Advancements in Macroscopic Modelling of Persistent and Non-Persistent Delays at Intersections

Xiaolin Gong

Institute of Transport and Logistics Studies

University of Sydney Business School

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

October 2024

Acknowledgements

When I began my PhD studies, I imagined the day I would write my acknowledgments, but words alone cannot capture the emotions I feel at this moment. My PhD journey commenced during the pandemic, with the first 1.5 years spent conducting research remotely from China. Here, I would like to express my heartfelt gratitude to many individuals who have offered their help and encouragement.

First and foremost, I would like to express my sincere gratitude to my supervisors, Professor Michiel Bliemer and Dr. Mark Raadsen. I am often grateful to have such patient and supportive supervisors. I thank them for their care and connections, which ensured I did not feel isolated or lost during my remote studies. I would like to thank their encouragement and unwavering support, which rekindled my motivation and confidence whenever I doubted myself. I am grateful to their guidance and assistance—every discussion, equation derived, feedback and comment—have been invaluable. Their passion for academic research will continue to inspire me.

Then, I want to extend my thanks to all the members of the Institute of Transport and Logistics Studies (ITLS) family. I am thankful to ITLS for providing such a warm and friendly atmosphere for research and learning, and for being supportive whenever I need assistance. I am grateful for every interesting discussion and every thought-provoking meeting with academics and fellow students.

Additionally, I am grateful to Professor Michiel Bliemer for generously funding my participation in international conferences. I would like to acknowledge the financial support from the Enhanced Business School Research Scholarship, the Research Travel Support Scheme from the Business School at the University of Sydney, and the Postgraduate Research Support Scheme from the University of Sydney for completing this thesis.

I am deeply grateful to my parents, Mr. Gong Xiangrong and Ms. Dong Wenqing, for their unwavering support and trust. To my partner, Mr. Ni Lifeng, I am grateful for being my “psychologist” throughout my PhD studies. His constant love and encouragement made this possible.

Last but not least, I would like to thank myself. You did it.

Authorship attribution statement

The work delivered in the body of this thesis, except otherwise acknowledged, is the result of my own research investigations. The following journal papers, which were submitted to journals during my PhD candidature, comprise this thesis. The authorship attribution statement for each paper is provided as below.

Chapter 2 includes the paper:

Gong, X., Raadsen, M.P.H. & Bliemer, M.C.J. (2024) A systematic review of node models for macroscopic network loading of traffic flows. Submitted to *Transportation Research Part B: Methodological.*

CRediT author statement:

Xiaolin Gong: Conceptualisation, Methodology, Investigation, Writing – original draft, Writing – Review & Editing, Visualisation. **Mark Raadsen**: Conceptualisation, Investigation, Writing – Review & Editing, Supervision. **Michiel Bliemer**: Conceptualisation, Investigation, Writing – Review & Editing, Supervision.

Chapter 3 includes the paper:

Gong, X., Bliemer, M.C.J. & Raadsen, M.P.H. (2024) Extended macroscopic node model for multilane traffic. Submitted to *Transportation Research Part B: Methodological*.

CRediT author statement:

Xiaolin Gong: Conceptualisation, Methodology, Software, Formal analysis, Writing – original draft, Writing – Review & Editing, Visualisation. **Michiel Bliemer**: Conceptualisation, Methodology, Formal analysis, Writing – original draft, Writing – Review & Editing, Visualisation, Supervision. **Mark Raadsen**: Conceptualisation, Methodology, Software, Writing – Review & Editing, Supervision.

Chapter 4 includes the paper:

Gong, X., Bliemer, M.C.J. & Raadsen, M.P.H. (2024) Incorporation of endogenous exit capacities in node models to account for varying movement exit speeds. Submitted to *Transportmetrica A: Transport Science*.

CRediT author statement:

Xiaolin Gong: Conceptualisation, Methodology, Formal analysis, Writing – original draft, Writing – Review & Editing, Visualisation. **Michiel Bliemer**: Conceptualisation, Methodology, Writing – Review & Editing, Supervision. **Mark Raadsen**: Methodology, Writing – Review & Editing, Supervision.

Chapter 5 includes the paper:

Bliemer, M.C.J. **Gong, X.** & Raadsen, M.P.H. (2024) Incorporation of non-persistent delays at signalised intersections in the link transmission model. Submitted to *Transportation Research Part B: Methodological*.

CRediT author statement:

Michiel Bliemer: Conceptualisation, Methodology, Formal analysis, Writing – original draft, Writing – Review & Editing, Visualisation, Supervision. **Xiaolin Gong**: Conceptualisation, Methodology, Software, Formal analysis, Writing – original draft, Writing – Review & Editing, Visualisation. **Mark Raadsen**: Conceptualisation, Methodology, Writing – Review & Editing, Supervision.

Part of this thesis has been presented at the following conference and symposiums:

- **Gong, X.**, Bliemer, M.C.J. & Raadsen, M.P.H. (2023). *Lane Choice Equilibrium in Macroscopic Multi-Queue Node Model for Multilane Intersections.* 9th International Symposium on Dynamic Traffic Assignment, Chicago. (Presentation)
- **Gong, X.**, Bliemer, M.C.J. & Raadsen, M.P.H. (2023). *Lane Choice Equilibrium in Macroscopic Multi-Queue Node Model for Multilane Intersections*. 102nd Transportation Research Board Annual Meeting, Washington, DC. (Poster)
- **Gong, X.**, Bliemer, M.C.J. & Raadsen, M.P.H. (2023). Incorporation of Non-Persistent Delays in Macroscopic Network Modeling. Transport Research Association for New South Wales Symposium 2023. (Presentation)
- **Gong, X.**, Bliemer, M.C.J. & Raadsen, M.P.H. (2022). Macroscopic Multi-queue Node Model for Signalised Multilane Intersections. Transport Research Association for New South Wales Symposium 2022. (Presentation)
- **Gong, X.**, Bliemer, M.C.J. & Raadsen, M.P.H. (2021). Extended node model with lane choice equilibrium. Transport Research Association for New South Wales Symposium 2021. (Presentation)
- **Gong, X.**, Bliemer, M.C.J. & Raadsen, M.P.H. (2020). Improving forecasting of intersection delays in macroscopic transport models. Transport Research Association for New South Wales Symposium 2020. (Presentation)

Xiaolin Gong

28th June 2024

Statement of originality

This is to certify that to the best of my knowledge; the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Xiaolin Gong

28th June 2024

As a supervisor for the candidature upon which this thesis is based, I confirm that the authorship attribution statements and statement of originality in this thesis are correct.

I confirm that this thesis is sufficiently well presented to be examined. I confirm that this thesis does not exceed the prescribed word limit or any extended word limit for which prior approval has been granted.

prof. dr. Michiel C.J. Bliemer	dr. Mark P.H. Raadsen
28th June 2024	28th June 2024

Abstract

The ability to forecast travel times in road networks is important for transport planning and traffic management. On urban roads, travel times are predominantly influenced by intersection delays, which consist of persistent and non-persistent delays. Non-persistent delays refer to temporary delays experienced by vehicles in short queues waiting for a red signal to turn green at a signalised intersection. Persistent delays arise when traffic demand exceeds capacity, forcing vehicles to wait in a persistent queue at an unsignalised intersection or a signalised intersection for more than one cycle. Traffic assignment and simulation models are widely used to forecast traffic flows and travel times in transport networks. Network loading (NL) constitutes a core component of traffic assignment and propagates traffic flow from origins to destinations. As opposed to microscopic models, macroscopic NL models are computationally efficient, suitable for large networks, and implementable in real-time traffic management applications. A macroscopic NL model consists of a link model and a node model. A link model propagates traffic states across a road segment like waves in water, whereas a node model manages the interaction of competing flows at intersections and imposes inflow and outflow restrictions, resulting in queues and delays on upstream links. Accurate forecasting of traffic conditions and travel times necessitates realistic link and node models.

Existing macroscopic NL models cannot effectively address both persistent and non-persistent delays. Persistent delays are inadequately modelled by existing node models due to their incapability of (i) accounting for the influence of approach lane configurations, and (ii) consideration of endogenous exit link capacities due to differences in exit speeds for different movements. Moreover, non-persistent delays are generally completely ignored in macroscopic NL. This thesis addresses these deficiencies through a comprehensive macroscopic intersection modelling approach. Firstly, a systematic literature review on the development of node models is conducted. Subsequently, this thesis proposes three advancements, namely: (i) an innovative macroscopic node model that allows for separate queues for each turning movement, while considering approach lane configurations, (ii) an approach to incorporate endogenous exit link capacities in node models, and (iii) a method to endogenously incorporate non-persistent delays into the NL propagation scheme. Each advancement is mathematically formulated, and efficient solution algorithms are developed. The capability of these approaches is tested through numerical examples, demonstrating their ability to improve travel time forecasts in urban intersections. The advancements proposed in this thesis not only enhance the current modelling techniques but also provide a foundation for future research in macroscopic transport modelling and real-time traffic management applications.

Keywords: macroscopic intersection modelling, node model, link transmission model, persistent delay, non-persistent delay

Table of Contents

List of Figures

Chapter 1 Introduction

Chapter 2 Literature Review

Chapter 3 Node Model for Multilane Traffic

Chapter 4 Node Model Incorporates Endogenous Exit Capacities

Chapter 5 Non-persistent Delays in LTM

List of Tables

Chapter 2 Literature Review

Chapter 3 Node Model for Multilane Traffic

Chapter 4 Node Model Incorporates Endogenous Exit Capacities

Chapter 5 Non-persistent Delays in LTM

1 Chapter 1 Introduction

This thesis is formed by my publications on the topic of macroscopic intersection modelling. There is a specific focus on developing novel methodologies to better incorporate persistent and non-persistent delays in node models and their deployment in macroscopic network loading models.

Each chapter reflects the contents of a paper on a topic within this thesis' theme. The introduction and conclusion chapters are not related to any submitted papers. They instead provide introductory context, motivation, and interpretation of the separate works that are included.

1.1 Background and scope

Urban transportation is the hub and artery of social and economic activities, but rapid population growth and increased car ownership have brought increased traffic problems such as congestion, accidents, and air pollution. To mitigate the impact of these issues, *traffic forecasting* provides a way to justify investment in transportation infrastructure to ensure demand and supply are coordinated now and in the future. Traffic forecasting is often applied within the field of *transport planning* to determine the long-term travel time savings of future road infrastructure for appraisal purposes. For example, predicting travel time savings after building a west connection tunnel or a new train line in Sydney to assess whether the investment is well spent and provides benefits. In contrast to long term planning, *traffic management* has similar objectives – reducing congestion and travel times but is focussed more on short-term predictions utilising existing infrastructure. Often this involves optimising traffic controls. For example, adjust signal timings at an intersection in the morning peak based on observed travel times over the past few hours, days, or weeks. In both cases, travel time (delay) is the most important output, used to make either long- or short-term decisions. An accurate prediction of this travel time is therefore paramount. This thesis provides several novel methodologies in the context of intersection modelling with the objective to predict traffic flows and travel times more accurately. This hopefully then will lead to uptake by practitioners to incorporate these methods in forecasting in both transport planning and traffic management applications to benefit policymakers and society in general.

Given the importance of accurately capturing delays in our models, let us first introduce the main types of delays and their terminology. When a vehicle traverses a road, and it experiences no delay it is generally assumed they travel with a speed equal to the road's speed limit. Any speed lower than this speed limit can be considered a delay. The mildest form of delay is termed *hypocritical delay*. This refers to delay caused by vehicles moving at speeds lower than the maximum legal speed in under-saturated (hypocritical) conditions, meaning there are no queues but due to surrounding traffic the vehicle is no longer able to travel at its desired speed. This in contrast to *hypercritical delay*. Hypercritical delay occurs only in over-saturated conditions, meaning that there is queue formation and congestion due to the traffic flow demand exceeding the road's capacity. Hypocritical and hypercritical delays are usually considered in the context of vehicles traversing a road, rather than vehicles traversing an intersection. The underlying cause of hypercritical delay on a road however is usually a downstream intersection (or bottleneck). We can therefore disentangle the causes of hypercritical delay further when considering the situation at intersections.

Intersection delays are often the primary source of delay in networks and have two principal components, namely persistent and non-persistent delay. *Persistent delay*, also referred to as overflow delay, occurs when traffic demand exceeds road capacity, leading to queue buildup until demand reduces, i.e., queuing delay on over-saturated highways. At signalised intersections, this situation arises when vehicles wait through multiple cycles before passing, in other words, when a queue does not dissipate fully within a single green cycle. In other scenarios, persistent delay may be the result from lane drops or other physical changes to the road. *Non-persistent delays* occur during short, temporary disruptions in traffic flow without causing systemic queues. At signalised intersections, this happens when vehicles temporarily wait for a red signal. While this does result in a queue, such non-persistent queues vanish entirely during the next green phase, also known as uniform delays. Non-persistent delays can also occur in other situations, such as waiting for pedestrians to cross or yielding at unsignalised intersections, also referred to as give-way delays. As long as these queues do not continue to grow but quickly resolve under stable demand, they are considered non-persistent. Fig. 1 illustrates non-persistent and persistent delays graphically through a cumulative inflow/outflow diagram for a hypothetical link with a traffic signal on its downstream border.

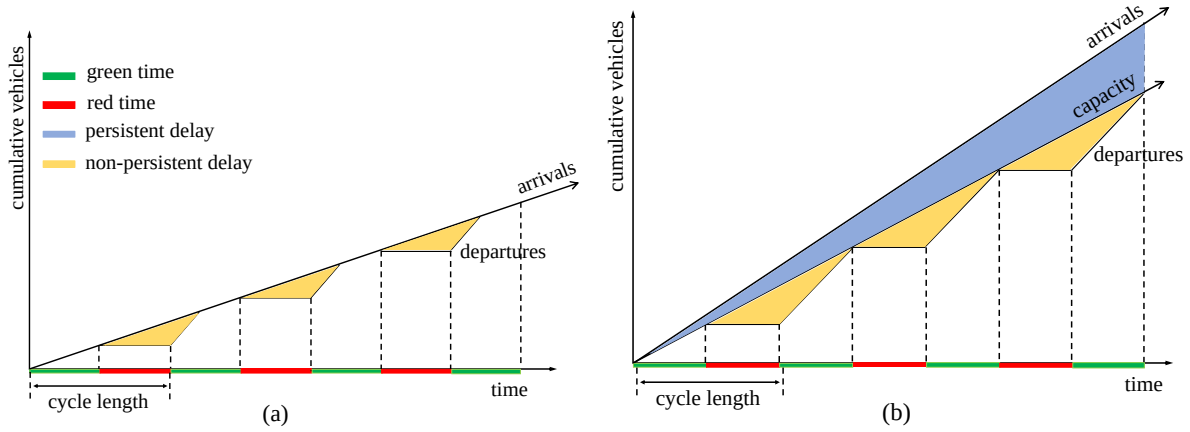


Fig. 1. Cumulative vehicles and delay accumulation at traffic signal for, (a) uncongested situation with non-persistent delay only, (b) congested situation, with both persistent and non-persistent delay.

This thesis focuses on improving intersection modelling in the context of both persistent and non-persistent delays. Some of the proposed methods can also be utilised to improve existing approaches regarding the prediction of congestion delays on motorways at on-ramps (merges) and off-ramps (diverges).

Intersection modelling does not occur in isolation but is part of a larger modelling framework, most commonly referred to as *traffic assignment* (TA). TA is an important component used in both transport planning and traffic management. It distributes traffic demand across the infrastructure network. Both the travel demand and network are typically inputs. This thesis focuses on private road transport (i.e., cars) within road infrastructure networks composed of road segments and intersections. TA models typically include a route choice model and a network loading model. The *route choice model* allocates traffic flows across feasible routes based on (generalised) route costs, while *the network loading (NL) model* propagates these path flows from origin to destination.

In transport planning, updated route costs generated by the NL model are often fed back into the route choice model, continuing this process until equilibrium is reached (route choice is deemed consistent with the network loading route cost). Conversely, for traffic management, determining equilibrium is generally not the objective, and a single (or limited number of) iteration(s) of the TA model typically suffice.

There are three main types of TA models: microscopic, mesoscopic, and macroscopic. *Microscopic* models describe the movement of individual vehicles through the network, whereas *macroscopic* models represent traffic as continuous flows, akin to water flowing through a pipe. *Mesoscopic* models reside somewhere in between. Microscopic models capture detailed interactions between vehicles but are computationally expensive, while macroscopic models are computationally attractive and can be applied on large-scale networks, but they lack the detail of microscopic models.

For transport planning purposes, *stability* and *equilibrium convergence* are important features to enable scenario comparison and cost-benefit analysis. Stability refers to the ability of traffic prediction in the absence of random components. Equilibrium convergence pertains to the model's ability to find a state (i.e., a fixed point) where the route choice model and the NL model are consistent (i.e., in equilibrium) regarding generalised travel costs and flows. Macroscopic TA models are more stable (due to lack of random components) and have more attractive mathematical properties, helpful in reaching equilibrium convergence, compared to microscopic models. Given our focus on – long-term, large-scale - transport planning this thesis solely focuses on macroscopic modelling. Further, in this work, we consider route choice as exogenous and known, with a primary focus on NL.

Macroscopic NL models describe traffic flows in either a static or dynamic manner. *Static NL models* lack an endogenous consideration of time and are therefore typically used only for strategic transport planning in large metropolitan areas. In such cases, planners are mainly interested in average network traffic conditions rather than detailed traffic flow dynamics. In contrast, *dynamic NL models*, capture traffic dynamics explicitly as part of the NL procedure and are therefore better suited for assessing temporal aspects of traffic states, such as queue build-up, queue dissipation, and spillback. Regardless of whether NL is static or dynamic, in its general form it comprises a *link model* – for the propagation of traffic flow on homogeneous roads – and a *node model* – for the distribution of competing flows at intersections, merges, diverges, and bottlenecks.

The contributions made in this thesis may benefit both static and dynamic NL approaches, include novel node model methodology, and extend existing link models to aid better intersection delay capturing. An overview of the TA modelling framework is provided in Fig. 2, where we highlighted the aspects of TA models containing contributions in bold.

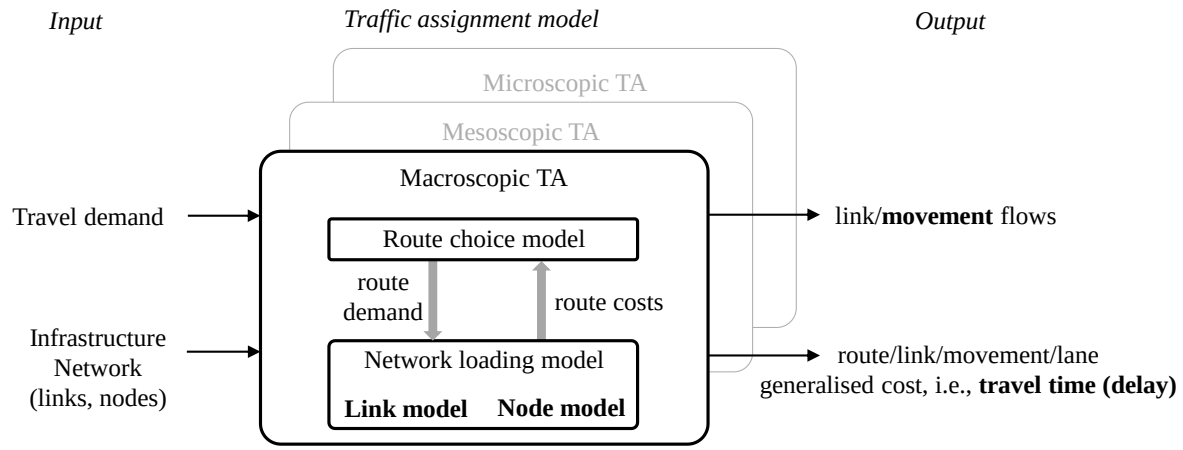


Fig. 2. TA modelling framework and research focus areas (in bold).

1.2 Research gaps

Within the scope of (static and dynamic) macroscopic NL models, it was found that, while attractive in terms of stability and convergence, there remain significant drawbacks particularly in their support for capable intersection delay modelling. This is especially apparent when looking at (non-)persistent delays at (signalised) multi-lane intersections. Fig. 3 loosely summarises the capabilities of exiting model types, including microscopic-dynamic, mesoscopic-dynamic, macroscopic-dynamic, and macroscopic-static models. These limitations are not just an academic issue, but a problem in practice as well as can be seen via the inclusion of the real-world macroscopic static model that is included here in the form of the Sydney Strategic Travel Model where flow exceeding capacity leads to over-saturated conditions on highways.

