 2020a. Evaluation of analytical approximation methods for the macroscopic fundamental diagram. Transportation Research Part C: Emerging Technologies 114, 1 – 19. doi:10.1016/j.trc.2020.02.003.
Tilg, G., Ul Abedin, Z., Amini, S., Busch, F., 2020b. Simulation-based design of urban bi-modal transport systems. Frontiers in Future Transportation 1.
Xu, G., Yu, Z., Gayah, V.V., 2020. Analytical method to approximate the impact of turning on the macroscopic fundamental diagram. Transportation Research Record 2674, 933–947. doi:10.1177/0361198120933274.
Yang, K., Menendez, M., Zheng, N., 2019. Heterogeneity aware urban traffic control in a connected vehicle environment: A joint framework for congestion pricing and perimeter control. Transportation Research Part C: Emerging Technologies 105, 439 – 455. doi:10.1016/j.trc.2019.06.007.
Yang, K., Zheng, N., Menendez, M., 2018. Multi-scale perimeter control approach in a connected-vehicle environment. Transportation Research Part C: Emerging Technologies 94, 32 – 49. doi:10.1016/j.trc.2017.08.014. iSTTT22.
Yildirimoglu, M., Sirmatel, I.I., Geroliminis, N., 2018. Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. Transportation Research Part B: Methodological 118, 106 – 123. doi:10.1016/j.trb.2018.10.007.
Zheng, N., Rérat, G., Geroliminis, N., 2016. Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. Transportation Research Part C: Emerging Technologies 62, 133–148. doi:10.1016/j.trc.2015.10.015.

Appendices

A. Nomenclature

Table 1 summarizes the nomenclature used in this paper. They are categorized into greek letters, as well as lower-case and capital latin letters.

Table 1: Nomenclature

Variable	Unit	Description
$\alpha \in A$	[-]	Turning ratio in set of turning ratios
$\tilde{\alpha}$	[-]	Factor impacting the queue growth due to turning flows

Table 1: Nomenclature (continued)

β	[-]	Time-dependent capacity constraint representing traffic signals
Δt	[h]	Numerical step length in the temporal dimension
Δx	[km]	Numerical step length in the spatial dimension
ΔN	[veh]	Cycle-based difference in supply and demand
δN	[veh]	Cycle-based difference in supply and demand at a signal phase change
κ	[veh/km]	Density
κ_{max}	[veh/km]	Jam density
κ_{opt}	[veh/km]	Optimal density
κ'	[veh/km]	Transformed density
θ	[-]	Ratio of the free-flow speed u and the backward wave speed w
τ	[s]	Time between beginning of active phase and t_e
σ	[h]	Duration of spillback impact
λ	[-]	Inflow factor at origin links
γ	[-]	Factor for approximating congestion propagation
c	[s]	Cycle length
c_f	[-]	Cycle when the final vehicle reaches the intersection
g	[s]	Duration of a green phase
i, j	[-]	Indices for corridors
k	[-]	Index for intersections and links
l	[km]	Link length
m	[-]	Index for moving observer
o	[s]	Offset
p	[-]	Valid path
q	[veh/h]	Flow
q_f	[veh/h]	Flow during the cycle c_f
q_{max}	[veh/h]	Link capacity
$\tilde{q}$	[veh/h]	Maximum cycle-average flow during green phase
$\tilde{q}_d$	[veh/h]	Flow $\tilde{q}$ constrained by demand only
$\tilde{q}_s$	[veh/h]	Flow $\tilde{q}$ constrained by supply only
$\tilde{q}_{k,s}^{j\rightarrow i}$	[veh/h]	Flow $\tilde{q}$ on C^i constrained by spillbacks from C^j
r	[s]	Duration of a red phase
t	[h]	Temporal dimension
t_b	[h]	Beginning of spillback impact
t_e	[h]	End of spillback impact
t_f	[h]	Time instant when the final vehicle leaves the intersection.
$t_{k,dq}$	[h]	Time instant when the queue starts to grow at the downstream end of the link k
$t_{k,uq}$	[h]	Time instant when the queue reaches the upstream end of the link k
u	[km/h]	Free-flow speed
v	[km/h]	Speed
w	[km/h]	Backward wave speed
x	[km]	Spatial dimension
z_p	[veh]	Cost of path p
B	[-]	Curve in time and space along which the boundary data N_B is given
$C \in \mathcal{C}$	[-]	Corridor in set of corridors
$\mathcal{G}$	[-]	Hypernetwork
$I \in \mathcal{I}$	[-]	Intersection in set of intersections
K	[veh/km]	Network-wide average density
K_k	[veh/km]	Maximum density on the link k.

Table 1: Nomenclature (continued)

K_{max}	[veh/km]	Network-wide jam density
$L \in \mathcal{L}$	[-]	Link in set of links
N	[veh]	Cumulative count
N_c	[veh]	Maximum number of vehicles passing an intersection during a cycle
N_e	[veh]	Cumulative count at the end of a green phase
N_s	[veh]	Cumulative count at the start of a green phase
$\mathcal{N}$	[-]	Physical road network
$P(x,t)$	[-]	Generic point in time and space
$\mathcal{P}$	[-]	Set of valid paths from B to P
Q	[veh/h]	Network-wide average flow
$R(v_m)$	[veh]	Average maximum passing flow for a moving observer m
Z_m	[veh]	Sum of all costs along the path of a moving observer m

B. Network variational theory formulations

This section provides a short summary of the mathematical formulation on which nVT (Tilg et al., 2021) builds. As described in Section 4.3.3, nVT proposes to exogenously account for the effects of source terms at the points P which are exactly at, and right downstream of an intersection I_k, i.e. at $x = \{x_k, x_k + \Delta x\}$.