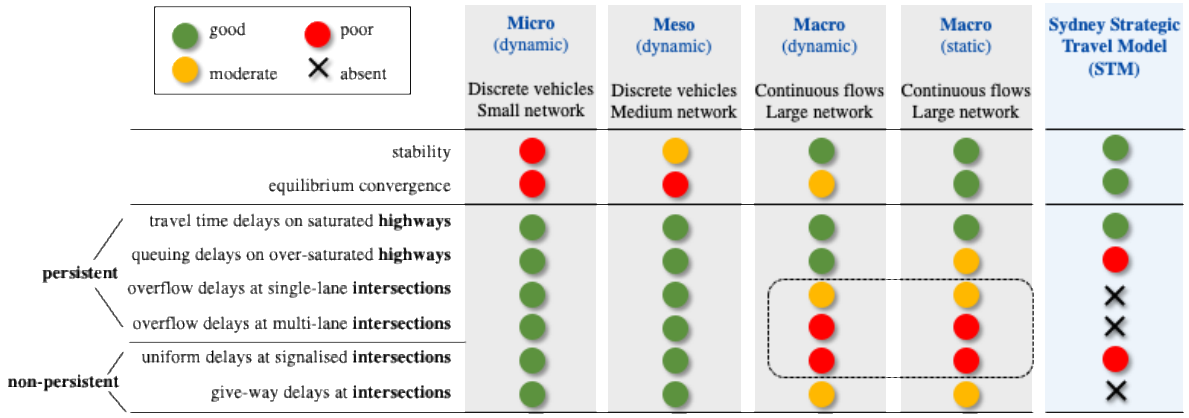


Fig. 3. Capabilities of state-of-the-art traffic assignment model types.

Let us now identify the four research gaps that are formulated as part of this work.

Gap 1 – Systematic review on node models

The first identified research gap pertains to the lack of a (systematic) review of existing node models. Substantial advancements in macroscopic node models have been made in the past decades, however, the literature lacks a paper providing an overview and classification of these developments.

Gap 2 – Multilane configuration

The second identified research gap concerns the inability of existing node models to accurately handle incoming links with multiple lanes.

Fig. 4 (a) shows an intersection with multiple approach lanes, where link 1 allows two movements, each, potentially with a separate queue. Traffic flow for each movement needs to be distributed across multiple approach lanes, considering that some lanes are movement-specific while others allow multiple movements. Fig. 4 (b) illustrates the representation of this intersection in in a typical macroscopic node model form, ignoring the underlying layout, and resulting in a single queue (if any) for all movements on the link. This is consistent with a link-level *first-in-first-out* (FIFO) assumption, where all vehicles exit their link in the same order as they entered. If one movement is blocked due to congestion or downstream spillback, then all other movements on the same link are also blocked. Clearly, in reality, multilane intersection approaches allow for separate queues for different movements. In other words, existing link-based FIFO based methods are too strict to be able to cater for multilane intersections effectively.

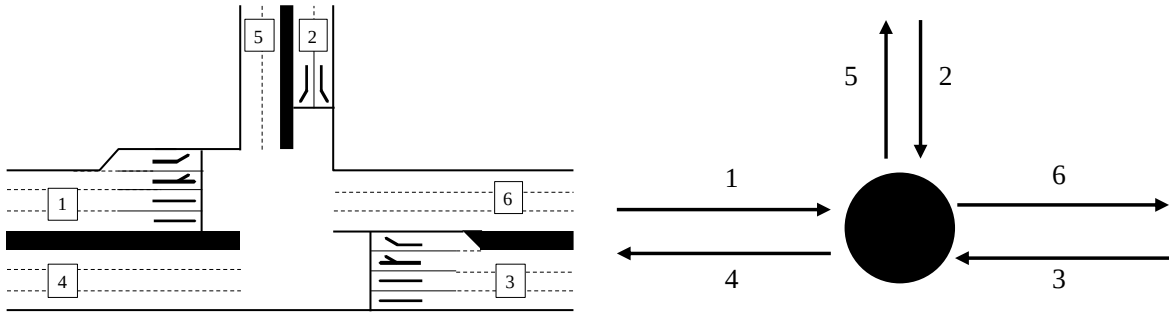


Fig. 4. (a) Example of an intersection with multiple approach lanes on urban roads, (b) typical macroscopic node model representation of same intersection.

We further note that circumventing this issue by altering the underlying network by replacing each link by multiple links (each lane becomes a link) is too naïve a solution. On a network level this exponentially increases the number of possible routes, making it near impossible to ever reach a route choice equilibrium.

Instead, there is a need for methodology such that route choice equilibrium can still be reached on the network level, while movement specific queues – and therefore relaxing link-based FIFO – can be modelled in a meaningful way.

Gap 3 – Achieve endogenous exit capacity in the node model

The third research gap involves the lack of consideration regarding endogenous exit capacities in node models. Current node models typically assume exogenous exit capacities, using fixed or pre-defined time-independent capacities. This assumption often neglects variations of exit capacities due to flow dynamics and variations in the composition of flow movements, such as reduced capacity due to a change in the proportion of vehicles making a sharp turn movement at low speed, resulting in less reliable predictions and reduced real-world applicability. Therefore, there is an opportunity to consider endogenous exit capacities in node models to improve their accuracy and practical relevance.

Consider a diverge node depicted in Fig. 5, featuring a single incoming link 1, two outgoing links 2 and 3, and two corresponding movements (1, 2) and (1, 3). If all vehicles on link 1 turn right, necessitating a reduction in speed, the exit capacity would be lower compared to a scenario where all vehicles proceed straight ahead at higher speeds. Consequently, the exit capacity on link 1 is influenced by the proportion of vehicles turning right and those going straight.
To quantify this, assume the saturation flow rate on link 1 is 1500 veh/h, corresponding to a saturation headway of 2.4 s. If half of the vehicles drive straight at high speed, maintaining a headway of 2.4 s, and the other half turn right, requiring a headway of 4.5 s, the exit capacity on link 1 would be $3600/(\frac{1}{2}\cdot 2.4+\frac{1}{2}\cdot 4.5)=1043$ veh/h. This scenario results in an approximately 30% reduction in capacity on link 1 due to the deceleration of vehicles making turns. In this thesis, a novel method will be developed to incorporate endogenous exit capacities within the node model, accounting for varying movement exit speeds. This approach aims to more accurately represent the impact of different vehicle manoeuvres on node capacity.

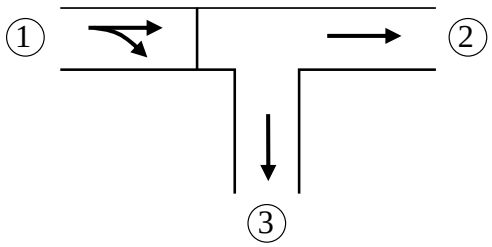


Fig. 5. Diverge node example with incoming/outgoing links and allowable movements.

Gap 4 – Incorporate non-persistent delay in NL models

The fourth gap relates to the challenge of integrating non-persistent delay modelling endogenously in macroscopic dynamic NL models with signalised intersections. Recall that non-persistent delay is the total of all the triangular areas between the arrival and departure flows in Fig. 1 (b). In that example, the arrival rate also exceeds capacity, resulting in a persistent queue building up in addition to the temporary non-persistent queue. Existing dynamic NL models can effectively capture persistent delays but typically ignore non-persistent delays at intersections, resulting in inaccurate overall travel delay forecasting.

This limitation is partially addressed by two distinct methodologies: the first method involves the explicit modelling of traffic control mechanisms, which includes the detailed specification

of red and green phases (Shirke et al., 2019). The second methodology encompasses the implicit consideration of average intersection delay, subsequently converting this delay into quantifiable vehicle queue numbers, a process outlined by Yperman & Tampère (2006) and Yperman (2007). The first approach, which explicitly simulates signal phases, offers an intuitive method for visualising non-persistent delays. Nevertheless, this method is seldom utilised in macroscopic transport modelling due to its high computational demands. The second approach, despite its effectiveness, is proposed in the form of an algorithm that lacks a theoretical foundation.

To illustrate the two approaches, consider a single lane link as shown in Fig. 6. Flow arriving at the intersection is near, but below, the link exit capacity and drops to a lower rate at time instant t'. Using kinematic wave theory (Lighthill & Whitham, 1955; Richards, 1956), traffic states and shockwaves are shown in Fig. 6(a) while considering red and green phases explicitly. Two primary shock waves are observed: the stopping shock wave, caused by the interaction between arriving and halted traffic, and the starting shock wave, generated as vehicles start moving at the saturation flow rate at the onset of the green phase (Wada et al., 2015). Non-persistent delay is then represented by the triangle areas. In the second approach, shown in Fig. 6(d) and representative for our study, the traffic signal is not explicitly considered but rather it accounts for the average uniform delay at the end of the link (shown in yellow) via a non-physical FIFO queue. This average delay is the same as if one would consider delays of the individual trajectories depicted in Fig. 6(a). In our proposed method, we account for fanning behaviour when flow increases, while adhering to FIFO and capacity constraints when flow decreases.

For completeness, two other approaches are shown in Fig. 6 too. Fig. 6(b) illustrates a typical DNL whereby non-persistent delay is ignored, and flow is treated as being in free-flow across the whole cycle. This would underestimate actual travel times in urban areas and in traffic assignment makes routes through urban centres more attractive than they really are. A possible other solution would be to simply time-shift outflows based on a non-persistent delay function, see Fig. 6(c). Such a naïve method will result in FIFO violations when flow rates drop and may also result in flow rates exceeding capacity, thereby resulting in unrealistic flow propagation. To numerically illustrate FIFO violations, imagine a vehicle arriving at a flow rate of 1000 veh/h with a 20s non-persistent delay. In 5s, another vehicle arrives at a reduced rate of 400 veh/h with a 10s non-persistent delay. The latter vehicle, despite arriving later, exits the link 5s earlier than the former vehicle, violating FIFO on a single lane.

This thesis demonstrates that it is possible to capture the average non-persistent delay without violating FIFO, while obviating the need to explicitly model signal phases, see Fig. 6(d). This average is the same as if one would consider delays of the individual trajectories depicted in Fig. 6(a). This thesis account for fanning behaviour when flow increases, while adhering to FIFO and capacity constraints when flow decreases. Importantly, the traffic flow dynamics in the DNL will remain consistent with the travel times applied in route choice, leading to a more stable TA, something that is not possible when adopting the current best-practice of adding in such delays in post-processing.

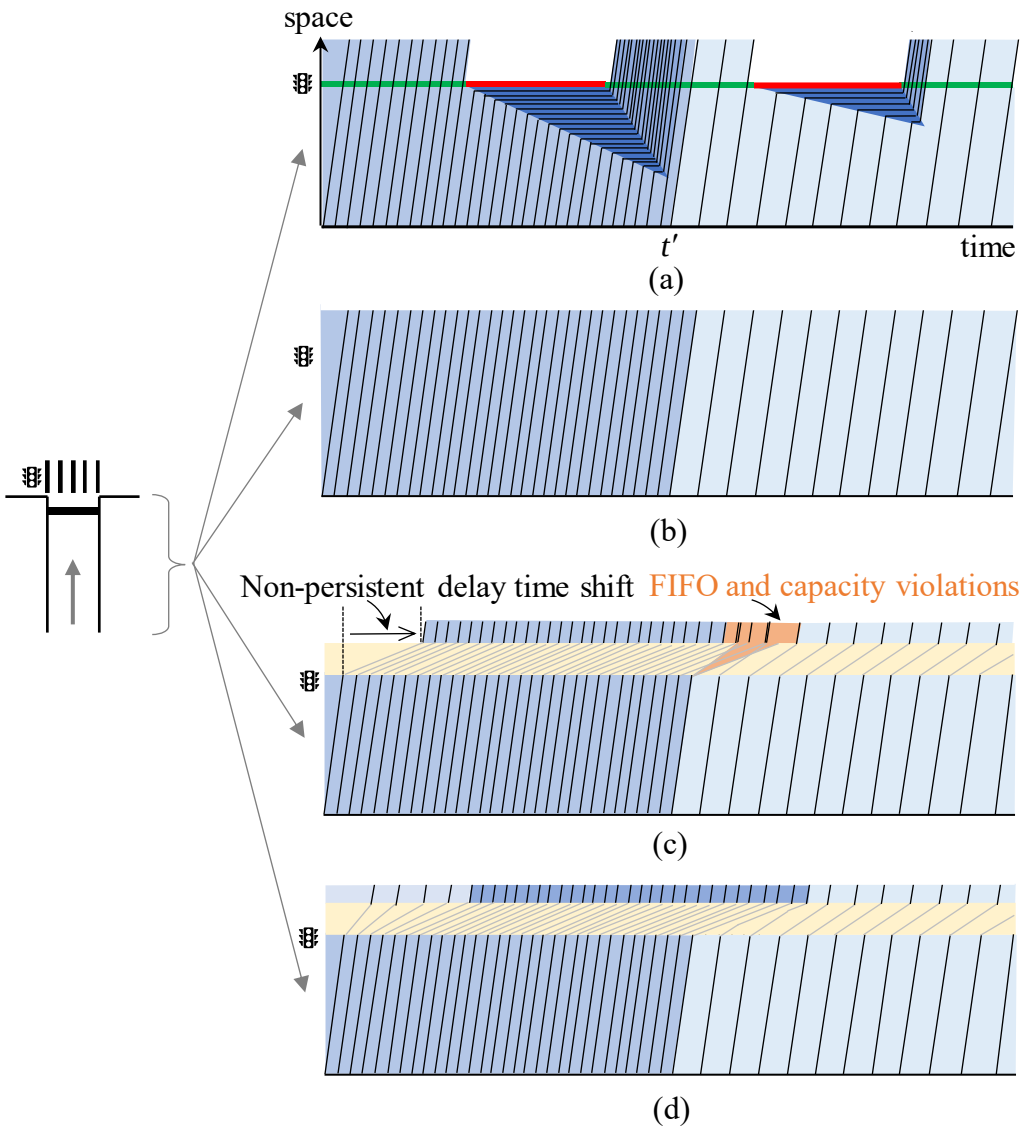


Fig. 6. Space-time diagram of first order traffic flow propagation at intersection with (a) explicitly simulating signal control (b) conventional DNL (c) naïve time-shift method (d) proposed novel DNL

1.3 Research questions

To address the four research gaps introduced in Section 1.2 and extract subsequent research objectives, we formulate one main research question (RQ) accompanied by four sub-research questions (SQs):

RQ: **How to improve persistent and non-persistent delay forecasts in macroscopic intersection modelling?**

SQ-1: What are the advancements and trends in the development of node models?

SQ-2: How can node models account for multilane configurations?

SQ-3: How to achieve endogenous exit capacities in node models?

SQ-4: How can non-persistent delays be incorporated within macroscopic dynamic NL models?

1.4 Research aims

The main aim is to address the RQ of enhancing the forecasting of both persistent and non-persistent delays in macroscopic intersection modelling. Specifically, this thesis aims to achieve this via the following four aims based on the proposed SQs.

Aim 1. Provide a comprehensive systematic review of the efforts in the development of node models.

This thesis aims to provide a comprehensive systematic review of the evolution of the node model within macroscopic NL frameworks. It seeks to classify existing node models according to distinct characteristics. Additionally, this thesis aims to highlight emerging trends, pinpoint future research directions, and identify opportunities, thereby providing guidance and insights for researchers and practitioners involved in the study of macroscopic node models.

Aim 2. Propose a multilane node model that considers approach lane configurations while relaxing stringent link-based FIFO to movement-based FIFO.

This thesis aims to formulate a novel (implicit lane-based) node model to relax link-based FIFO when incoming links of nodes have multiple lanes. The proposed model intends to account for the approach lane configuration and satisfies movement-based FIFO by introducing an equilibrium lane choice principle. Additionally, this thesis seeks to develop a general solution algorithm to determine equilibrium flows within the extended lane-based node model.

Aim 3. Propose a novel method to incorporate endogenous exit capacities within node models.

This thesis aims to propose a novel approach to incorporate endogenous exit capacities within node models to account for varying turn movement speeds that occur when vehicles traverse intersections. The primary goal of this approach is to moderate flow rates endogenously, rather than dynamically adjusting exit capacities. The presented approach aims to be compatible with existing node model frameworks as well as provide a solution algorithm.

Aim 4. Propose a novel approach to embed non-persistent delay generated at signalised intersections in link transmission models.

This thesis aims to introduce a novel methodology that directly incorporates non-persistent delays into various forms of the Link Transmission Model (LTM) formulation, ensuring the (average) impact of traffic lights is accounted for while satisfying the FIFO principle. By embedding non-persistent delays directly into the NL propagation scheme, the thesis aims to achieve consistency between route choice and flow propagation within transportation networks. In addition, this thesis intends to develop a solution algorithm to effectively implement the proposed approach.

1.5 Thesis structure and contributions

This section outlines the key content of each chapter, illustrating the coherence and connections between them, and providing an overview of the thesis structure as shown in Fig. 7. This chapter, Chapter 1, offered background information, introduced research questions, aims, and contributions. The core of the thesis is comprised of Chapters 2 - 5, with each dedicated to addressing SQs 1 - 4, respectively, in the form of individual papers. Chapter 6 concludes the thesis drawing conclusions and provides suggestions for future research.

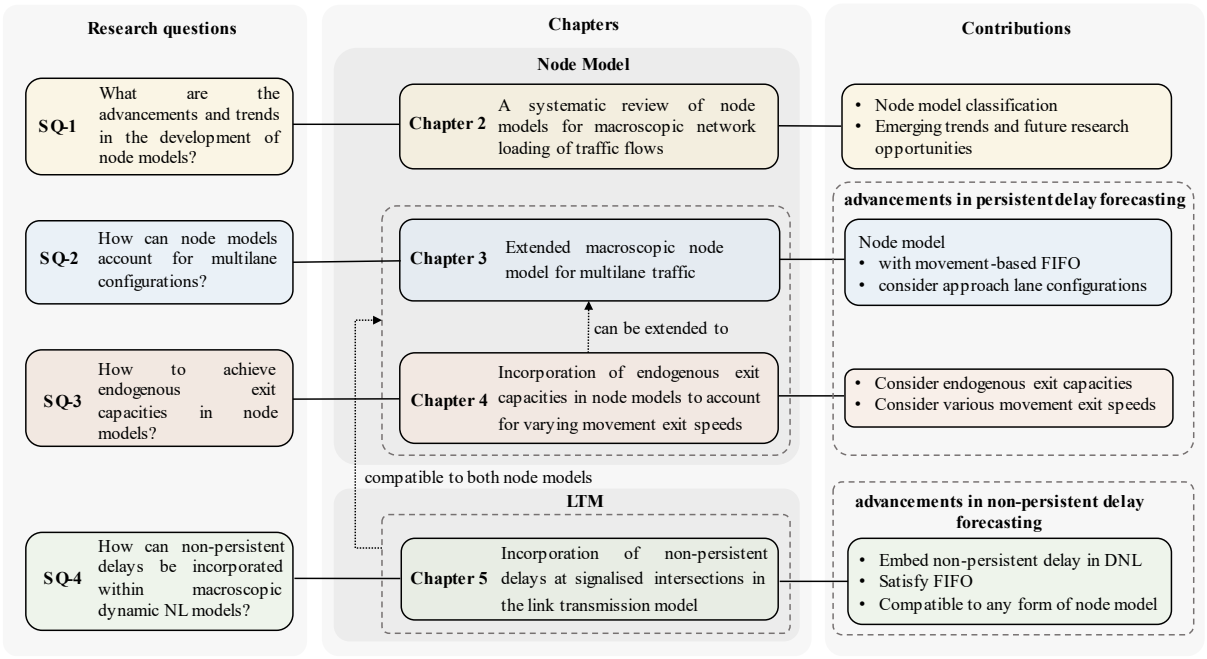


Fig. 7. Overview and structure of the paper-based chapters of this thesis.

Chapter 2: A systematic review of node models for macroscopic network loading of traffic flows. This chapter addresses SQ-1 through a systematic literature review on the development of node models within macroscopic NL frameworks. It identifies key characteristics of node models and systematically classifies and interprets these characteristics across existing studies. The chapter utilises six general principles to classify existing node models and proposes five additional extension categories to these principles, to discuss other relevant features of existing node models. The primary output of this chapter is a comprehensive classification table, which includes the formulation and detailed explanation of these principles and extensions, supported by relevant references. Additionally, this chapter identifies future research directions, offering insights for researchers and practitioners in the field of macroscopic node models.

Chapter 3: Extended macroscopic node model for multilane traffic. This chapter addresses SQ-2 by formulating a novel lane-based node model designed to relax link-based FIFO constraints when incoming links of nodes have multiple lanes. A lane equilibrium principle, for distributing sending flows across lanes, is proposed to take into account the approach lane configuration and ensuring movement-based FIFO principles. This method is formalised using a variational inequality problem formulation. The chapter also proposes a general solution

algorithm to solve the extended node model. In addition, numerical examples and a case study are provided, offering comparisons to the conventional link-based node model.

Chapter 4: Account for endogenous exit capacities in node models with varying movement exit speeds. This chapter addresses SQ-3 by proposing a method that accounts for endogenous exit capacities within node models. This method proposes a concept - saturation movement units - based on different movement exit speeds. Instead of dynamically adjusting exit capacities, it considers exit capacities by moderating the flow rates endogenously. This is achieved by scaling up inflows to the node model in pre-processing and subsequently scaling down outflows in the post-processing after running the node model. Additionally, this chapter provides a solution algorithm for the proposed method. Various numerical examples are provided to demonstrate the viability of this approach.

The models proposed in Chapters 3 and 4 are methodologically independent of each other. However, this independence does not preclude potential mutual adoption between the two models. The link-based node model with endogenous exit capacities discussed in Chapter 3 can be extended to lane-based node model by adopting the method proposed in Chapter 4. Conversely, the lane-based node model with exogenous exit capacities presented in Chapter 4 can be enhanced to achieve endogenous exit capacities by adopting the methodology outlined in Chapter 3.

Chapters 3 and 4 concentrate on enhancing persistent delay prediction within the node model. Non-persistent delay, however, is commonly overlooked in the existing dynamic NL models (at least in an endogenous context). Thus, Chapter 5 focuses on improving the forecasting of non-persistent delays within a well-known macroscopic dynamic NL model - LTM.

Chapter 5: Incorporation of non-persistent delays at signalised intersections in the link transmission model. This chapter addresses SQ-4 by introducing novel methodology to directly integrate non-persistent delays into the LTM formulation, effectively averaging the impact of within-cycle delays of traffic lights without violating the FIFO principle. A virtual link is implicitly incorporated into the link model representation, wherein Webster's uniform delay is reformulated as a hypocritical branch of its associated fundamental diagram form. Numerical examples are provided to illustrate the practical application of the proposed method. The adjusted LTM formulation presented in this chapter is applicable to any general node model, existing, or as introduced in Chapter 3 and 4.

2 Chapter 2 Literature Review

A (systematic) review of existing node models is absent in the field of macroscopic transport modelling. This chapter systematically reviews 48 node model papers selected through the PRISMA (2015) method. This chapter updates and evaluates six general principles for node models and proposes five extensions, highlighting additional features of existing models. It classifies node models based on the identified characteristics and provides a comprehensive synthesis of existing research, identifying key advancements and future research opportunities. This chapter provides valuable insights for researchers and practitioners in the field of macroscopic node models in traffic management and urban planning by offering insights into best practices and innovative approaches. This chapter also serves as the literature review of the entire thesis. It provides a research background and theoretical foundation for methodology research Chapters 3, 4, and 5.

This chapter includes the following paper:
Gong, X., Raadsen, M.P.H. & Bliemer, M.C.J. (2024). A systematic review of node models for macroscopic network loading of traffic flows. Submitted to *Transportation Research Part B: Methodological*.

Manuscript: A Systematic Review of Node Models for Macroscopic Network Loading of Traffic Flows

Abstract

Node models in network loading are used to distribute competing flows arising at motorway ramps, junctions, and intersections within macroscopic traffic models, influencing congestion and queuing delays. Despite decades of research on macroscopic node model development, a comprehensive literature review on their characteristics, categories, emerging trends, and further research opportunities does not yet exist. This study fills this gap by conducting a systematic literature review on the node models in macroscopic network loading frameworks. We identify representative characteristics of node models and then systematically classify and interpret those characteristics across existing node model studies present in the literature. We propose six general principles for node models and explore five extension categories characterising additional features. This paper makes two contributions to the field. Firstly, it provides a comprehensive classification of node model research, grounded in the proposed principles and extension categories. This classification is substantiated by relevant references and culminates in the development of a node model classification table. Secondly, it identifies future research directions and opportunities, providing guidance and insights for researchers and practitioners engaged in the study of macroscopic node models.

Keywords: Macroscopic network loading, node model, traffic simulation, traffic assignment

2.1 Introduction

2.1.1 Background

Traffic flow modelling is commonly undertaken within the framework of Traffic Assignment (TA), which is a process that allocates predicted traffic demand across the infrastructure network, thereby generating traffic flow and travel time forecasts. The framework of macroscopic TA is characterised by two fundamental components: the *route choice* model and the *network loading* (NL) model. These two models are interdependent. The route choice model allocates traffic flows among viable pathways according to generalised costs, while the NL model disseminates flows along the selected routes from the origin to the destination via links and nodes. Links denote uniform road segments characterised by attributes including length, speed limit, and number of lanes. Nodes represent junctions or intersections, facilitating specific turn movements and possibly regulated by traffic signals. Generalised travel costs encompass a broad metric reflecting the aggregate cost imposed on a traveller, encompassing for example travel time, tolls, reliability perception, comfort, safety, and environmental considerations. For transportation planning, travel time frequently serves as the primary cost metric to achieve equilibrium in the TA model. Traffic flow can be analysed at macroscopic, mesoscopic, and microscopic levels of aggregation. Macroscopic models aggregate individual vehicle dynamics to a flow-based representation to delineate overall traffic patterns, such as congestion and queue formation, making them ideal for large-scale network assessments where traffic characteristics like speed, density, and flow are pivotal. Microscopic models, on the other hand, delve into the specifics of individual vehicle behaviour, including vehicle following, lane changing, and gap acceptance, providing a detailed, albeit computationally intensive, view of traffic dynamics. Mesoscopic models strike a balance, treating traffic as packets or individual vehicles under macroscopic laws, thereby offering a more detailed analysis with reduced complexity compared to microscopic models. Given our focus on strategic and large-scale applications, this review's scope is chosen to be primarily focused on node models within a macroscopic NL context.

A macroscopic NL model consists of a link model and a node model. The link model describes how flows (and queues) propagate along links, resulting in sending flows on each downstream link boundary and receiving flows on each upstream link boundary. *Sending flows* represent downstream flows looking to exit the link, while *receiving flows* are the maximum allowable inflow rates available to enter a link upstream. The node model component then determines the accepted outflow of incoming links by taking sending flows and receiving flows from link models as inputs. This accepted outflow then serves as the inflow rate for the outgoing links of each node. In macroscopic NL models, time may be modelled either statically or dynamically. Static models assume that travel demand and network conditions remain constant over time without considering traffic flow dynamics. Traditional static (capacity restrained) models lack a node model and are consequently less relevant to this review paper. Node models in static NL were introduced in the context of capacity-constrained static NL models. Bifulco & Crisalli (1998) for example introduced an early method consistent with an implicit node model with infinite receiving flows. Their NL approach was generalised and replaced with an explicit first order node model in Bliemer et al. (2014). Conversely, dynamic models are distinguished by their ability to encapsulate the temporal evolution of traffic conditions, thereby accounting for fluctuations in traffic densities, queue formations, and their subsequent resolutions. The two most well-known dynamic NL models have been extensively explored in the literature, namely the *Cell Transmission Model* (CTM) (Daganzo, 1994, 1995) and the *Link Transmission Model*

(LTM) (Yperman, 2007; Gentile, 2010; Friesz et al., 2013; Han et al., 2016; Himpe et al., 2016; van der Gun et al., 2017; Raadsen & Bliemer, 2019; Bliemer & Raadsen, 2019).
Node models are integral to the majority of existing macroscopic NL frameworks, underpinning their ability to simulate traffic flows and congestion patterns effectively. Consequently, substantial research efforts have been dedicated to the development of macroscopic node models, aiming to enhance their accuracy and efficiency in reflecting real-world traffic dynamics. These efforts have led to advancements and improvements in the features of node models, edging closer, albeit still simplified, compared to their microscopic model counterparts. To the best of our knowledge, a comprehensive review of macroscopic node models remains absent from the literature. Therefore, in this paper, we undertake a systematic review of the literature on node models in the context of macroscopic network loading.

2.1.2 Research scope and objectives

In transportation research, and in this work, we make a distinction between a *node* (or junction), defined as a point where two or more roads converge, typically characterised by its lack of spatial dimension, and, an *intersection* model, that incorporates a spatial aspect such as length, width, and conflict points. This paper specifically focuses on node models and deliberately excludes intersection models as they exist in microscopic and mesoscopic contexts. However, it is noteworthy that some studies employ intersection models as proxies for node models devoid of spatial dimensions. To ensure comprehensiveness, our search strategy included terms related to both node and intersection models.

The systematic literature review aims to consolidate current knowledge on the development and application of node models within DNL frameworks. Specifically, it aims to: (a) identify the key characteristics of generic node models, (b) evaluate the methodologies and technologies used in these models, (c) assess the capability of various node models in replicating actual traffic flows, and (d) highlight emerging trends and potential areas for future research in the field. By doing so we hope to provide valuable insights to researchers, policymakers, and traffic engineers engaged in the planning and management of urban traffic networks. Furthermore, we discuss and identify various opportunities for future research directions.

2.1.3 General principles and extensions

Early contributions to node models have been made by Daganzo (1994, 1995) and Lebacque (1996), focusing on clarifying the complex interactions and establishing boundary conditions within traffic flow dynamics at the network. Since then, various node models have been proposed in subsequent literature. Although they may seem very different, they share the same principles. Following the incremental transfer (IT) principle introduced by Daganzo et al. (1997), Tampère et al. (2011) proposed the Generic Class of First-order Macroscopic Node Models (GCNM), setting seven generic requirements that all node models should satisfy to be considered consistent and effective, thereby defining the GCNM model criteria. The seven requirements of a GCNM are: *1. general applicability, 2. non-negativity of flows, 3. conservation of flow, 4. satisfaction of invariance principles, 5. satisfaction of demand and supply constraints, 6. flow maximisation, 7. obeying link-based First-In-First-Out (FIFO)*. A node model that satisfies all seven requirements is a generic first-order node model. Since then, attempts have been made to relax some of the original requirements, most notably, requirements regarding FIFO and flow maximisation. As a result, in this work, the original

seven requirements have been reorganised into six categories of principles, each with sub-categorisation where appropriate.

The six categories of general principles were chosen to be the following: (i) basic traffic flow principles, (ii) FIFO compliance, (iii) operational principles, (iv) intersection applicability, (v) priority, and (vi) intersection performance. The degree to which a macroscopic node mode satisfies these principles/aspects determines its capability and suitability for various application contexts. Assessing the capabilities across existing node models also reveals gaps and subsequent potential future research opportunities.

In addition to these six general principles, we classify recent advancements in node model research, by proposing five extension categories for node model classification, namely (i) signal control, (ii) conflicts, (iii) multiuser class, (iv) intersection layout, and (v) stochasticity. The complete set of principles, its attributes, and extensions are summarised in Table 1.

Table 1. General principles and extensions of macroscopic node models.

General principles	Basic traffic flow	• Non-negative flow • Flow conservation
	FIFO	Link / Movement / Lane / No
	Operational	• Invariance principle • Demand & supply constraints
	Intersection applicability	Connect / Merge / Diverge / General
	Priority	Exogenous parameters/ Demand proportional/ Capacity proportional
	Intersection performance	Flow maximisation / Holding free / Not holding free
Extensions	Account for	Signal control
		Conflicts
		Multiclass
		Intersection layout
		Stochasticity

Tampère et al. (2011) also introduced a concept they termed the *supply constraint interaction rule (SCIR)*. This is not so much a requirement, but the overarching concept of how vehicle behaviour translates to aggregate flows traversing the node, for example, assuming cooperative behaviour will yield different results to assuming selfish behaviour in a model, such decision influences SCIR, hence the inclusion of the term 'interaction'. In the context of this work we do consider SCIR explicitly, but instead attempt to be (largely) captured through the intersection performance principle classification, see also Section 2.3.6.
In the remainder of this paper, we discuss each of the principles and their attributes in more detail. We provide examples of the literature that support all or a subset of each principle/extension to facilitate this discussion. The full classification of all relevant literature identified as part of our systematic review is provided in Table 2. This is the main contribution of this work.

2.1.4 Outline

Section 2.2 details the methodology used for the literature search, and synthesis the findings related to node models in DNL. Section 2.3 presents the six general principles derived from the review. Section 2.4 discusses the five extensions to node models. Section 2.5 discusses the applicability of the node model with proposed principles and extensions and identifies potential avenues for future research. It also concludes the review, summarising key insights and implications for the field of traffic engineering and management.

Table 2. Node model classification overview (all studies are compliant with non-negativity and flow conservation).

Node model	*General Principles*							*Extensions*				
	Intersection applicability	*FIFO*	*Priority*	*Intersection performance*	*Operational*			*Intersection layout*	*Signal control*	*Conflicts*	*Multiclass*	*Stochasticity*
					Invariance Principle	*Demand Constr.*	*Supply Constr.*					
Newell (1993)	Merge	Link	On-ramp	Max. flow	■	■	■	□	–	–	–	–
Daganzo (1994)	Diverge	Link	-	Max. flow	■	■	■	–	–	–	–	–
Daganzo (1995)	Merge†	Link	Exogenous	Max. flow	■	■	■	–	–	–	–	–
Holden & Risebro (1995)	General	Link	-	Max. flow	■	■	■	–	–	–	–	–
Lebacque (1996)	General	Link or Mov.	Capacity	Not holding free	■	■	■	–	–	–	–	–
Daganzo (1997)	Diverge	Lane	-	Max. flow	■	■	■	□	–	–	–	–
Newell (1999)	Diverge	Lane	-	Max. flow	■	■	■	□	–	–	–	–
Herty & Klar (2003)	Merge, Diverge	Lane	-	Max. flow	■	■	■	□	–	–	–	–
Jin & Zhang (2003)	Merge	Link	Demand	Max. flow	–	■	■	–	–	–	–	–
Jin & Zhang (2004)	General	No	Demand	Not holding free	–	■	■	–	–	–	–	–
Ni & Leonard (2004)	Diverge	Mov.	-	Max. flow	■	■	■	–	–	–	–	–
Coclite et al. (2005)	General	Link	-	Max. flow	■	■	■	–	–	–	–	–
Ni & Leonard (2005)	Merge	Link	Capacity	Max. flow	■	■	■	–	–	–	–	–
Lebacque & Khoshyaran (2005)	Diverge, Merge	Link	Capacity	Max. flow	■	■	■	–	–	–	–	–
Ni et al. (2006)	Diverge, Merge†	Link	Capacity	Max. flow	■	■	■	–	–	–	–	–
Laval & Daganzo (2006)†††	Merge/Diverge	Lane	Capacity	Max. flow	■	■	■	□	–	□	–	–

Gentile et al. (2007)	General	Link	Capacity	Not holding free	■	■	■	–	–	–	–	–
Yperman (2007)	General	Link	Exogenous	Not holding free	■	■	■	–	□	–	–	–
Bliemer (2007)	General	Link	Demand	Holding free	–	■	■	–	–	–	■	–
Durlin & Henn (2008)	General	Link	Demand	Holding free	■	■	■	–	□	–	–	–
Nie et al. (2008)	General	Link	Demand	Holding free	–	■	■	–	–	–	–	–
Jin (2010)	Merge	Link	Exogenous	Max. flow	■	■	■	–	–	–	–	–
Tampère et al. (2011)	General	Link	Capacity	Holding free	■	■	■	–	□	–	–	–
Flötteröd & Rohde (2011)	General	Link	Exogenous	Holding free	■	■	■	–	–	–	–	–
Gibb (2011)	General	Link	Capacity	Holding free	■	■	■	–	–	–	–	–
Osorio et al. (2011)	General	Link	Capacity	-	–	■	■	–	–	–	–	■
Gao (2012)	General	Lane	Exogenous	Not holding free	■	■	■	□	–	–	–	–
Jin (2012)	General	Link	Demand	Holding free	■	■	■	–	–	–	–	–
Corthout et al. (2012)	General	Link	Capacity	Holding free	■	■	■	–	–	□	–	–
Jin & Zhang (2013)	Diverge	Link	-	Max. flow	–	■	■	–	–	–	–	–
Smith et al. (2013)	Merge†	Link	Exogenous	Max. flow	–	■	-	–	–	–	–	–
Han et al. (2014)	General (no merge competition)	Link	-	Not holding free	■	■	■	–	■	–	–	–
Bliemer et al. (2014)	General	Link	Capacity	Holding free	■	■	■	–	–	–	–	–
Shiomi et al. (2015)†††	Connect	Lane	Capacity	Max. flow	–	■	■	□	–	□	–	–
Jin (2015)	Diverge	Link	-	Max. flow	■	■	■	–	–	–	–	–
Han & Gayah (2015)††††	General (no merge competition)	Lane	-	Not holding free	■	■	■	□	■	–	–	–

Smits et al. (2015)	General	Link	Capacity	Holding free	■	■	■	–	–	–	–	–
Jabari (2016)	General	Link	Exogenous	Holding free	■	■	■	–	□	–	–	–
Himpe et al. (2016)	General	Link	Capacity	Holding free	■	■	■	–	–	–	–	–
Wright et al. (2016)	General	Link††	Exogenous	Holding free	■	■	■	□	–	–	■	–
Wright et al. (2017)	General	Link††	Exogenous	Holding free	■	■	■	□	–	–	■	–
Flötteröd & Osorio (2017)	Diverge, Merge	Link	Exogenous	-	–	■	■	–	–	–	–	■
Thonhofer et al. (2018)	General	Link	Demand	Holding free	–	■	■	–	□	–	–	–
Guo et al. (2021)	General	No	-	Holding free (HDV)/Max. flow (CAV)	–	■	■	–	–	–	□	–
Lu & Osorio (2022)	General	Link	Exogenous	-	–	■	■	–	–	–	–	■
Zhang et al. (2023)	Diverge, Merge	Link	Capacity	-	–	■	■	–	–	–	□	■
Yahyamozdarani & Tampère (2023)	General	Link	Capacity	Holding free	■	■	■	–	■	–	–	–
Gong et al. (2024)	General	Mov.	Capacity	Holding free	■	■	■	■	–	–	–	–

Demand Constr. = demand constraint; Supply Constr. = supply constraint; Max. flow = flow maximisation; Mov. = movement

†: merge node with max two incoming links which is not generally applicable, we listed properties of the equal delay merge that were explored qualitatively

††: partial link-based FIFO

†††: Special case applied within the link (connect node) on lane level, but from lane-level perspective, this is a general node, i.e., multi-lane in, multi-lane out node model. Implementations use the incremental transfer principle resulting in effectively a capacity-based priority when merging flows.

††††: Signalised for general intersections but effectively split to lane-to-lane with a fixed capacity based on green time, i.e., connect node model with pre-processing

■: developed; □: partially developed; –: undeveloped; -: not available

2.2 Method

For this study, we adopt the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, 2015). This technique requires the detailing of the search process, clearly indicating the sources, screening steps, and justification for choices made. This process is outlined below.

2.2.1 Search criteria

The *inclusion criteria* required studies to be: (a) peer-reviewed articles, book chapters, and conference papers, (b) research explicitly addressing the development or application of macroscopic node models within a network loading context, and (c) publications providing novel methodology including demonstrable results via examples or case studies. *Exclusions* were also identified, and studies were discarded when they were found to be: (a) review articles, (b) incomplete, i.e., texts not available in full, and (c) lacking discernible node model techniques/impact, underlying assumptions, mathematical formulations, and/or implementation details.

2.2.2 Search strategy

We comprehensively searched four academic databases: Science Direct, Web of Science, Scopus, and Google Scholar. The exact keyword specification used was ("macroscopic") and ("node model" OR "intersection model" OR "junction model") AND ("link transmission model" OR "network loading" OR "traffic simulation" OR "traffic assignment"). There were no restrictions placed on the year of publication to be able to capture the most relevant developments in the field over time. The search process performed between February to March 2024. All the identified and unique articles underwent title and abstract screening. All articles that passed the title and abstract screening were reviewed by studying their full texts. Articles whose full texts could not be retrieved were excluded, and so were duplicate studies. To ensure comprehensive literature coverage, we employed a forward and backward snowballing method by reviewing the reference lists of all included studies. This process was utilised to identify any potentially relevant studies that were not captured through the initial database searches, minimising the risk of omitting pertinent research.

Table 3. Searching string, inclusion, and exclusion criteria.

Source	Web of science (21), Science direct (144), Scopus (22), Google Scholar (499)
Search field	Title, abstract, keywords
Search terms	("macroscopic") AND ("node model" OR "intersection model" OR "junction model") AND ("link transmission model" OR "network loading" OR "traffic simulation" OR "traffic assignment")
Refined by	Document type: peer-reviewed articles, book chapters, and conference papers Timespan: no restriction

2.2.3 Search results

Fig. 1 presents a breakdown of the search and screening process. The searches in both databases resulted in 686 potentially eligible studies. These were imported to Covidence (2014) Covidence, and 290 duplicates were removed, leaving 396 unique records. The titles and abstracts of these studies were screened to check their relevance to the aims of this review based on the eligibility criteria. The screening of titles and abstracts resulted in 35 potentially eligible studies for a full-text review. After forward and backward snowballing, an extra 13 papers were found eligible to be included in the review set. These 48 studies constitute the final dataset utilised in this review, and none should be excluded. An additional 12 references are cited in this paper to provide contextual background and methodological justification.