For points at $x = x_k + \Delta x$ the inflows from or outflows to adjacent corridors need to be considered. Downstream traffic states, as well as capacity constraints are taken into account as in the classical VT formulation. The corresponding formulation is:

$$\begin{aligned} N^i(x_k + \Delta x, t) = \min(&N^i(x_k, t - \Delta t) \cdot (1 - \alpha^{ij}) + N^j(x_k, t - \Delta t) \cdot \alpha^{ji}, \\ &N^i(x_k + 2\Delta x, t - \theta\Delta t) + \Delta x \kappa_{max}, \\ &N^i(x_k + \Delta x, t - \Delta t) + \beta). \end{aligned} \tag{31}$$

For points at $x = x_k$, one needs to properly account for congestion propagation from downstream. Upstream traffic states, as well as capacity constraints are taken into account as in the classical VT formulation. As nVT assumes a FIFO diverge, congestion from the adjacent corridor is considered as well. For the sake of readability, the equation is split into two:

$$\begin{aligned} \Delta N^{ii} &= \left(N^i(x_k, t - (1+\theta)\Delta t) \cdot (1 - \alpha^{ij}) + N^j(x_k, t - (1+\theta)\Delta t) \cdot \alpha^{ji}\right) - N^i(x_k + \Delta x, t - \theta\Delta t), \\ \Delta N^{ij} &= \left(N^i(x_k, t - (1+\theta)\Delta t) \cdot \alpha^{ij} + N^j(x_k, t - (1+\theta)\Delta t) \cdot (1 - \alpha^{ji})\right) - N^j(x_k + \Delta x, t - \theta\Delta t). \end{aligned} \tag{32}$$

$$\begin{aligned} N^i(x_k, t) = \min(&N^i(x_k - \Delta x, t - \Delta t), \\ &N^i(x_k, t - (1+\theta)\Delta t) + \frac{1}{1 - \alpha^{ij}} \cdot \max\left(0, \Delta x \kappa_{max} \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot \alpha^{ji}\right) - \Delta N^{ii}\right), \\ &N^i(x_k, t - (1+\theta)\Delta t) + \frac{1}{\alpha^{ij}} \cdot \max\left(0, \Delta x \kappa_{max} \cdot \left(1 - \frac{r}{(1+\theta)\Delta t} \cdot (1 - \alpha^{ji})\right) - \Delta N^{ij}\right), \\ &N^i(x_k, t - \Delta t) + \beta). \end{aligned} \tag{33}$$

We refer the interested reader to Tilg et al. (2021) for additional details on the derivation of these equations.

C. Derivation of $\tilde{\alpha}$

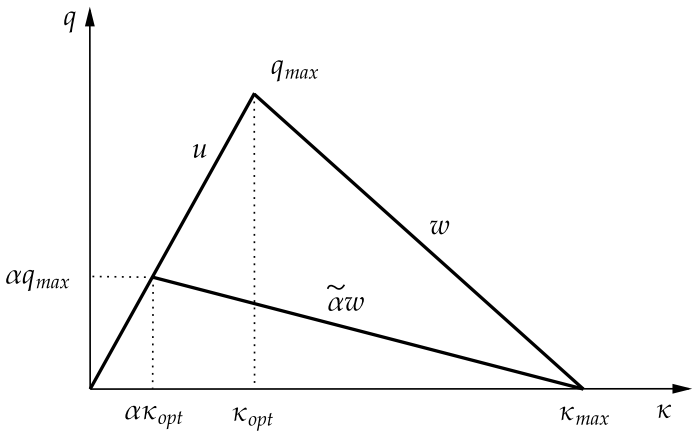


Figure 16: Illustration of $\tilde{\alpha}$ in the FD.

The maximum flow q_{max} can be written as $q_{max} = u\kappa_{opt}$ and $q_{max} = w(\kappa_{max} - \kappa_{opt})$. From this follows that:

$$u = w\frac{\kappa_{max} - \kappa_{opt}}{\kappa_{opt}}. \tag{34}$$

Furthermore, from the fundamental diagram as depicted in Figure 16 we can write:

$$\alpha q_{max} = (\kappa_{max} - \alpha\kappa_{opt})\tilde{\alpha}w. \tag{35}$$

Then, we can reformulate this equation to:

$$\tilde{\alpha} = \frac{\alpha q_{max}}{w(\kappa_{max} - \alpha\kappa_{opt})} \tag{36}$$

By putting $q_{max} = u\kappa_{opt}$, diving by k_{opt}, and reformulating the denominator, and finally utilizing eq.(34), we get the following:

$$\begin{aligned}
\tilde{\alpha} &= \frac{\alpha q_{max}}{w(\kappa_{max} - \alpha\kappa_{opt})} \\
&= \frac{\alpha u\kappa_{opt}}{w(\kappa_{max} - \alpha\kappa_{opt})} \\
&= \frac{\alpha u}{\frac{w}{\kappa_{opt}}(\kappa_{max} - \alpha\kappa_{opt})} \\
&= \frac{\alpha u}{\frac{w}{\kappa_{opt}}(\kappa_{max} - \kappa_{opt} + (1-\alpha)\kappa_{opt})} \\
&= \frac{\alpha u}{\frac{w}{\kappa_{opt}}(\kappa_{max} - \kappa_{opt}) + \frac{w}{\kappa_{opt}}(1-\alpha)\kappa_{opt}} \\
&= \frac{\alpha u}{\frac{w}{\kappa_{opt}}(\kappa_{max} - \kappa_{opt}) + w(1-\alpha)} \\
&= \frac{\alpha u}{u + w(1-\alpha)}
\end{aligned} \tag{37}$$

This concludes the derivation of $\tilde{\alpha}$ as used in the paper.