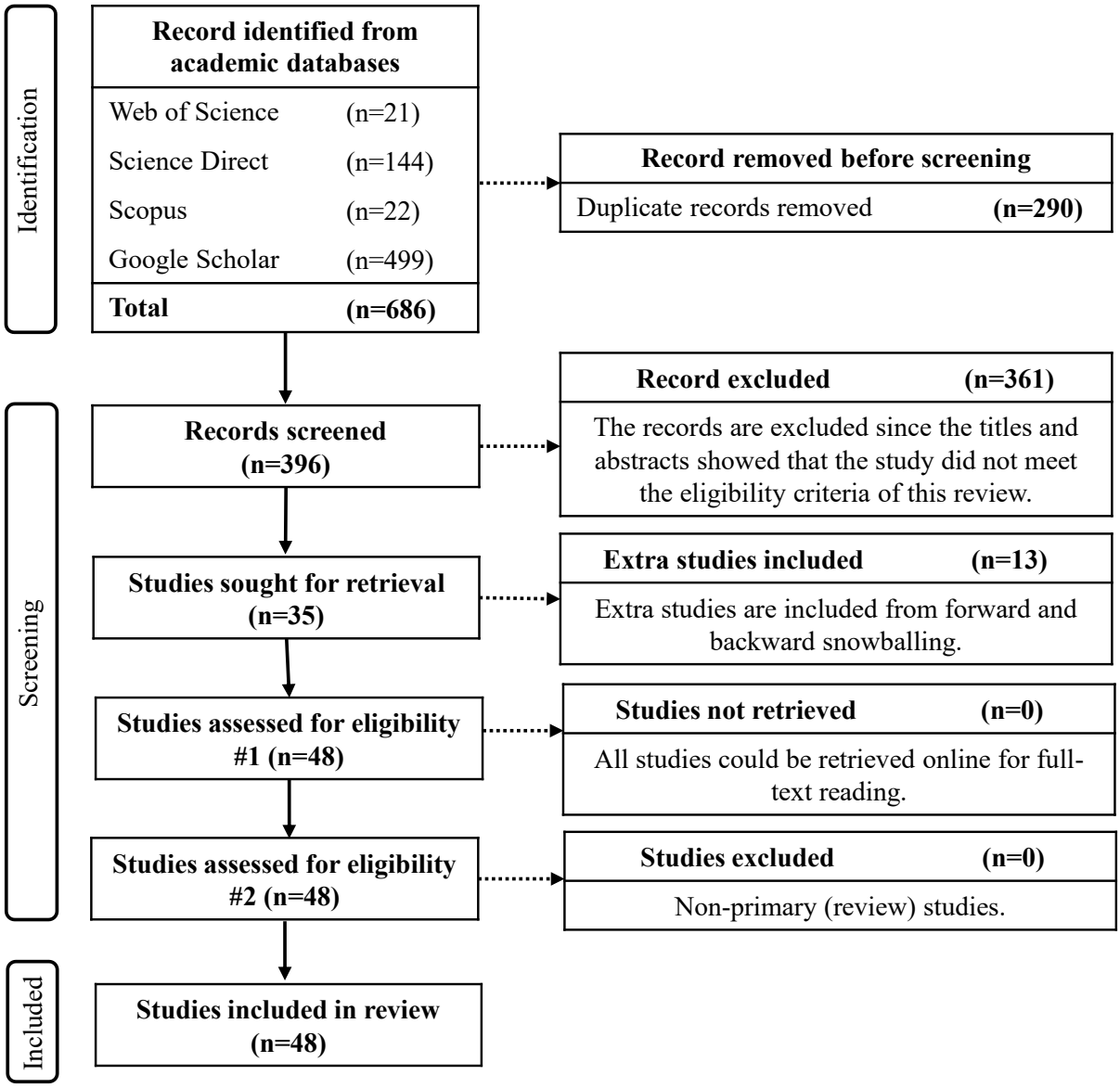


Fig. 1. PRISMA flow diagram of the article inclusion process (PRISMA, 2015).

2.3 General principles

We begin our discussion based on the six general principles outlined in Table 1. Each subsection introduces one of the principles in detail and discusses the literature relevant to it. Recall that the full classification of all studies and how they adhere to these principles are summarised in Table 2.

2.3.1 Basic traffic flow principles

Basic traffic flow principles require each macroscopic node model to (i) avoid negative traffic flows, typically resulting in the model imposing a non-negativity constraint in their methodology, and (ii) ensure that no vehicles are lost, i.e., adhere to flow conservation.
Non-negativity is a requirement because we assume traffic does not move backward on a road (when modelled) in which case the lowest flow feasible is zero flow by definition. *Flow conservation* mandates that at any node, all flow entering the node – inflow - must necessarily be equal to the sum of all flow exiting the node – outflow – where we assume that the node is not an origin or destination of any traffic. Violating flow conservation is not allowed because it would lead to generating ghost traffic or incorrectly removing traffic from the network. All existing node models in this study are compliant with these two requirements. We have therefore chosen to not explicitly list them in Table 2 for brevity.

2.3.2 FIFO principles

The First-in-first-out (FIFO) principle, indicates that vehicles must exit in the same order they entered. This principle can be applied on multiple granularity levels, such as links, lanes, or movements. Especially in a macroscopic context, it is an important concept since many macroscopic network dynamic loading models rely on this principle to be able to construct a solution to their traffic propagation. It is therefore typically considered a desirable property for the node models used in such a context.

In the context of the GCNM (Tampère et al., 2011), link-based FIFO is introduced as part of what the authors refer to as 'the conservation of turning fractions', stating that vehicles on a link exit in the entered order. Imposing link-based FIFO as a requirement has important consequences, for example, it implies that whenever flow on a link becomes supply-constrained, i.e., there is too little capacity available on the exit link to accommodate all the demand, and a queue starts to form on the incoming link, all traffic flows wanting to exit this link are equally impacted (Daganzo, 1995), including the flows that desire to exit through another non-saturated link. Consequently, link-based FIFO guarantees that only a single queue forms on each node's incoming link, irrespective of approach lane configuration. This approach is widely adopted in existing node models (Daganzo, 1994, 1995; Lebacque & Khoshyaran, 2005; Bliemer, 2007; Jin, 2010; Tampère et al., 2011; Flötteröd & Rohde, 2011; Han et al., 2014; Smits et al., 2015; Jin, 2015; Jabari, 2016; Thonhofer et al., 2018; Guo et al., 2021; Lu & Osorio, 2022; Yahyamozdarani & Tampère, 2023; Zhang et al., 2023). Nonetheless, such models may restrict traffic flow too stringently, compared to reality. This limitation is known but has been challenging to resolve. Laval & Daganzo (2006) and Shiomi et al. (2015) mimic a lane-based FIFO by modelling each lane as if it were a link and allowing for lane changes per fixed length "cell" akin to how one would model a diverge node, albeit without explicitly labelling this as a node model and this behaviour occurring within a link. This adaptation can

be considered a "workaround" since the node model remains link FIFO, only each link is now guaranteed to always be a single lane. Similarly, Gao (2012) and Han & Gayah (2015) propose lane-based node models that explicitly consider the influence of directional turning lanes and physical queues in different lanes. Their models explicitly account for directional queues in each turning lane (by modelling them as links) and implement a lane-based FIFO. Wright et al. (2016, 2017) noted that such approaches significantly increase model complexity, run times, and complicate achieving route choice equilibrium in general networks exponentially because of the numerous multilane intersections. Instead, Wright et al. (2016, 2017) introduced a partial link-based FIFO mechanism where some queues only affect specific movements but not others. Their approach however is not endogenous, meaning that the interaction and partial blocking need to be pre-specified and does not consider the endogenous nature of interactions that occur between flows when approach lanes become congested. Recently, Gong et al. (2024) proposed a model that integrates the concept of a lane choice equilibrium within an implicitly expanded node representation, relaxing link-based FIFO to movement-based FIFO. In this approach, FIFO holds across all used lanes for the same movement and the distribution of flows and interactions due to congestion are endogenously dealt with within the node model itself.

While link-based FIFO has the benefit of yielding a single queue per link, it is the least accurate reflection of reality (Jin & Zhang, 2004; Ni & Leonard, 2004). Conversely, lane-based FIFO provides the most accurate representation of flow dynamics, but it increases computational complexity and complicates the route choice problem at the network level. Movement-based FIFO is an alternative to both. It is less restrictive than link-based FIFO and can avoid the route choice complexities on the network level that come with lane-based FIFO. Yet they do yield multiple queues per link which poses new challenges in network loading.

2.3.3 Operational principles

Operational principles describe the functioning and efficiency of the node model. The operational principle consists of (i) demand and supply constraints, and (ii) the invariance principle. When demand constraints are satisfied, then the flow traversing a node via an incoming link is guaranteed to not exceed the demand that wants to exit this link at its downstream boundary. Conversely, satisfying the supply constraint means that flow entering an outgoing link of a node does not exceed the available supply – capacity in unsaturated conditions - at the link's upstream boundary. The node model can only operate appropriately if both operational principles are satisfied. Given the intuitiveness of this principle, most models adhere to both demand and supply constraints. Examples of exceptions to this rule are Bifulco & Crisalli (1998) and Smith et al. (2013), they do not satisfy the supply constraint because they assume there is no limit on how much flow may enter a link despite enforcing a capacity on link exit.

The *Invariance principle* was first proposed by Lebacque & Khoshyaran, (2005). Like demand and supply constraints it considers two distinct situations to which it applies. Firstly, it states that – in congested conditions - the accepted outflow of a node's incoming link does not change when increasing its demand. This is reasonable because the number of vehicles that one allows to exit the link at the front of the queue is not governed by how many vehicles are entering the queue at the back. Similarly, the invariance principle states that for an uncongested situation on an exit link of the node, increasing the capacity of that link should not affect the flow into that link.

Demand-constrained node models – that do not adhere to the invariance principle – are more common than one might think and were used in practice as well. In those models, the desired demand of an incoming link determined its bargaining power to claim exit link capacity. While flawed, these methods did pioneer useful other features such as more general applicability as we discuss in Section 2.3.4. Examples of demand-constrained node models from this period can be found in Jin & Zhang (2003, 2004), Bliemer (2007), Yperman (2007), Nie et al. (2008), Durlin & Henn (2008), and Thonhofer et al. (2018).

Earlier diverge-only node models as well as the majority of (general) node models used in practice today all comply with the invariance principle (Daganzo, 1994, 1995; Lebacque, 1996; Ni & Leonard, 2004, 2005; Flötteröd & Rohde, 2011; Gibb, 2011; Tampère et al., 2011; Corthout et al., 2012; Bliemer et al., 2014; Smits et al., 2015; Himpe et al., 2016; Wright et al., 2017; Gong et al., 2024).

Some specific approaches based on finite capacity theory also do not adhere to the invariance principle (Osorio et al., 2011; Flötteröd & Osorio, 2017; Zhang et al., 2023). The main drawback of these approaches in practice is that they cause instability in the solutions of the node model causing flip-flopping regarding its accepted flow rates at the link boundary at increasingly smaller time intervals. When applying node models in a time-discretised network loading context, however, this flip-flopping dies out once its time interval to switch state drops below the adopted network loading time step, as demonstrated by Lebacque & Khoshyaran (2005). This mitigating feature however should not be considered a reason to adopt such an approach today as better and more robust approaches now exist.

2.3.4 Intersection applicability principles

The applicability principle comprises how universally usable a node model is regarding the intersection layouts that exist. The simplest layout being a (i) *connect node* with a single incoming and single outgoing link, see Fig. 2 (a). The first node models typically only supported (ii) *diverges,* nodes with a single incoming and multiple outgoing links, and (iii) *merges,* where two (or more) incoming links converge to form a single outgoing link. The most general of node models is a (iv) *cross* node, supporting both multiple incoming links and multiple outgoing links, referred to as a general node throughout this paper. Fig. 2 provides the simplest possible example of each of these types of nodes.

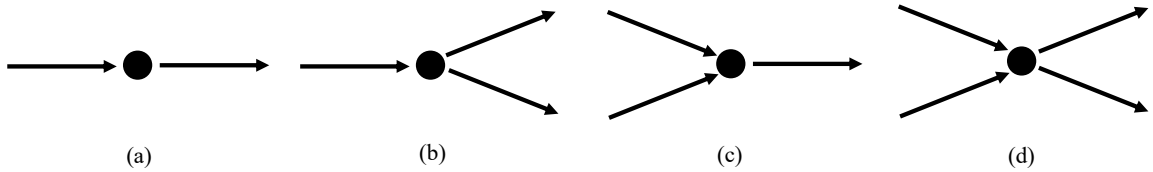


Fig. 2. The simplified (a) connect, (b) diverge, (c) merge, (d) general node model, with directional links indicated as arrows.

Daganzo's (1994, 1995) simplified first-order node model represents one of the earliest efforts in node model research. Specifically designed for motorways, Daganzo's model is applicable solely to merge and diverge nodes. Notably, their merge node is restricted to two incoming links to allow for a closed-form solution. When merge nodes have more than two incoming

links the close-form solution approach is no longer possible and finding a solution typically requires some kind of iterative algorithm. Diverge node models have been explored in several studies (Daganzo, 1994; Ni & Leonard, 2004; Lebacque & Khoshyaran, 2005; Gentile et al., 2007; Jin & Zhang, 2013) and are relatively straightforward when considering link-based FIFO conditions.

Enhancements to Daganzo's early merge node formulation include Jin & Zhang's (2003) introduction of a fairness requirement in merging priorities. Other research (Lebacque & Khoshyaran, 2005; Ni & Leonard, 2005; Gentile et al., 2007; Jin, 2010) have proposed various merging schemes, though these may yield unrealistic outcomes under specific scenarios. More recent research of Smith et al. (2013) investigated how merge nodes are affected by downstream bottlenecks, utilising spatial queuing models to depict traffic behaviours with and without spillback. Flötteröd & Osorio (2017) developed a diverge and merge node model in the context of a stochastic modelling framework (see also Section 2.4.5), while Lu & Osorio (2022) built on this to propose a general node model compatible with various stochastic link models.

The earliest general node models were developed by Holden & Risebro (1995) and Lebacque (1996), supporting any number of incoming and outgoing links. Further developments in this area were made by Jin & Zhang (2004), Bliemer (2007), Gentile et al. (2007), and Yperman (2007), expanding applicability to various traffic scenarios and/or network loading approaches. Building on these foundations, the GCNM by Tampère et al. (2011) set a foundational framework for macroscopic node models, influencing most current research in this area, demonstrating its broad applicability and influence in the field since (Flötteröd & Rohde, 2011; Tampère et al., 2011; Gibb, 2011; Smits et al., 2015; Himpe et al., 2016; Jabari, 2016; Wright et al., 2016, 2017; Guo et al., 2021; Yahyamozdarani & Tampère, 2023; Gong et al., 2024).

2.3.5 Priority principles

Priority principles refer to the strategies and approaches taken to determine the bargaining power for claiming the node's outgoing link capacity at the expense of others (incoming links, lanes, movements). Since it has not always been clear cut what approach is deemed most realistic, various approaches have emerged over the years. From the literature, three main methods to determine priority emerged as a result: exogenous-based parameters, demand proportional, or capacity proportional, where the latter two would both be classified as endogenous-based approaches. Note that priority only matters for nodes with multiple incoming links such as merge or general nodes.

Daganzo (1994, 1995) pioneered the use of an exogenous priority parameter to set the priority of each of its two competing incoming links for his simplified merge node model. This priority could be set to any value, although a common approach would be to set it based on the incoming link's capacity normalised to a value between zero and 1. This approach has been adopted in numerous other studies since (Flötteröd & Rohde, 2011; Wright et al., 2016, 2017; Flötteröd & Osorio, 2017).

Lebacque (1996) proposed a density proportional approach in a merge node context. Subsequently, Ni & Leonard (2005) proposed a capacity proportional priority merging rule, where they modified Daganzo's merge model by substituting the exogenous priority with the endogenous priority that is determined by incoming link capacities. This concept was further

generalised to general intersections by Gentile et al. (2007) and is currently the defacto standard for priority modelling (Lebacque & Khoshyaran, 2005; Ni et al., 2006; Gibb, 2011; Osorio et al., 2011; Tampère et al., 2011; Corthout et al., 2012; Bliemer et al., 2014; Shiomi et al., 2015; Smits et al., 2015; Himpe et al., 2016; Yahyamozdarani & Tampère, 2023; Zhang et al., 2023; Gong et al., 2024). Smits et al. (2015) converted capacity to a headway representation, but in doing so kept the same underlying principle for priority intact.

Recalling the demand-constrained models discussed in Section 2.3.3, Jin & Zhang (2003, 2004) integrated a fairness-driven demand-proportional distribution into Daganzo's merge model, where departing capacities are divided according to the demand on incoming links. Bliemer (2007) adapted this rule for general intersections, with several studies following this methodology (Durlin & Henn, 2008; Nie et al., 2008; Thonhofer et al., 2018). However, as mentioned demand-proportional distributions typically violate the invariance principle, leading to undesirable and unrealistic flow fluctuations.

2.3.6 Intersection performance principles

In node models, sending and receiving flows are defined as the maximum possible flows that incoming and outgoing links can offer, or accept, respectively. The utilisation of the receiving flows dictates the intersection performance qualification. The use of the term flow maximisation in the context of node models has been used loosely and interchangeably with the notion of holding free. To avoid any confusion, we adopt a very strict definition in this context. Firstly, we adopt the definition of a holding free solution, see Jabari (2016). They define holding free (H-F) such that no traffic is withheld unnecessarily when there remains available supply on their desired outgoing link. 'Unnecessarily' in this context, refers to the situation that traffic may still be withheld – even if there is an available outgoing supply, if it is blocked by traffic moving to another outgoing link on the same incoming link to, for example, satisfy FIFO. A holding free solution is also flow maximising when it concerns a diverge or merge node, but not necessarily for general nodes.

Most node models in the literature are in fact holding free studies (Daganzo, 1995; Lebacque, 1996; Lebacque & Khoshyaran, 2005; Flötteröd & Rohde, 2011; Tampère et al., 2011; Corthout et al., 2012; Himpe et al., 2016; Gong et al., 2024), with a few exceptions, most notably Lebacque (1996) and Jin & Zhang (2004). For example, approaches that allocate a fixed portion of the supply to each incoming link without any redistribution facilities cannot be considered holding free.

In this work we opt to define a narrow definition of what a flow maximising node model is, to avoid any confusion. All general node models that have a fixed rule to distribute the available supply across competing flows for an outgoing link and enforce FIFO, are not considered flow maximising. This is because the auxiliary flows towards other outgoing links in such situations get partially held back and they then determine whether the flow is maximised or not *across the entire node*. Hence, only when the model accounts for these secondary effects a node model is considered flow maximising. Note that this is why for merge and diverge nodes the distinction between H-F and flow maximising cannot occur. Practically, this means that cooperative behaviour is required to achieve flow maximisation. Therefore, most of the node models, despite claiming to be flow maximising, are in fact not under this definition, and are listed as holding-free instead. Flow maximisation is, therefore, a sufficient condition for a holding free solution, but not a necessary condition (Jabari, 2016).

Lastly, we point out the definition of flow maximisation is rather strict here, while looser in other places in the literature. This is hot helpful, and we may benefit from classifying this principle/requirement differently in the future. For example, when assuming selfish behaviour a model could still be considered flow maximising within this context, yet it will most likely have less flow throughput compared to a system optimum approach.

2.4 Extensions

Beyond the general principles discussed in Section 2.3, five categories of extensions have been identified within the field of macroscopic node model research. Each category is discussed to highlight key contributions to date. These five categories comprise signal control, conflict management, multiclass support, intersection layout consideration, and stochasticity modelling. Each of the extension categories is also incorporated in the classification table, recall Table 2.

2.4.1 Account for signal control

While most node model research has traditionally focused on unsignalised intersections, the inclusion of signal control can improve the accuracy of simulating flows and delays at signalised intersections. Especially in urban contexts such considerations are important to better capture reality, and we therefore see more research in this area in recent years. Node models that embed traffic signals in their approach, typically attempt to model some form of average impact caused by the reduction in capacity due to phasing and signal cycles on the accepted flows (Yperman, 2007; Durlin & Henn, 2008; Tampère et al., 2011; Han et al., 2014; Han & Gayah, 2015; Jabari, 2016; Thonhofer et al., 2018; Yahyamozdarani & Tampère, 2023). These models often treat traffic signals as constraints that affect the capacity of incoming links. For instance, Han et al. (2014) developed a continuum model to approximate variable signal phases, and subsequent improvements were made by Han & Gayah (2015) to address some of its limitations. However, many continuous signal models, which rely merely on the proportion of green time per link (e.g., Tampère et al., 2011; Han et al., 2014; Han & Gayah, 2015; Thonhofer et al., 2018; Zhang et al., 2023), fail to account for the distinct active phases of different links and their interactions, which led to Yahyamozdarani and Tampère (2023) to introduce a refined continuous signalised node model that calculates average turn flows over a cycle, while considering such interactions based on signal timings, enhancing the model's accuracy and relevance.

2.4.2 Account for conflicts

Most intersection models within macroscopic NL frameworks currently ignore the impact of situations where there are simultaneous cross-movement conflicts, e.g., where streams of traffic simultaneously (get green and) traverse a node and cross paths. In such a situation there exists what Tampère et al. (2011) refers to as *internal node supply constraints*. When considered these constraints reduce node flows accordingly. In one of the few studies to explore conflicts explicitly, Corthout et al. (2012) introduced additional supply constraints and explored how solution non-uniqueness could arise when different priorities are assigned to turns from the same incoming link under varying constraints. This phenomenon highlighted a fundamental aspect of modelling traffic behaviour, where increased realism through detailed constraints

may lead to undesirable mathematical properties such as non-unique flow solutions. We come back to this discussion in Section 2.5.

2.4.3 Account for multiclass

Urban roads feature a variety of vehicle types and drivers, where each combination of those yields different aggregate flow patterns. To be able to capture these types of intricacies node models ideally support vehicle and/or user classes endogenously. Doing so may lead to additional complexities to consider, for example considering bicycles may require a different type of FIFO relaxation where there is FIFO within a user class but not across classes.
Current research on node models predominantly addresses homogeneous traffic; however, studies on mixed traffic are emerging. These studies typically focus on two areas: one dealing with multiple vehicle types (Bliemer, 2007; Wright et al., 2016, 2017) and the other with autonomous and human-driven vehicles (Guo et al., 2021; Zhang et al., 2023). Bliemer (2007) integrated queuing and spillback in a multiclass setting within a macroscopic dynamic NL framework. Wright et al. (2016, 2017) introduced an additional requirement to the existing seven requirements defined by Tampère et al. (2011), tailored for multiclass/multi-commodity traffic flows. This requirement states that "supply restrictions on a flow from any given input link are imposed on commodity components of this flow proportionally to their presence in this link". Guo et al. (2021) and Zhang et al. (2023) developed a model that specifically addresses the dynamics of mixed traffic, incorporating stochastic elements and dual-queue mechanisms to capture the interactions between different types of vehicles. Additionally, Zhang et al. (2023) developed a stochastic dynamic NL model for mixed traffic scenarios involving autonomous and human-driven vehicles.

2.4.4 Account for intersection layout

In urban settings, approach links often have multiple lanes supporting different movements. This means that traffic must consciously choose their lane to make their desired turn. Modelling this behaviour and the subsequent interactions that occur therefore becomes increasingly important to capture reality. Doing so not only enhances the model's accuracy on the node level but may impact the ability to correctly capture spillback effects such as route choice behaviour on the network level. Most existing node models implicitly assume all lanes support all movements, making them suitable for situations where an incoming link has a single lane. In those situations, this is a natural assumption with the added benefit that link-FIFO, lane-FIFO, and movement FIFO work out to be the same. However, with multilane incoming links, the situation becomes more complicated, and the all-lane-all-movement assumption may be problematic. Firstly, because it is no longer realistic in comparison to a viable intersection layout. Secondly, because the common assumption of link-based FIFO likely is too restrictive. For instance, in a two-lane incoming link designated for different turning directions, FIFO may not hold practically as one lane may experience queuing while the other proceeds unimpeded. Gao (2012), Han & Gayah (2015), Shiomi et al. (2015) explicitly model approach lanes to the node by considering lane-based FIFO. However, this approach necessitates the creation of separate approach lanes within the network, leading to an exponential increase in the number of routes. Consequently, solving route choice problems becomes significantly more challenging. Wright et al. (2016, 2017) attempt to address this issue by introducing a partial FIFO mechanism. In their model, approach lane configurations are simplified using a single exogenous parameter, bypassing a detailed analysis of infrastructure and driver behaviour impacts. This method, while not computationally intensive, relies on a subjective parameter

that may lack a physical underpinning. Such a fixed, exogenous, approach potentially leads to miscalibration. In response, Gong et al. (2024) recently developed a node model that explicitly considers the approach lane configuration, thus allowing for different queues per movement, without the need to model each lane separately. This approach also relaxes the link-based FIFO to a movement-based FIFO methodology that leverages an endogenous lane choice equilibrium approach, offering an arguably more realistic representation of multilane traffic dynamics at intersections than previously available.

2.4.5 Account for stochasticity

The last extension category we discuss investigates the various ways arrival flow rates may be modelled. Within macroscopic node models, we distinguish between two principal types: deterministic and probabilistic. The *Deterministic* class of models implicitly assume a uniform arrival pattern of flow. Due to this assumption, there is no need for a random element, hence their name. This type of model is the most common in the context of macroscopic node models Notable examples include the formulations by Daganzo (1995), Lebacque (1996), Lebacque & Khoshyaran (2005), Tampère et al. (2011), Flötteröd & Rohde (2011), Corthout et al. (2012), Smits et al. (2015), Jabari (2016), Yahyamozdarani & Tampère (2023), Gong et al. (2024) . Stochastic models attempt to account for non-uniform arrival patterns that we observe in reality to a more or lesser extent. The idea is that there are always fluctuations in arrival rates whether recurrent or non-recurrent and these fluctuations may have a significant impact on the resultant flows from the node models and should therefore be considered. To enhance and/or investigate network reliability and robustness, probabilistic node models employing stochastic arrival patterns have therefore been developed. They explicitly account for the unpredictable nature of traffic and driver behaviour by embedding assumed, for example, distributions in the modelling of traffic flow arrival rates. Notable examples can be found in Osorio et al. (2011) their differentiable stochastic dynamic network loading model, Flötteröd & Osorio (2017) their analytical stochastic LTM for various node types, or Lu & Osorio's (2022) general stochastic node model for integrated network analysis. Zhang et al. (2023) further examined stochastic arrivals' effects in mixed traffic, developing a node model that probabilistically describes congestion propagation.

2.5 Discussions and research opportunities

This paper classifies the current state of node models in macroscopic NL frameworks. This was done based on six general principles and a further with five extension categories. Findings were discussed and results were summarised in a classification table (Table 2).

Much progress has been made in the past decades. Still, no models provide an approach that classifies as the most advanced possible considering all the aspects that were explored (across both principle and extension categories). To guide our discussion on research opportunities we first provide another view on the findings in this work in the form of a Sankey diagram, see Fig. 3. As can be seen in Fig. 3, recent studies often provide their main contributions by investigating aspects of one or more of the extension categories discussed, especially integrations of signal control and the consideration of intersection layout impact were found to be fruitful avenues of research. Both add complexity but come with improved modelling capabilities. The incorporation of stochasticity within node models has been another emerging theme, providing insights into the variability and unpredictability of traffic flows despite the macroscopic context in which these models operate.

This brings us to the many research opportunities that remain. While acknowledging this is subjective, we feel, the following subjects provide the most promising opportunities for further exploration.

1) At an intersection, two primary types of flow interactions create conflicts: vehicle to vehicle and vehicle to other road users, such as pedestrians. Both types of conflicts can lead to traffic flow reduction at intersections. Despite their significance, this aspect is largely overlooked in current node models. Addressing conflicts accurately, efficiently, and robustly is important for improving the accuracy and utility of node models. However, to date, this issue remains inadequately addressed. The main reason being that incorporating conflicts can lead to non-uniqueness, which may be one of the reasons this topic has been largely avoided so far. Therefore, future research endeavours could focus on addressing conflicts within node models while ensuring the attainment of a unique solution.

2) Most existing node models focus on a single extension category (if they incorporate any at all). A more holistic approach, which integrates multiple extension aspects simultaneously, would be highly beneficial. For instance, combining stochasticity with conflict scenarios, or integrating multiple vehicle classes with signalised intersections and intersection layout configurations, could provide a more comprehensive understanding and robust modelling framework.

3) The aspect of flow maximisation is not adequately defined in the current literature. While some adopt a strict optimisation-based definition, most node models do not adhere to this yet still consider themselves to be flow-maximising (even though they may not be). It may be more helpful to define these models in terms of their underlying assumed behaviours, specifically how they achieve flow maximisation under these conditions. This is especially prudent given the anticipated shift from selfish behaviour to cooperative behaviour with the emergence of connected and autonomous vehicles.

4) Some extensions, such as conflicts and priority, which represent elements of micro-level modelling, may improve the accuracy but also increase the complexity of macroscopic modelling. Therefore, another more philosophical aspect of this type of research is the appropriate juncture at which to transition from macroscopic to microscopic modelling. Macroscopic models offer broad overviews and are invaluable for large-scale traffic forecasting and planning. However, if the application requires intricate behavioural insights and interactions at smaller scales, there is likely a point at which one must wonder if it is sensible to attempt to integrate this into a macroscopic model. Instead, crossing over to microscopic modelling may make more sense. Identifying the threshold where the benefits of micro-level detail outweigh the convenience and scope of macroscopic modelling is an interesting and largely unexplored discussion in the literature. This discussion is essential for guiding applications in both theoretical and practical contexts to achieve optimal results in real-world scenarios.

We hope that the findings of this work will be beneficial to researchers and practitioners in developing more comprehensive and integrated macroscopic node models used for both policy-making and urban transport planning, contributing to more effective and informed decision-making processes.

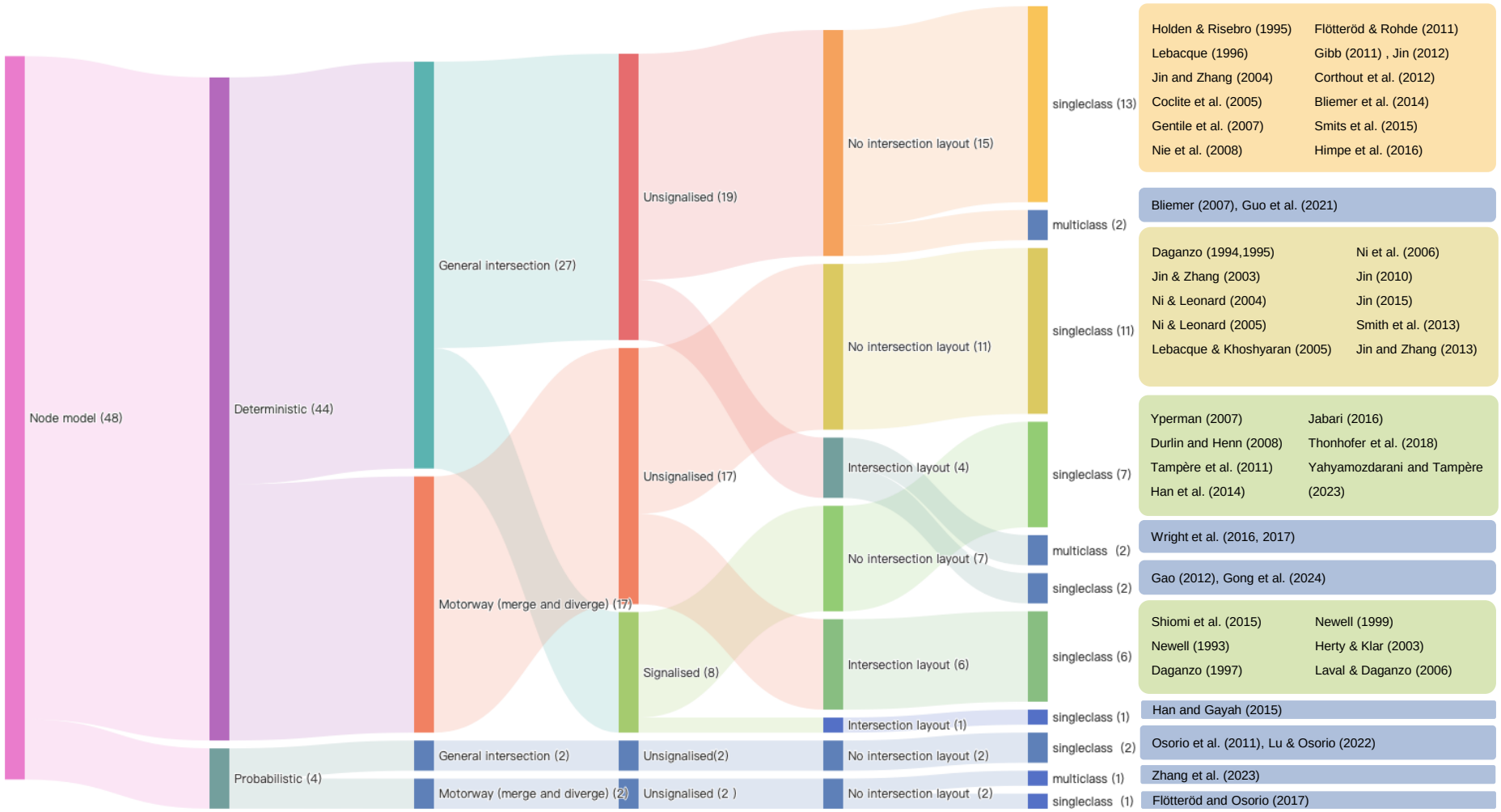


Fig. 3. Classification of node model research

3 Chapter 3 Node Model for Multilane Traffic

Existing node models typically follow link-based FIFO and ignore approach lane configurations. However, this assumption is too stringent for multilane intersections. This chapter introduces a novel (implicit) lane-based node model that explicitly considers approach lane configurations. This is achieved by first constructing a lane-expanded node model that disaggregates each incoming link into individual lanes, implicitly treating each lane as a separate link. A lane choice equilibrium principle is then introduced to aggregate lanes into movements, thereby relaxing link-based FIFO to movement-based FIFO. The proposed lane-based node model satisfies the seven general principles proposed by Tampère et al. (2011), making it a generic node model. It is noteworthy that the proposed model does not create explicit approach lanes within the link, thus avoiding the exponential growth of routes within the network. This chapter introduces a solution algorithm designed to identify lane choice equilibrium flows within the node model. Numerical examples are provided to validate the model's capability by comparing it to the conventional link-based node model.

This chapter includes the following paper:
Gong, X., Bliemer, M.C.J. & Raadsen, M.P.H. (2024) Extended macroscopic node model for multilane traffic. Submitted to *Transportation Research Part B: Methodological.*

Manuscript: Extended Macroscopic Node Model for Multilane Traffic

Abstract

In a macroscopic assignment model, traffic flows are distributed onto the network by means of a network loading model. The network loading propagates flows along links via a link model and through junctions or intersections via a node model. Most of the travel time delays are caused by queues forming at junctions or intersections, especially in urban networks. Therefore, the efficiency and accuracy of the underlying node model is paramount in capturing these delays (and flows). Existing link-based macroscopic node models make the simplifying assumption that first-in-first-out (FIFO) holds at the link level, which is often unrealistic when a link has multiple approach lanes near an intersection or junction. In this work we propose to relax this assumption such that FIFO holds at the movement level. We do so by developing several model extensions. First, a novel lane-based formulation of the node model is proposed. Secondly, we formulate an equilibrium problem and a general solution algorithm to allocate sending flows to lanes. This allows us to explicitly consider approach lane configurations that contain important information about the layout of an intersection or junction. We show that the conventional link-based node model is a special case of our newly proposed model in case each approach lane on an incoming link allows all possible movements. Various numerical examples are provided, demonstrating the capabilities of the proposed extensions to the node model.

Keywords: Macroscopic traffic assignment, link-based FIFO, multi-queue node model, lane choice problem

3.1 Introduction

3.1.1 Background

Traffic assignment (TA) plays a critical role in transport planning and traffic management. It distributes (predicted) travel demands to the infrastructure network, producing traffic flows and travel time estimates. In this study we focus on macroscopic TA models. Macroscopic TA models differ from microscopic TA models because they model traffic as aggregate flows (veh/h) rather than individual vehicles. Macroscopic models are most commonly adopted in a strategic planning context, where there is a long planning horizon as well as a large(r) geographic area to cover. TA models are typically composed of two interrelated models, namely a *route choice* model – possibly extended with mode and/or departure time choice, and a *network loading* (NL) model. The route choice model allocates traffic flows across feasible routes based on (generalised) route costs, while the NL model propagates the resultant path flows along the chosen routes obtained from the route choice model. These two models are interdependent. To obtain a solution, these models attempt to find an *equilibrium*. The notion of a user equilibrium (UE) is a commonly used equilibrium construct in strategic TA applications. According to Wardrop's first principle, in UE travellers selfishly keep updating their route until they can no longer reduce their travel time, which results in a situation where all *used* routes have the same minimum travel time (Wardrop, 1952).

Road infrastructure networks in macroscopic TA models are represented by *links* and *nodes*. Links describe homogeneous road segments defined by spatial characteristics such as length, speed, and number of lanes, while nodes represent locations where links come together and reflect motorway merges, diverges, or urban intersections. Nodes have permissible movements (such as left, straight, and right) for each incoming link and potentially have associated traffic signal controls. The link model describes how flows (and queues) propagate along links. This results in sending flows on each downstream link boundary and receiving flows on each upstream link boundary. *Sending flow* is the amount of flow that would like to exit a link, i.e., the node's incoming link's demand. Receiving flow is the maximum amount of flow that may enter a link, i.e., the node's outgoing link's supply. The node model governs the interaction between sending flows and receiving flows, resulting in *realised* flows through the node for all permissible movements. Realised flows may be lower than the sending flows if demand exceeds supply. Since the node model dictates the applied flow restrictions at the link boundaries, it plays a pivotal role in the modelled congestion and queuing delays.

Node models are applied in dynamic NL model such as the cell transmission model (Daganzo, 1994, 1995) and link transmission model (Yperman, 2007; Gentile, 2010; Friesz et al., 2013; Han et al., 2016; Himpe et al., 2016; Van der Gun et al., 2017; Bliemer & Raadsen, 2020), but also in some capacity-constrained static NL models (Bliemer et al., 2014; Raadsen & Bliemer, 2023) proposed a node model that can be applied in capacity-constrained static NL model. Most contemporary macroscopic node models adhere to seven specific requirements identified by Tampère et al. (2011). These requirements are: (i) *general applicability*, (ii) *flow maximisation*, (iii) *non-negativity of flows, (iv) conservation of vehicles*, (v) *satisfaction of demand and supply constraints*, (vi) *conservation of turning fractions*, and (vii) *satisfaction of the invariance principle*. Node models that satisfy all requirements fall within the *Generic Class of First-order Macroscopic Node Models* (GCNM) (Tampère et al., 2011). Requirement (vi), conservation of turning fractions, and hence requiring first-in-first-out (FIFO) to be

satisfied on each link, is very stringent because if any movement becomes congested, all other movements from the same link are expected to be equally impacted. In other words, a single homogeneous queue is constructed even if, in reality, different queues and delays per movement and/or lane would emerge. Hence, such a *whole link* approach is typically only realistic for nodes where incoming links only have a single lane. Knowing that major urban intersections and motorways junctions typically have multiple lanes, there is a practical need to address these existing limitations. Existing lane-based formulations also ignore the approach lane configuration, which describes the movements that are permitted on each lane, despite the fact that this configuration has a major impact on realised flows. An approach lane can be *reserved* for a single movement, or it can be *shared* by multiple movements.

3.1.2 Illustrative example

Consider a diverge node shown in Fig. 1(a) with a single incoming link $i \in \{1\}$, three outgoing links $j \in \{2,3,4\}$, and three movements, namely (1,2), (1,3) and (1,4). A conventional link-based node model does not consider any approach lane configuration and implicitly assumes that all lanes can be used for all movements, see Fig. 1(b). In this example we assume the approach lane configuration shown in Fig. 1(c) with two shared lanes, where lane 1 allows left and straight movements and lane 2 allows straight and right movements. Our novel lane-based node model takes this configuration explicitly into account. Assume a capacity of 1000 veh/h for all lanes.

Assume a sending flow on link 1 of 1500 veh/h, where 600 veh/h turn left, 600 veh/h go straight, and 300 veh/h turn right. Fig. 2(a) graphically illustrates this situation for the link-based node model, where all sending flow can exit the node without interruption. Fig. 2(b) shows the same situation for our lane-based node model, which has the same total realised flow (because of free-flow conditions). Based on an equilibrium principle that will be discussed later, the distribution of flow across lanes is such that both lanes have equal total flow, which means that for the straight movement (1,3), 150 veh/h choose the left lane while 450 veh/h choose the right lane.

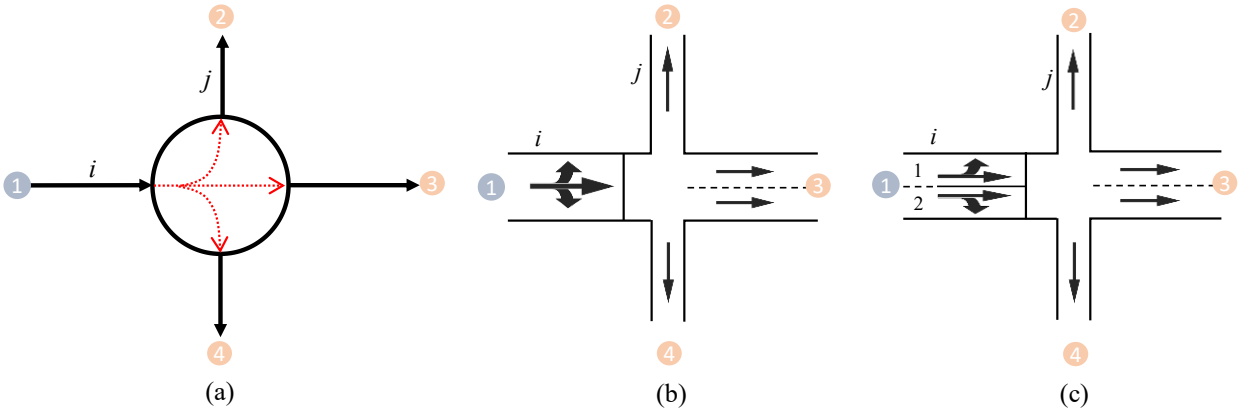


Fig. 1. Diverge node example, (a) model representation with incoming/outgoing links and allowable movements, (b) no approach lane configuration considered in link-based node model, (c) approach lane configuration considered in lane-based node model.

Now consider the same example but with link 2 experiencing an inflow restriction of 400 veh/h due to spillback of a downstream queue. In the link-based node model, only 2/3 of the sending

flow towards link 2 can exit, which also impedes the exit flows to *all* other outgoing links such that only 1000 veh/h can exit while a queue builds up at a rate of 500 veh/h on link 1, see Fig. 2(c). In contrast, in our lane-based node model only lane 1 will be congested while drivers going straight or turning right choose to use lane 2 to exit unimpeded. This results in a total realised flow of 1300 veh/h, see Fig. 2(d), a difference of 300 veh/h compared to the existing link-based node model. In our approach, lanes will only ever be considered within the node model; it is important that they are not expanded as separate links in the network as this would exponentially grow the number of routes in the network and would make equilibrium traffic assignment infeasible on large networks.

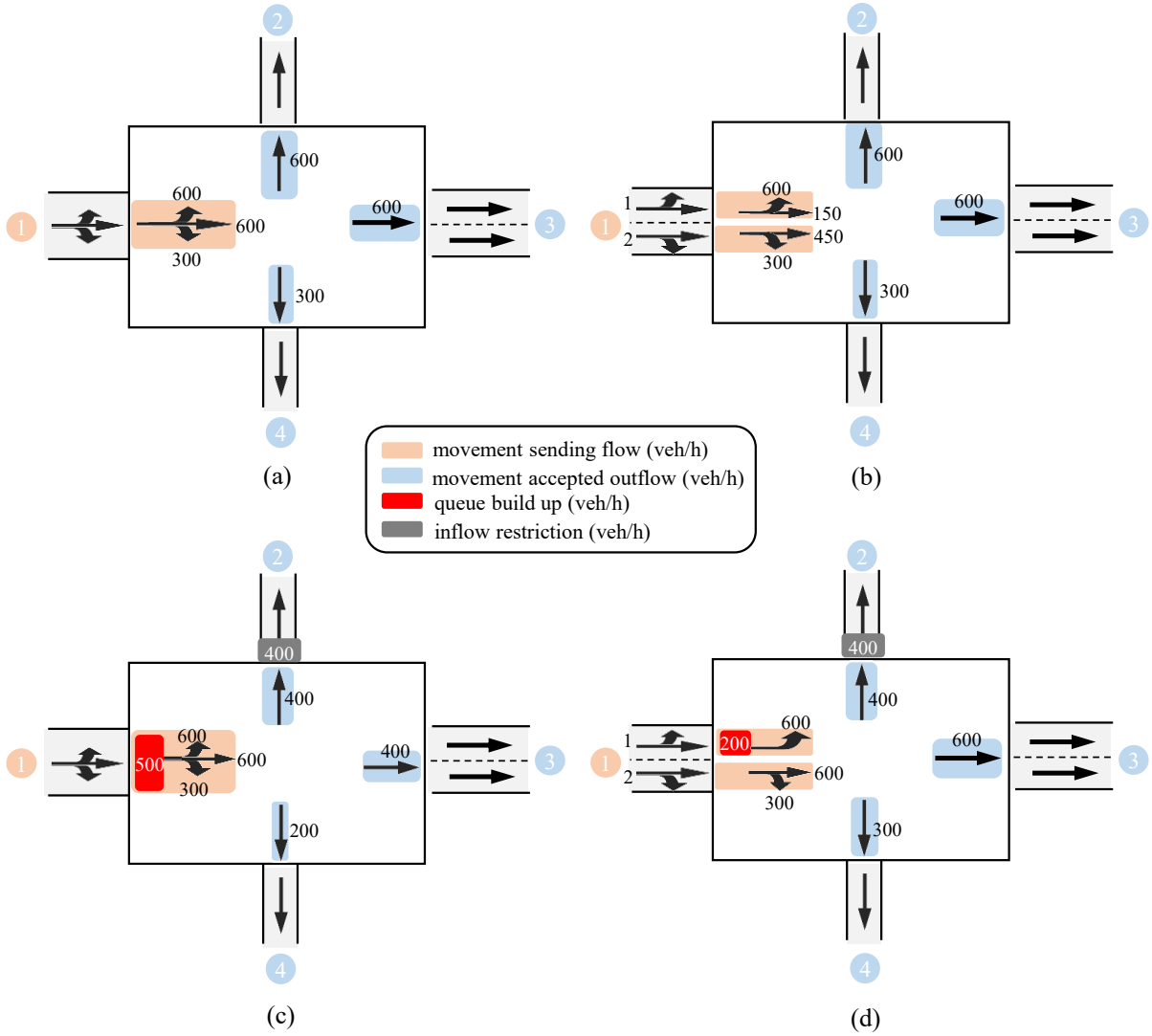


Fig. 2. Diverge node example, (a) link-based node model results in free-flow, (b) lane-based node model results in free-flow, (c) link-based node model results in congestion, (d) lane-based node model results in congestion.

3.1.3 Contribution and outline

The node model proposed in this paper extends the existing link-based formulation of the node model to a lane-based formulation. This lane-based formulation explicitly accounts for the approach lane configuration while distributing flows across lanes based on an equilibrium principle. The result is a node model that relaxes link-based FIFO to movement-based FIFO in which each movement may have a separate queue.

In this work we make the following main contributions: (1) formulation of a novel lane-based node model to relax link-based FIFO for nodes where incoming links have multiple lanes; (2) introduction of a novel equilibrium principle to distribute sending flows across lanes that considers the approach lane configuration and satisfies movement-based FIFO, formalised by means of a variational inequality problem formulation; (3) development of a general solution algorithm to solve the extended node model. In addition, we also provide numerical examples and a case study to illustrate the viability and usefulness of our extended node model.

The remainder of the paper is structured as follows. Section 3.2 provides an overview of the current state-of-the-art in macroscopic node models. Section 3.3 presents the foundations of the conventional link-based node model. The lane-based node model is formulated in Section 3.4, and in Section 3.5 the equilibrium distribution of sending flows across lanes is formalised. Properties of this equilibrium solution are discussed in Section 3.6. A solution algorithm for solving the lane-based node model with equilibrium sending flows is provided in Section 3.7. A case study is presented in Section 3.8 to demonstrate the proposed node model extensions and compare its results with those of the conventional link-based node model in Tampère et al. (2011). Section 3.9 provides concluding remarks and discusses limitations and potential avenues for future research.

3.2 Literature review

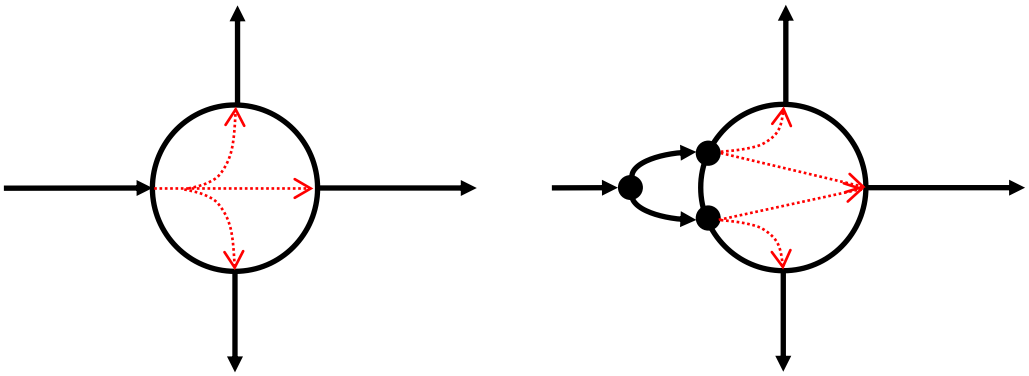
The simplified first order node models proposed by Daganzo (1994, 1995) are among the earliest macroscopic node models. These models describe merges and diverges separately and are therefore not generally applicable. Daganzo's FIFO diverge rule has been adopted by the majority of subsequent node models as with a single incoming link it is straight forward to solve due to the lack of competition for outgoing supply. A node model for merges however is more challenging as multiple sending flows are competing for a single supply on the outgoing link. To overcome this, a fairness rule to determine the merging priority was proposed by Jin and Zhang (2003). Various alternative schemes with different merging rules have been proposed since (Lebacque & Khoshyaran, 2005; Ni & Leonard, 2005; Jin, 2010).

Among the first general node models, supporting multiple incoming and outgoing links, are the models proposed by Jin and Zhang (2004) and Bliemer (2007). The merging behaviour in these node models however is demand-proportional, meaning that whenever the sending flow changes, the resulting realised flow will also change. While this is logical in uncongested conditions, in congested conditions this is not realistic because adding more vehicles to the end of a queue should not impact the amount of flow that is accepted at the front. This undesirable property is an example of violating the *invariance principle* (Lebacque & Khoshyaran, 2005). Tampère et al. (2011) highlighted such shortcomings of existing node models and proposed seven requirements defining a General Class of first order Node Models (GCNM), see Section 3.1.1. Simultaneously, an alternative node model adhering to the same principles but with more

detailed flow constraints was proposed by Flötteröd and Rohde (2011) and leveraged the incremental transfer principle proposed by Daganzo et al. (1997). These two node models are both capacity-proportional, meaning that the incoming links' capacity is used to dictate the priority between competing sending flows. GCNMs used in existing static and dynamic NL models conform to link-based FIFO.

The above-mentioned models all yield a unique solution in terms of flows. When considering internal conflicts and priorities among movements, however, non-uniqueness might occur as was found by Corthout et al. (2012). To address non-uniqueness, Jabari (2016) proposed a technique to resolve the issue by prohibiting conflicting movements from entering the node simultaneously. This method, however, violates the flow maximisation requirement of GCNM. To better capture the randomness and heterogeneity of traffic conditions and drivers' behaviour, Lu and Osorio (2022) proposed a probabilistic node model to capture the variability and uncertainty of traffic flow at intersections. They introduced a node model that considers the intricate dynamics of intersections, such as movements, signal timings, and queue spillbacks. However, their proposed node model also violates the flow maximisation requirement and does not explicitly incorporate approach lane configurations.

Conservation of turning fractions is one of the key requirements of GCNM and is closely related to its link-based FIFO restriction, yielding a single queue regardless of the approach lane configuration. The example in Section 3.1.2 demonstrated that the conventional link-based node model can yield unnecessarily reduced traffic flows when some but not all movements are congested. To date, as far as the authors are aware, only two attempts have been made to address this issue. First, an approach detailed by Shiomi et al. (2015) explicitly expands all lanes to links to circumvent this problem and instead "trick" the node model into treating each lane as a link. This is illustrated in Fig. 3, which provides a lane-expanded representation of the same multilane intersection in Fig. 1(a). This way, existing node models can be applied while link-based FIFO is effectively relaxed to lane-based FIFO. While attractive at first glance, this naïve approach makes it almost impossible to find a route choice equilibrium in practical networks that have hundreds or thousands of multilane junctions or intersections because the number of routes in the network would increase exponentially. For example, drivers travelling from west to east in Fig. 3 have two lanes to choose from, and if they encounter just five of such intersections there would already be $2^5=32$ different routes in the lane-expanded network to choose from. As an alternative to explicit lane expansion, Wright et al. (2017) proposed a partial link-based FIFO mechanism that relaxes the strict link-based FIFO assumption to achieve a situation where some queues only affect specific movements but not others. However, they do not model the interactions between lanes to determine the magnitude of this interaction, instead they require the user to provide a single exogenous parameter that drives this effect. Recent work of Yahyamozdarani and Tampère (2023) extended the conventional link-based node model with a focus on traffic signal control but did not explicitly consider approach lanes.



(a) (b)

Fig. 3. Link model representation, (a) Traditional, (b) Lane-expanded approach that does not scale to practical applications.

To the best of our knowledge, our lane-based node model is the first to capture the impact of multilane junctions and intersections endogenously, in a behaviourally justifiable way, without unwanted side-effects, and does so in an efficient way making it suitable for large scale applications. Realised flows account for movement-specific queues and delays that better reflect reality compared to conventional link-based node models.

3.3 Conventional link-based node model

This section describes conventional link-based node models. Let $G=(N,A)$ be a directed graph representing a road network, where N is the set of nodes and A is the set of directed links. Each node $n \in N$ has a set of incoming links $A_n^{\text{in}} \subset A$ and a set outgoing links $A_n^{\text{out}} \subset A$. Node models are applied for each node and for each time instant, but for notational convenience we omit the node and time indices from all variables in the remainder of this paper.

Consider a specific node and time instant. For convenience we renumber and use index $i=1,\ldots,I$ to denote incoming links and $j=1,\ldots,J$ to denote outgoing links, respectively, where I and J are the cardinality of sets A_n^{in} and A_n^{out}, respectively. A movement is indicated by a combination of an incoming and outgoing link, (i, j). Sending flows describe the potential outflow for a movement. Receiving flows describe the potential inflows into each outgoing link. In other words, sending flows reflect movement demand while receiving flows reflect supply, and the purpose of the node model is to determine realised link inflows and outflows based on the interaction between demand and supply.

Let M_i be the number of lanes on incoming link i and let $M=\sum_i M_i$ be the total number of lanes on all incoming links. Without loss of generality, we assume that lanes on each incoming link are numbered consecutively (from left to right or right to left). Lane exit capacities are exogenously given by $M\times 1$ vector $\mathbf{c}=[c_m]$ with positive elements. The layout of the intersection or junction that the node represents is described by a $I\times M$ matrix of link-lane incidence indicators $\boldsymbol{\delta}=[\delta_{im}]$, where $\delta_{im}=1$ if lane m is on incoming link i and zero otherwise, and a $M\times J$ matrix of approach lane configuration indicators $\boldsymbol{\eta}=[\eta_{mj}]$, where $\eta_{mj}=1$ if lane m permits travel in the direction of outgoing link j and zero otherwise. For each movement with positive sending flow we assume that there exists at least one lane that permits this movement.

The conventional link-based node model does not consider individual lanes, therefore all variables are lane-aggregated. All lane-aggregated variables in this paper are indicated with a dot ($\bullet$) as superscript. Such a conventional node model considers the following variables as input: (i) an $I \times J$ matrix of sending flows for each incoming link and direction, $\mathbf{s}^\bullet = [s_{ij}^\bullet]$, (ii) a $1 \times J$ vector of link-specific receiving flows, $\mathbf{r}^\bullet = [r_j^\bullet]$, and (iii) an $I \times 1$ vector of link exit capacities to determine the merge priorities, $\mathbf{c}^\bullet = [c_i^\bullet]$, where $c_i^\bullet = \sum_m \delta_{im} c_m$ is the sum of lane exit capacities on link i. The first two inputs come from the link model, e.g., a cell or link transmission model. The conventional node model does not consider approach lane configuration $\boldsymbol{\eta}$. Instead, it implicitly assumes $\eta_{mj} = 1$ for all lanes m and all directions j, i.e., it assumes that each approach lane can be used for all directions. Observe that this is not physically possible for links with multiple lanes as it would create conflicts on the road. This essentially invalidates the conventional link-based model except for nodes where all incoming links have a single lane.

The node model output is given by an $I \times J$ matrix of direction-specific link outflow rates $\mathbf{v}^\bullet = [v_{ij}^\bullet]$ via an implicit function $\Gamma(\cdot)$ that describes a general node model,

$$\mathbf{v}^\bullet = \Gamma(\mathbf{s}^\bullet, \mathbf{r}^\bullet, \mathbf{c}^\bullet). \tag{1}$$

Define total link sending flows $s_i^\bullet = \sum_j s_{ij}^\bullet$ and total link outflow rates $v_i^\bullet = \sum_j v_{ij}^\bullet$ for all $i = 1, \ldots, I$. In the conventional link-based node model all movements on the same link are constrained equally. In other words, for each incoming link i and direction j with positive sending flow it holds that $s_{ij}^\bullet / v_{ij}^\bullet = s_i^\bullet / v_i^\bullet$. This can be rewritten as $s_{ij}^\bullet / s_i^\bullet = v_{ij}^\bullet / v_i^\bullet$, which indicates link-based FIFO: all vehicles on incoming link i, independent of direction, exit the link in the same order as they entered the link.

Node model $\Gamma(\cdot)$ implicitly determines a set E of supply constrained (congested) links, $E \subseteq \{1, \ldots, I\}$. If link $i \in E$ then it holds that $v_i^\bullet < s_i^\bullet$ and a queue exists at the end of link i, while if link $i \notin E$ then the link is demand constrained (i.e., in free-flow) and it holds that $v_i^\bullet = s_i^\bullet$. First order node models such as Tampère et al. (2011) automatically inflate the sending flow to link capacity if the link is supply constrained, i.e., $s_i^\bullet$ is replaced by $c_i^\bullet$ if $i \in E$.

To illustrate, consider the three-legged multilane intersection in Fig. 4. The input variables are provided in Table 1, assuming a lane capacity of 1000 veh/h. The zeros on the diagonals of $\mathbf{s}^\bullet$ and $\mathbf{v}^\bullet$ indicate that U-turns are not permitted. Note that the total sending flow into link 3 equals $2400 + 1500 + 3900$ veh/h, which exceeds its receiving flow of 3000 veh/h. The node model of Tampère et al. (2011) generates accepted outflows $v_{ij}^\bullet$ as shown in Table 1, where flows out of links 1 and 2 are (capacity) proportionally reduced such that the flow into link 6 equals 3000 (considering inflated sending flows to capacity).

Fig. 5(a) illustrates link-based queues in a conventional link-based node model, where all movements on the same link face the same FIFO queue. In our novel lane-based node model, which will be presented in the next section, all links are first disaggregated into FIFO lanes with lane-specific queues as shown in Fig. 5(b). After that, an equilibrium lane flow distribution will be determined in Section 3.5 with the property that *FIFO holds across all used lanes for the same movement*. In other words, for each movement, queues on used lanes can be merged into a single movement-based FIFO queue, see Fig. 5(c). For this particular example, when

outlink 2 is congested while outlink 3 is uncongested, drivers on link 1 going straight only using lanes 3 and 4, while lanes 1 and 2 are only used by drivers that turn left.

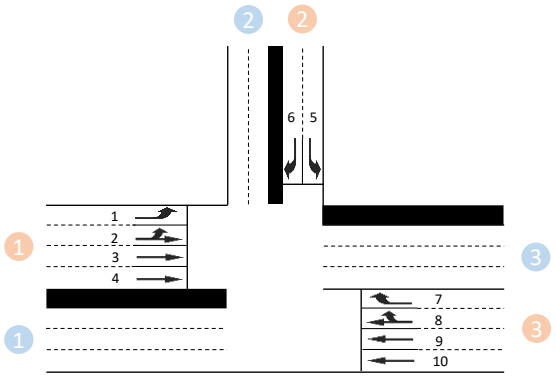


Fig. 4. Three-legged multilane intersection.

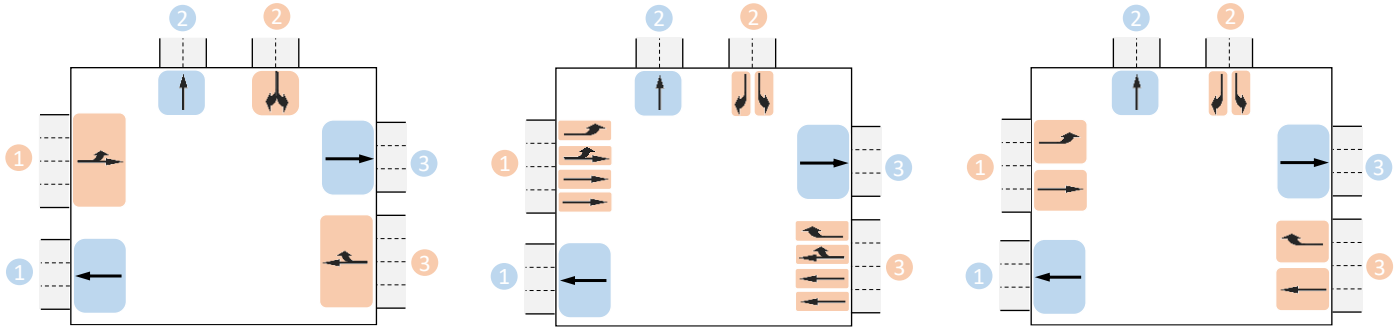


Fig. 5. (a) Link-based queues, (b) Lane-based queues, (c) Movement-based queues

Table 1. Inputs and outputs for three-legged interaction in Fig. 4 using conventional link-based node model.

j	1	2	3		1	2	3
i		$s_{ij}^{\bullet}$		$c_i^{\bullet}$		$v_{ij}^{\bullet}$	
1	0	1200	2400	4000	0	942	1884
2	400	0	1500	2000	298	0	1116
3	1000	500	0	4000	1000	500	0
$r_j^{\bullet}$	3000	2000	3000				

3.4 Lane-based node model

The first node model extension disaggregates each incoming link into lanes by constructing a lane-expanded node, where each lane is essentially considered a separate link – within the node model – to allow separate queues for each lane (Fig. 5(b)). The lane-based node model considers the same three inputs as the link-based node model, namely sending flows, receiving flows, and exit capacities, but now considers an $M \times J$ matrix of sending flows for each lane and direction, $\mathbf{s} = [s_{mj}]$, and the vector of *lane* exit capacities $\mathbf{c}$. The receiving flows remains link-based since only approach lanes on incoming links are considered explicitly, not lanes on outgoing links. This results in the following lane-based node model,

$$\mathbf{v} = \Gamma(\mathbf{s}, \mathbf{r}^{\bullet}, \mathbf{c}), \tag{2}$$

where $\mathbf{v} = [v_{mj}]$ is an $M \times J$ matrix of direction-specific lane outflows. Define total lane m sending flow s_m and total lane m outflow rate v_m as

$$s_m = \sum_j \eta_{mj} s_{mj}, \quad \forall m = 1, \ldots, M, \tag{3}$$

$$v_m = \sum_j \eta_{mj} v_{mj}, \quad \forall m = 1, \ldots, M. \tag{4}$$

Note that function $\Gamma(\cdot)$ is identical to the function in the link-based model, only the dimensions of the inputs and outputs have been modified. Therefore, the same algorithm to solve the link-based node model can be used for solving the lane-based node model. Node model $\Gamma(\cdot)$ now implicitly determines a set E of supply constrained (congested) *lanes*, $E \subseteq \{1, \ldots, M\}$, where $v_m < s_m$ if $m \in E$ and $v_m = s_m$ if $m \notin E$, i.e., if lane m is demand constrained (free-flow). Lane sending flow s_m is now automatically scaled to *lane* capacity c_m within the node model if $m \in E$. In the lane-based node model, *all movements on the same lane are constrained equally*, that is, for each lane m and direction j with positive sending flow it holds that $s_{mj} / v_{mj} = s_m / v_m$. This can be rewritten as $s_{mj} / s_m = v_{mj} / v_m$, which indicates lane-based FIFO. Note that lane-based FIFO is a necessary but not sufficient condition for movement-based FIFO. To achieve movement-based FIFO, a lane choice equilibrium is required, which will be introduced in the next section.

While receiving flows $\mathbf{r}^{\bullet}$ and lane exit capacities $\mathbf{c}$ are direct input into the node model, movement-specific lane sending flows $\mathbf{s}$ needs to be constructed first by distributing sending flows $\mathbf{s}^{\bullet}$ across the lanes. Define the set of movements with positive sending flow by set W, $W = \left\{(i, j), i = 1, \ldots, I; j = 1, \ldots, J \mid s_{ij}^{\bullet} > 0\right\}$. The set of feasible sending flows for each lane and direction, denoted by S, consists of all $\mathbf{s}$ that satisfy the following constraints:

$$\sum_m \delta_{im} \eta_{mj} s_{mj} = s_{ij}^{\bullet}, \quad \forall (i, j) \in W, \tag{5}$$

$$s_{mj} \begin{cases} \geq 0, & \text{if } \eta_{mj} = 1, \\ = 0, & \text{if } \eta_{mj} = 0, \end{cases} \quad \forall m = 1, \ldots, M; \forall j = 1, \ldots, J. \tag{6}$$

Eq. (5) describes flow conservation constraints, while Eq. (6) describes non-negativity constraints as well as approach lane constraints where the sending flow on a lane in a specific

direction is set to zero if this lane does not permit movement in that direction. If all lanes on a link are reserved for a single movement, then the sending flow distribution on this link becomes trivial. For example, link 2 in Fig. 4 would yield $s_{5,3} = s^{\bullet}_{2,3}$ (all sending flow on link 2 in the direction of link 3 is assigned to lane 5) and $s_{6,1} = s^{\bullet}_{2,1}$ (all sending flow on link 2 in the direction of outgoing link 1 is assigned to lane 6), while $s_{5,1} = s_{5,2} = s_{6,2} = s_{6,3} = 0$. However, in the presence of lanes that are shared by different movements, such as on links 1 and 3 in Fig. 4, an infinite number of ways to distribute the sending flow exists, hence in the next section we present additional criteria to select $\mathbf{s} \in S$ with desirable properties while conforming to expected underlying human behaviour.

Table 2 presents inputs and outputs of the lane-based node model for the three-legged multilane intersection in Fig. 4 where sending flows $\mathbf{s}^{\bullet}$ (as shown in Table 1) are distributed across lanes, resulting in movement-specific lane sending flows $\mathbf{s}$. For example, sending flow $s^{\bullet}_{1,2} = 1200$ is distributed across lanes 1 and 2, namely $s_{1,2} = 1080$ and $s_{2,2} = 120$. This distribution is based on an equilibrium in which drivers select lanes such that they maximise their outflow potential, or minimise their generalised lane cost h_m, as will be discussed in the next section. It is clear that in this case lanes 1 to 5 are congested (supply constrained) as their outflows are smaller than their sending flows, while all other lanes experience free-flow conditions (demand constrained).

Table 2. Inputs and outputs for three-legged interaction in Fig. 4 using lane-based node model.

	j	1	2	3		1	2	3	
i	m		s_{mj}		c_m		v_{mj}		h_m
1	1	0	1080	0	1000	0	1000	0	1.080
1	2	0	120	720	1000	0	111	667	1.080
1	3	0	0	840	1000	0	0	778	1.080
1	4	0	0	840	1000	0	0	778	1.080
2	5	0	0	1500	1000	0	0	778	1.928
2	6	400	0	0	1000	400	0	0	0.400
3	7	0	375	0	1000	0	375	0	0.375
3	8	250	125	0	1000	250	125	0	0.375
3	9	375	0	0	1000	375	0	0	0.375
3	10	375	0	0	1000	375	0	0	0.375
	$r^{\bullet}_j$	3000	2000	3000					

3.5 Equilibrium flow distribution of sending flows across lanes

To determine lane-based sending flows $\mathbf{s}$ endogenously in the node model, we propose the novel concept of a (deterministic) *lane choice equilibrium*, similar to Wardrop's equilibrium law for route choice. In exiting a link, drivers are assumed to select the lane that provides *maximum outflow potential*, noting that flow maximisation is one of the seven requirements of GCNM. In a lane choice equilibrium, no driver can improve their outflow potential by unilaterally changing lanes. As a result, in equilibrium, all used lanes for a movement have the same outflow potential that is larger than the outflow potential of any unused lane. While true equilibria are rarely observed in practice, it provides a meaningful way to distribute traffic flow

based on rational decision making. Further, if the analyst is interested in a (deterministic) user equilibrium at the network level, which is often the case in strategic transport planning applications where multiple scenarios are to be compared, lane choice equilibrium in each node is a necessary condition and implicitly assumed.

Note that a node model only considers vehicles at the end of the link (i.e., the sending flow), therefore our model considers lane choice when exiting the link, not lane choice when entering a queue on a link (which would be part of the link model). Consider a driver waiting at the end of link i to exit in the direction of link j. Suppose that all available lanes for this movement are supply constrained (congested). Then this driver can maximise their outflow potential by choosing the lane with the highest outflow relative to sending flow, v_m / s_m. The inverse of lane outflow potential, s_m / v_m, expresses the marginal increase in lane queuing delay and can be thought of as a generalised lane exit cost. Now suppose that all available lanes are demand constrained (free-flow). Then the driver can maximise their outflow potential by choosing the lane with the highest capacity relative to sending flow, c_m / s_m. The inverse of lane outflow potential in this case, s_m / c_m, is again some generalised lane exit cost where relatively busy lanes are less attractive because there is more competition for exit capacity. If available lanes are a mix of supply and demand constrained, then clearly the strategy to maximise outflow potential means trying to switch to a demand constrained lane.

We define the following generalised lane exit cost function for lane m, which is identical for all directions j,

$$h_{mj} = h_m = \left(\frac{s_m}{c_m}\right)\left(\frac{\tilde{s}_m}{v_m}\right), \qquad \forall m = 1, \ldots, M, \forall j = 1, \ldots, J, \tag{7}$$

where the adjusted lane sending flow $\tilde{s}_m$ is commonly scaled to capacity in case of congestion,

$$\tilde{s}_m = \begin{cases} s_m, & \text{if } m \notin E, \\ c_m, & \text{if } m \in E, \end{cases} \qquad \forall m = 1, \ldots, M. \tag{8}$$

Combining Eqs. (7) and (8), together with the fact that $s_m = v_m$ if lane m is demand constrained, leads to

$$h_m = \begin{cases} \dfrac{s_m}{c_m}, & \text{if } m \notin E, \\ \dfrac{s_m}{v_m}, & \text{if } m \in E, \end{cases} \qquad \forall m = 1, \ldots, M. \tag{9}$$

Since drivers are assumed to minimise generalised lane exit cost h_m, it follows, by definition, that demand constrained lanes with free-flow conditions are preferred over supply constrained lanes with congested conditions since $s_m / c_m \leq 1$ and $s_m / v_m > 1$.
We can now commence with formulating the lane choice equilibrium problem more formally. We start by defining L_{ij} as the set of lanes that are available for movement (i, j),

$$L_{ij} = \left\{ m = 1, \ldots, M \mid \delta_{im} \eta_{mj} = 1 \right\}, \quad \forall (i, j) \in W. \tag{10}$$

Let $\bar{h}_{ij}$ be the minimum generalised lane exist cost for movement (i, j),

$$\bar{h}_{ij} = \min_{m \in L_{ij}} h_m, \quad \forall (i, j) \in W. \tag{11}$$

We then introduce the lane choice equilibrium conditions for equilibrium sending flows $\bar{\mathbf{s}}$ in which all used lanes have equal generalised link exit cost. This cost is lower than or equal to the generalised link exit cost for any unused lane:

$$\begin{cases} h_{mj} = h_m = \bar{h}_{ij}, & \text{if } \bar{s}_{mj} > 0, \\ h_{mj} = h_m \geq \bar{h}_{ij}, & \text{if } \bar{s}_{mj} = 0. \end{cases} \quad \forall m = 1, \ldots, M, \text{ where } \delta_{im} = 1; \forall j = 1, \ldots, J. \tag{12}$$

Looking again at the three-legged intersection in Fig. 4 and the equilibrium lane-specific flows in Table 2, it is easy to see that lanes 1 to 5 are supply constrained because $h_m > 1$ while all other lanes are demand constrained because $h_m \leq 1$. Lanes 1 and 2 can be used for movement $(1,2)$, i.e., $L_{1,2} = \{1, 2\}$, and the lane sending flows for this movement are in equilibrium because $h_1 = h_2$. Similarly, lane sending flows are in equilibrium for movement $(1,3)$ because $L_{1,3} = \{2, 3, 4\}$ and $h_2 = h_3 = h_4$. Similar arguments can be made to show that the lane sending flows are also in equilibrium for the other movements.

According to Proposition 1, equilibrium movement-specific lane-based sending flows $\bar{\mathbf{s}}$ can be determined by solving the following variational inequality (VI) problem:

$$\sum_i \sum_j \sum_m \delta_{im} h_m(\bar{\mathbf{s}}) \left(s_{mj} - \bar{s}_{mj} \right) \geq 0, \quad \forall \mathbf{s} \in S. \tag{13}$$

Proposition 1. Movement-specific lane sending flows $\bar{\mathbf{s}}$ constitute a lane choice equilibrium that maximises outflow potential if and only if $\bar{\mathbf{s}}$ solves VI problem (13).

Proof: The proof follows the same line of reasoning as for Theorem 4.2 in Chen (1999) and is applied separately for both demand constrained movements and supply constrained movements, see Appendix B. ■

As will be shown in the next section, in equilibrium, FIFO is satisfied across all lanes used in the same movement, hence lane-specific outflows can be aggregated to movement-specific outflows without loss of queuing delay information,

$$v_{ij}^{\bullet} = \sum_m \delta_{im} v_{mj}, \quad \forall i = 1, \ldots, I; \forall j = 1, \ldots, J. \tag{14}$$

This allows for direct comparisons with outflows resulting from the link-based node model. For example, Table 3 shows lane-aggregated outflows of the lane-based node model, which can be directly compared to the outflows of the link-based node model in Table 1. Subtantial differences can be observed: outflow rates of the lane-based node model are 18% larger for movements (1,2) and (1,3), 34% larger for movement (2,1), while they are 30% smaller for movement (2,3).

Table 3. Lane-aggregated inputs and outputs of lane-based node model.

j	1	2	3		1	2	3
i		$s_{ij}^{\bullet}$		$c_i^{\bullet}$		$v_{ij}^{\bullet}$	
1	0	1200	2400	4000	0	1111	2223
2	400	0	1500	2000	400	0	778
3	1000	500	0	4000	1000	500	0
$r_j^{\bullet}$	3000	2000	3000				

3.6 Equilibrium properties

In this section we discuss properties of equilibrium solution $\bar{\mathbf{s}}$ via a series of propositions and proofs.

According to Proposition 2, movement-based FIFO is satisfied in the equilibrium solution. This is an important property as it avoids the need to keep track of lane inflows and outflows to determine link travel times in network loading. Rather, it suffices to keep track of inflows and outflows aggregated across movements.

Propositions 3 and 4 states that a solution exists and that it is unique, which ensures general applicability with consistent solutions.

Finally, Proposition 5 states that the conventional link-based node model is a special case of our lane-based node model under equilibrium lane-based sending flows. It shows that the lane-based node model collapses to the link-based node model if each approach lane on an incoming link permits all movements. This means that the proposed node model is a true extension of the link-based node model by relaxing the (unrealistic) assumption of $\boldsymbol{\eta} = \mathbf{1}$ for multilane incoming links.

Proposition 2. If movement-specific lane-based sending flows $\bar{\mathbf{s}}$ satisfy equilibrium conditions (12) then FIFO is satisfied not only at the lane-level but also at the movement-level.

Proof: According to Eq. (9) it holds that $h_m = s_m / v_m$ if lane m is supply constrained, where $s_m = \sum_j \eta_{mj} s_{mj}$ and $v_m = \sum_j \eta_{mj} v_{mj}$. Let $\bar{\mathbf{v}}$ denote the output of the lane-based node model based on equilibrium sending flows $\bar{\mathbf{s}}$. It then follows from the equilibrium conditions (12) that for any two used lanes m and m', $h_m = \bar{s}_m / \bar{v}_m = \bar{s}_{m'} / \bar{v}_{m'} = h_{m'}$, which can be rewritten as $\bar{s}_m / \bar{s}_{m'} = \bar{v}_m / \bar{v}_{m'}$. This implies that, in equilibrium, FIFO is satisfied not only within each lane but also across lanes, such that in equilibrium movement-based FIFO is satisfied. Note that FIFO is satisfied by definition when lanes are demand constrained as drivers can exit with zero delay. ■

Proposition 3. There exists a solution to VI problem (13).

Proof: The set of feasible movement-specific lane sending flows, S, is a convex compact set. It is also non-empty because we assumed that for each movement with positive sending flow there exists at least one approach lane that permits this movement. If generalised lane exit cost functions $h_m(\cdot)$ are continuous then the VI problem has a solution, see Nagurney (1998). Function $h_m(\cdot)$ exhibits a discontinuity when it changes from demand constrained to supply constrained or vice versa. However, to prove existence of a solution to the VI problem, it

suffices that $h_m(\cdot)$ is continuous across lane choice set L_{ij} for each movement (i, j). According to equilibrium conditions (12), all used routes in L_{ij} have equal generalised lane exit costs. This can only be the case if all used lanes are demand constrained or all used lanes are supply constrained. If they are all demand constrained, then $h_m = s_m / c_m$ applies to all relevant lanes and since this function is continuous there will exist a solution to the VI problem. If all used lanes are supply constrained, then $h_m = s_m / v_m$ applies to all relevant lanes. Lane exit flows v_m are derived via node model function $\Gamma(\cdot)$, which is continuous if the invariance principle is satisfied (Tampère et al., 2011) and the sending flow is non-zero (Gibb, 2011). Since we assumed that all lane exit capacities are positive, lanes will have positive flow unless they are only available for movements with zero sending flow. Movements with zero sending flow can simply be omitted from consideration in the node model. Hence also in the case that all used lanes are supply constrained function $h_m(\cdot)$ is continuous and a solution to the VI problem exists. ■

Proposition 4. The solution to VI problem (13) is unique.

Proof: Since a solution exists according to Proposition 3, a sufficient condition for the equilibrium solution to the VI problem to be (locally) unique is that the generalised lane exit cost function $h_m(\cdot)$ is (locally) strictly monotone for each lane m (Nagurney, 1998). Since in equilibrium all used lanes for each movement are either all demand constrained or all supply constrained, we prove that the solution is unique in both situations. If lane m is demand constrained then $h_m = s_m / c_m$ and it is easily observed that this is a strictly monotone function. If lane m is supply constrained, then $h_m = s_m / v_m$ where v_m is determined via node model function $\Gamma(\cdot)$. We need to prove that an increase in sending flow s_m results in an increase in h_m, which means that v_m should increase less than the increase in s_m. When lane m supply constrained it holds that $s_m > v_m$. Due to satisfaction of the invariance principle, an increase in the sending flow will not lead to an increase in outflow. Therefore, $h_m(\cdot)$ is strictly monotone both for demand and supply constrained conditions. When increasing the sending flow on a lane, the lane may switch from demand constrained to supply constrained (but not vice versa), which results by definition in an increase in $h_m(\cdot)$. Since $h_m(\cdot)$ is strictly monotone for each lane m, function $\mathbf{h}(\cdot) = [h_m(\cdot)]$ is also strictly monotone. Therefore, there will only be a single solution to VI problem (13). ■

Proposition 5. If $\boldsymbol{\eta} = \mathbf{1}$ the lane-based node model has equilibrium sending flows shown in Eq. (15) and yields identical outflow per movement as a capacity-proportional link-based node model such as described in Tampère et al. (2011) or Flötteröd and Rohde (2011).

$$\overline{s}_{mj} = \sum_i \delta_{im} \left(\frac{c_m}{c_i^\bullet} \right) s_{ij}^\bullet, \quad \forall m = 1, \ldots, M; \forall j = 1, \ldots, J. \tag{15}$$

Proof: If each approach lane on an incoming link allows all movements, then in equilibrium all lanes on an incoming link are either demand constrained or supply constrained and each should have the same generalised lane exit cost. Consider an incoming link i and consider all lanes on this link, i.e., all lanes m where $\delta_{im} = 1$. If these lanes are all demand constrained, then they all have the same generalised lane exit cost h_i,

$$\frac{\overline{s}_m}{c_m} = \frac{1}{c_m} \sum_j \left(\frac{c_m}{c_i^\bullet} \right) s_{ij}^\bullet = \sum_j \frac{s_{ij}^\bullet}{c_i^\bullet} \equiv h_i. \tag{16}$$

If these lanes are all supply constrained then the outflow of each lane is capacity proportional, which means that for each two lanes m and m' on link i it holds that $v_m / v_{m'} = c_m / c_{m'}$. Using Eq. (15) we can derive that

$$\frac{\bar{s}_m}{\bar{s}_{m'}} = \frac{\left(\frac{c_m}{c_i^\bullet}\right)\sum_j s_{ij}^\bullet}{\left(\frac{c_{m'}}{c_i^\bullet}\right)\sum_j s_{ij}^\bullet} = \frac{c_m}{c_{m'}} = \frac{v_m}{v_{m'}}. \tag{17}$$

This means that all lanes m on link i have the same generalised link exit cost $s_m / v_m \equiv h_i$. Since we have shown that for any incoming link i the generalised lane exit costs are equal across all lanes, for demand as well as supply constrained conditions, we conclude that the sending flows in Eq. (15) are indeed an equilibrium.

Let $v_{ij}^\bullet$ be the outflow for movement (i, j) resulting from the link-based node model. Since the lane-based node model uses the same capacity-proportional function $\Gamma(\cdot)$ where only its input has scaled down from link flows to lane flows while maintaining the same split proportions on each lane, it follows that output from the lane-based node model is $v_{mj} = (c_m / c_i^\bullet) v_{ij}^\bullet$ for any lane m on incoming link i. Therefore, the outflow for movement (i, j) in the lane-based node model equals the outflow from the link-based node model since

$$\sum_m \delta_{im} v_{mj} = \sum_m \delta_{im} \left(\frac{c_m}{c_i^\bullet}\right) v_{ij} = \frac{v_{ij}}{c_i^\bullet} \sum_m \delta_{im} c_m = \frac{v_{ij}}{c_i^\bullet} c_i^\bullet = v_{ij}. \tag{18}$$

■

3.7 Algorithm

3.7.1 Iterative framework

The iterative algorithm proposed in this section determines equilibrium movement-specific lane sending flows $\bar{\mathbf{s}}$. A schematic interpretation of how the lane-based node model variables relate to the algorithm structure is provided in Fig. 6. To compare, we show the schematic diagram of the link-based node model in Fig. 7 where it becomes clear that the same node model function $\Gamma(\cdot)$ is used with the same input and output, except for approach lane configuration $\boldsymbol{\eta}$, and where the link based model relies on the lane aggregated sending flows and capacities.

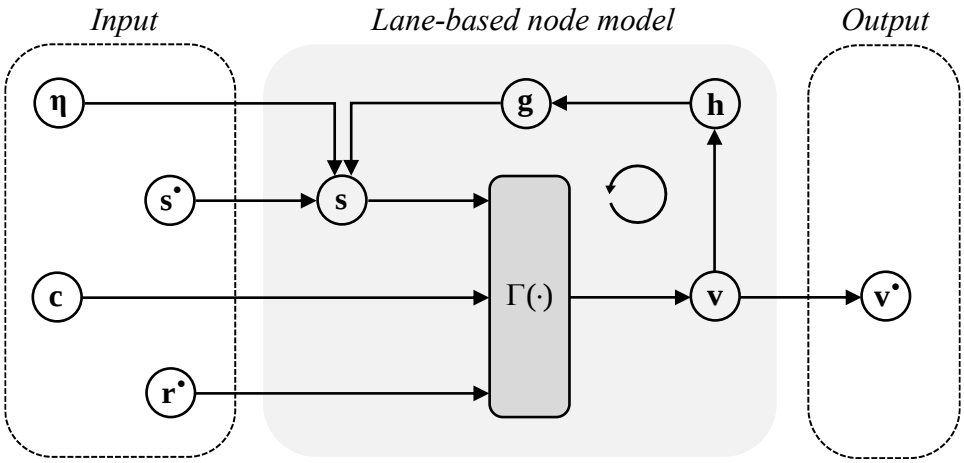


Fig. 6. Schematic diagram of relationships in lane-based node model with equilibrium lane sending flow distribution.

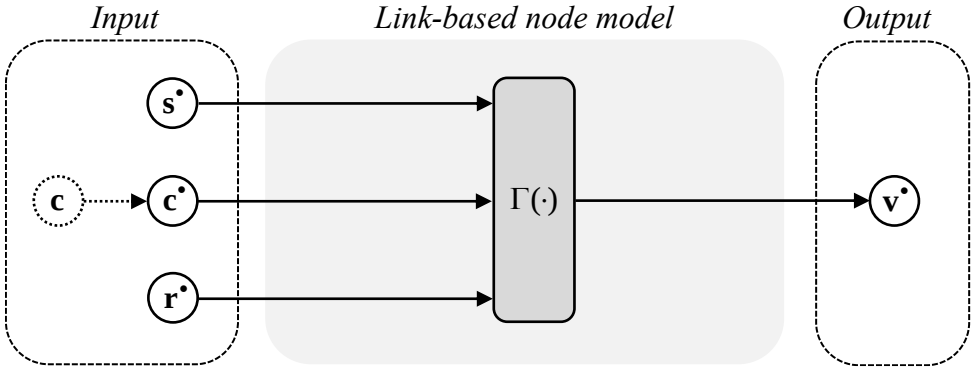


Fig. 7. Schematic diagram of relationships in conventional link-based node model.

3.7.2 System of linear equations

We now formulate a system of linear equations that determines sending flow distribution **s**. Firstly, following Eq. (12), in equilibrium it should hold that $h_m = h_{m+1}$ if lanes m and $m+1$ are both available and used for the same movement (making use of the fact that lanes were assumed to be consecutively numbered on each link). Using Eqs. (9) and (3) we can rewrite $h_m = h_{m+1}$ as

$$\delta_{im} g_m \sum_j \eta_{mj} s_{mj} - \delta_{i,m+1} g_{m+1} \sum_j \eta_{m+1,j} s_{m+1,j} = 0, \quad \forall m = m'_{ij}, \ldots, m'_{ij} + |L_{ij}| - 2; \forall (i,j) \in W, \tag{19}$$

where $m'_{ij} = \min\{L_{ij}\}$ is the lane with the lowest number in set L_{ij}, and g_m is the gradient of h_m,

$$g_m = \begin{cases} \dfrac{1}{c_m}, & \text{if } m \notin E, \\ \dfrac{1}{v_m}, & \text{if } m \in E, \end{cases} \qquad \forall m = 1, \ldots, M. \tag{20}$$

Note that Eq. (19) only needs to be satisfied if multiple lanes exist for the same movement, i.e., if $|L_{ij}| \geq 2$. Further, any movement-specific lane sending flow allocation $\mathbf{s} \in S$ needs to satisfy flow conservation constraints Eq. (5), and non-negativity and approach lane constraints Eq. (6). Combining the linear equality constraints in Eqs. (5) and (6) with the linear equality constraint in Eq. (19), and temporarily ignoring the inequality constraints for non-negativity in Eq. (6) – which we will consider later – allows us to formulate the following system of linear equations,

$$\mathbf{B}\,\mathrm{vec}(\mathbf{s}) = \mathbf{p}, \tag{21}$$

where we use notation $\mathrm{vec}(\mathbf{x})$ to denote the operator that transforms a matrix $\mathbf{x}$ into a column vector by vertically stacking the columns of the matrix, $\mathbf{B}$ is a block diagonal matrix with blocks for each incoming link i, $\mathbf{b}_1, \ldots, \mathbf{b}_I$, and $\mathbf{p}$ is a column vector with stacked vectors $\mathbf{p}_1, \ldots, \mathbf{p}_I$. Because $\mathbf{B}$ is block diagonal, its inverse is also block diagonal, which means that the system of equations can be solved for each incoming link i separately,

$$\mathrm{vec}(\mathbf{s}_i) = \mathbf{b}_i^{-1}\mathbf{p}_i, \tag{22}$$

where $\mathbf{s}_i$ refers to the rows (lanes) in matrix $\mathbf{s}$ that correspond to link i.

Consider incoming link i. Flow conservation constraints Eq. (5) appear for each movement $(i, j) \in W$ as a row in $M_i J \times M_i J$ matrix $\mathbf{b}_i$ by taking the transpose of the following vectorised $M_i \times J$ matrix,

$$\mathrm{vec}\begin{pmatrix} \mathbf{0} & \delta_{i,m'_{ij}}\eta_{m'_{ij},j} & \mathbf{0} \\ \mathbf{0} & \delta_{i,m'_{ij}+1}\eta_{m'_{ij}+1,j} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \delta_{i,m'_{ij}+|M_i|-2}\eta_{m'_{ij}+|M_i|-2,j} & \mathbf{0} \end{pmatrix}, \tag{23}$$

noting that the associated value of $s_{ij}^{\bullet}$ appears in the corresponding row in $\mathbf{p}_i$. In a similar fashion, equilibrium constraints in Eq. (19) are added for each movement $(i, j) \in W$ whenever $|L_{ij}| \geq 2$. These appear across multiple rows in $\mathbf{b}_i$ by taking the transpose of the following vectorised $M_i \times J$ matrix for each lane $m \in L_{ij}$ (except for the last lane in this set),

$$\mathrm{vec}\begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \delta_{im}\eta_{m,1}g_m & \cdots & \delta_{im}\eta_{m,j}g_m & \cdots & \delta_{im}\eta_{m,J}g_m \\ -\delta_{i,m+1}\eta_{m+1,1}g_{m+1} & \cdots & -\delta_{i,m+1}\eta_{m+1,j}g_{m+1} & \cdots & -\delta_{i,m+1}\eta_{m+1,J}g_{m+1} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{24}$$

and with associate value in the same row in $\mathbf{p}_i$ of zero. The approach lane equality constraints in Eq. (6) are trivial to add in matrix $\mathbf{b}_i$.

For example, consider incoming link 1 in the three-legged multilane intersection in Fig. 4 with equilibrium flows reported in Table 2. Eq. (25) shows what $\mathbf{b}_1\,\mathrm{vec}(\mathbf{s}_1) = \mathbf{p}_1$ looks like, where the first two rows reflect the flow conservation constraints, the next three rows reflect the (supply constrained) equilibrium constraints, and the final seven rows reflect approach lane configuration constraints. Note that the last seven rows in $\mathbf{b}_1$ could be ignored, together with seven corresponding columns, so for computational efficiency one only needs to consider five

rows and five columns to compute $s_{1,2}$, $s_{2,2}$, $s_{2,3}$, $s_{3,3}$, and $s_{4,3}$. However, for the mathematical construction of $\mathbf{b}_i$ and vector $\mathbf{p}_i$ we need to initially consider all sending flows $\mathbf{s}$.

$$\left.\begin{array}{r} s_{1,2}+s_{2,2}=s_{1,2}^{\bullet} \\ s_{2,3}+s_{3,3}+s_{4,3}=s_{1,3}^{\bullet} \\ \hline \frac{s_{1,2}}{v_1}-\frac{s_{2,2}}{v_2}-\frac{s_{2,3}}{v_2}=0 \\ \frac{s_{2,2}}{v_2}+\frac{s_{2,3}}{v_2}-\frac{s_{3,3}}{v_3}=0 \\ \frac{s_{3,3}}{v_3}-\frac{s_{4,3}}{v_4}=0 \\ \hline s_{1,1}=0 \\ s_{2,1}=0 \\ s_{3,1}=0 \\ s_{4,1}=0 \\ s_{3,2}=0 \\ s_{4,2}=0 \\ s_{1,3}=0 \end{array}\right\} \rightarrow \left(\begin{array}{cccccccccccc} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & \frac{1}{v_1} & -\frac{1}{v_2} & 0 & 0 & 0 & -\frac{1}{v_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{v_2} & 0 & 0 & 0 & \frac{1}{v_2} & -\frac{1}{v_3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{v_3} & -\frac{1}{v_4} \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{c} s_{1,1} \\ s_{2,1} \\ s_{3,1} \\ s_{4,1} \\ s_{1,2} \\ s_{2,2} \\ s_{3,2} \\ s_{4,2} \\ s_{1,3} \\ s_{2,3} \\ s_{3,3} \\ s_{4,3} \end{array}\right) = \left(\begin{array}{c} s_{1,2}^{\bullet} \\ s_{1,3}^{\bullet} \\ \hline 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array}\right) \tag{25}$$

The outcome of Eq. (22) may result in negative movement-specific lane sending flows since we, so far, ignored the non-negativity constraints in Eq. (6). If for a certain incoming link i sending flow $s_{mj} < 0$, then lane m should not be used for the corresponding movement. To enforce $s_{mj} = 0$ we set $\eta_{mj} = 0$, then reconstruct matrix $\mathbf{b}_i$ and vector $\mathbf{p}_i$ and solve the system of linear equations again for link i. This process is repeated until all movement-specific lane sending flows are non-negative for all incoming links.

After applying the lane-based node model utilising current solution $\mathbf{s}$, lane outflows $\mathbf{v}$ and set E are updated. This also yields updated generalised link exit costs $\mathbf{h}$ and gradients $\mathbf{g}$, which then may require solving an updated system of linear equations, as discussed next.

3.7.3 Generic solution scheme

The iterative generic solution scheme is provided in Algorithm 1, where each iteration is tracked by iteration counter k. In Step 1, we initialise supply constrained lane set E by assuming that all lanes are demand constrained as this is the most likely outcome for most nodes in a transport network. In Step 2, movement-specific lane sending flows are determined for each incoming link i separately by solving the linear system of linear equations as discussed above. Gradients $\mathbf{g}$ in the first iteration are based on lane exit capacities $\mathbf{c}$ since $E = \varnothing$. If any negative movement-specific lane sending flows are found, then the approach lane configuration $\tilde{\boldsymbol{\eta}}$ is adjusted in Step 2(c) to ensure that no flow is sent to that lane for that movement. Note that such approach lane configuration adjustments may differ in each iteration k, therefore, at the beginning of Step 2 we always revert to the original approach lane configuration as a starting point. In Step 3, the lane-based node model is applied and generalised link exist costs $\mathbf{h}$ are updated. If the following gap function $G^{(k)}$ becomes smaller than a small positive threshold ε,

$$G^{(k)} = \frac{\sum_i \sum_j \sum_m \delta_{im} s_{mj}^{(k)} \left(h_m^{(k)} - \bar{h}_{ij}^{(k)} \right)}{\sum_i \sum_j s_{ij}^{\bullet} \bar{h}_{ij}^{(k)}}, \tag{26}$$

then an equilibrium solution $\bar{\mathbf{s}}$ has been found. If $G^{(k)} > \varepsilon$ then the algorithm has not yet converged and Steps 2 and 3 are repeated.

Algorithm 1. Main solution scheme.

Input: Sending flows $\mathbf{s}^{\bullet}$, inlink lane capacities $\mathbf{c}$, receiving flows $\mathbf{r}^{\bullet}$, approach lane configuration $\boldsymbol{\eta}$, gap threshold ε.

Output: Equilibrium movement-specific lane sending flows $\bar{\mathbf{s}}$.

Step 1: *Initialisation*
Set iteration counter $k := 1$.
Set $E = \varnothing$.

Step 2: *Update movement-specific lane sending flows*
Determine gradients $\mathbf{g}$ using Eq. (20).
Set adjusted approach lane configuration $\tilde{\boldsymbol{\eta}} := \boldsymbol{\eta}$.
Set link index $i := 1$.
(a) Construct matrix $\mathbf{b}_i$ and vector $\mathbf{p}_i$ using $\tilde{\boldsymbol{\eta}}$, following Eqs. (23) and (24)
(b) Determine movement-specific lane flows, $\mathbf{s}_i^{(k)}$, using Eq. (22).
(c) Let $s_{mj}^{\min}$ be the smallest value in $\mathbf{s}_i^{(k)}$. If $s_{mj}^{\min} < 0$ then set $\tilde{\eta}_{mj} := 0$ and return to Step 2(a), otherwise continue.
(d) If $i = I$ then continue to Step 3, otherwise set $i := i + 1$ and return to Step 2(a).

Step 3: *Update generalised lane exit costs*
Use the lane-based node model to compute $\mathbf{v}^{(k)} = \Gamma(\mathbf{s}^{(k)}, \mathbf{r}^{\bullet}, \mathbf{c})$, and update supply constrained lane set *E,* see Appendix A.
Determine adjusted lane sending flows, $\tilde{s}_m^{(k)}$, using Eq. (8).
Update generalised lane exit costs, $\mathbf{h}^{(k)}$, using Eq. (9).

Step 4: *Equilibrium convergence check*
Determine gap $G^{(k)}$ using Eq. (26).
If $G^{(k)} < \varepsilon$, then terminate and set $\bar{\mathbf{s}} := \mathbf{s}^{(k)}$, otherwise, set $k := k + 1$ and return to Step 2.

3.7.4 Algorithm performance

While the solution scheme in Algorithm 1 may seem complex, repeatedly solving a small system of linear equations takes very little computation time. Also, typically, most incoming links at intersections or junctions are demand constrained and for those cases the algorithm terminates after just a single iteration. There also exist numerous implementation strategies to further optimise computational efficiency.

To provide an idea of the performance under various conditions, and some strategies to reduce computational cost by adapting the general algorithm presented earlier, consider Fig. 8. It illustrates the use cases for a specific incoming link of a node. When this link only has a single lane (situation A), then Step 2 in Algorithm 1 can be skipped for this link and the equilibrium solution is trivial. If an incoming link has multiple lanes but set *E* remains empty in Step 3, i.e.,

if all lanes are demand constrained (situation C), then convergence is also reached immediately for this link. Similarly, when no shared approach lanes exist (situation B), the initial solution is also the final solution, only now due to lack of lane interactions. Only in case of *congestion on shared lanes* – which is expected to occur only for a small number of nodes – finding a lane choice equilibrium *may* require more effort, yet it is in those situations with significant delays that it is important to get an accurate delay prediction as they, to a large extent, drive route choice on a network level. In such cases, we distinguish between single or multiple movements being affected. In most situations, the resolution of the system of equations in Step 2 typically converges to equilibrium solutions within a limited number of iterations. Only when multiple shared movements become congested (Situation E), more iterations may be required to achieve convergence.

By tailoring the complexity of the algorithm to the likelihood of occurrence as well as the importance of finding a more accurate solution, computational overhead is minimised and only employed when needed to maximise its practical applicability. Incoming links may also be pre-processed to determine whether there are *potentially* congested movements. For example, if the input results in $\sum_i s_{ij}^{\bullet} > r_j^{\bullet}$ for some outgoing links j, then one or more lanes will be supply constrained and one may initialise E as a non-empty set to speed up convergence. Given the focus of this work on presenting theory and a generally applicable solution scheme, we leave the implementation of those many possible optimisations for future research to not detract from the main focus of this work.

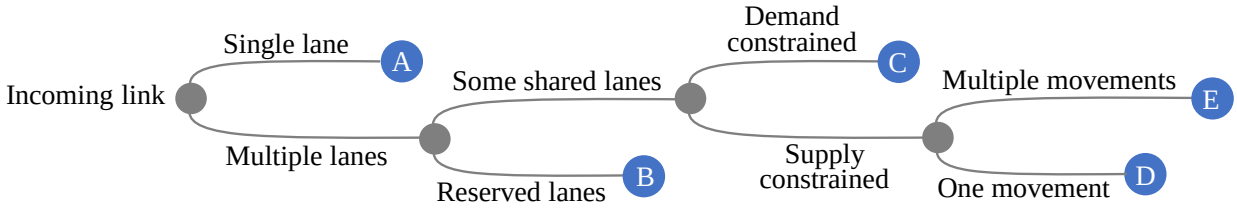


Fig. 8. Possible classifications of situations that may occur when solving lane-based node model.

3.8 Case study

We present a case study adapted from the example originally provided in Tampère et al. (2011), see Fig. 9(a). The original example's link capacities, sending and receiving flows have been doubled to facilitate multiple approach lanes, resulting in the inputs as per Table 4. In this section, we limit ourselves to discussing results, convergence performance, and comparing outcomes of our lane-based node model to outcomes of the link-based node model. Computational details of this case study can be found in Appendix C.

Each approach lane is assumed to have a capacity of 1000 veh/h and we assume an approach lane configuration as per Fig. 9 (b), where links 2 and 4 have four lanes for three movements, and links 1 and 3 have two lanes for three movements.

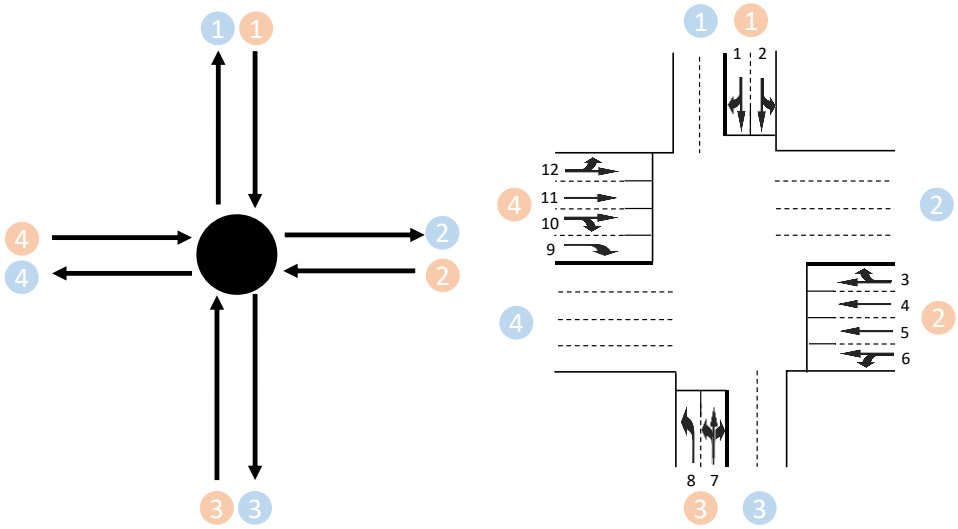


Fig. 9. (a) Intersection example from Tampère et al. (2011), (b) assumed lane configuration.

Table 4. Node model inputs for case study.

i \ j	1	2	3	4	$c_i^{\bullet}$
	$s_{ij}^{\bullet}$				
1	0	100	300	600	2000
2	200	0	600	3200	4000
3	200	200	0	1200	2000
4	200	1600	1600	0	4000
$r_j^{\bullet}$	2000	4000	2000	4000	

The assumed lane configuration was chosen such that the node model becomes challenging to solve and triggers all aspects of the iterative solution scheme. The equilibrium solution in the lane-based node model is given in Table 5. We can observe that both lanes on link 1 are demand constrained and hence outflows are not restricted. Flow for straight movement (1,3) is only assigned to lane 2 since lane 1 already has a large flow turning right. All lanes on link 2 have equal generalised lane exit costs and are supply constrained since outgoing links 3 and 4 have limited receiving flows. Flow for right turning movement (2,3) has no choice but to use lane 3 and will similarly be congested. Both lanes on link 3 are supply constrained where shared lane 7 is used by all three allowable movements. Finally, link 4 has two lanes that are supply constrained, namely lanes 9 and 10, and two lanes that are demand constrained, namely lanes 11 and 12. In equilibrium, shared lane 10 is only used for right turning movement (4,3), while shared lane 12 is used for both allowable movements, i.e., straight and left.

The convergence gap by iteration is provided in Fig. 10, illustrating that in this deliberately complex setup, 14 iterations are required to achieve a gap $\varepsilon < 10^{-10}$. As we argued in Section 3.7, in most cases, only very few iterations will be required to reach this level convergence. Further, note that a gap larger than $\varepsilon = 10^{-10}$ typically suffices in a practical TA setting. We would advise to dynamically tune the gap threshold to the network equilibrium gap present in

the current TA iterations. In other words, in earlier TA iterations where the (network level) gap is large, it is not necessary to aim for strict gap thresholds at each node, as it does little to contribute to improve the TA equilibrium at that point in time, saving computation time in the process.

Table 5. Equilibrium accepted outflows and generalised lane exit costs with $\varepsilon = 10^{-10}$

	j	1	2	3	4		1	2	3	4	
i	m	$\bar{s}_{mj}$				c_m	$\bar{v}_{mj}$				$\bar{h}_m$
1	1	0	0	0	600	1000	0	0	0	600	0.600
1	2	0	100	300	0	1000	0	100	300	0	0.400
2	3	200	0	0	837	1000	143	0	0	598	1.399
2	4	0	0	0	1037	1000	0	0	0	741	1.399
2	5	0	0	0	1037	1000	0	0	0	741	1.399
2	6	0	0	600	289	1000	0	0	429	207	1.399
3	7	200	200	0	400	1000	185	185	0	371	1.079
3	8	0	0	0	800	1000	0	0	0	741	1.079
4	9	0	0	800	0	1000	0	0	636	0	1.259
4	10	0	0	800	0	1000	0	0	636	0	1.259
4	11	0	900	0	0	1000	0	900	0	0	0.900
4	12	200	700	0	0	1000	200	700	0	0	0.900
	$r_j^{\bullet}$	2000	4000	2000	4000						

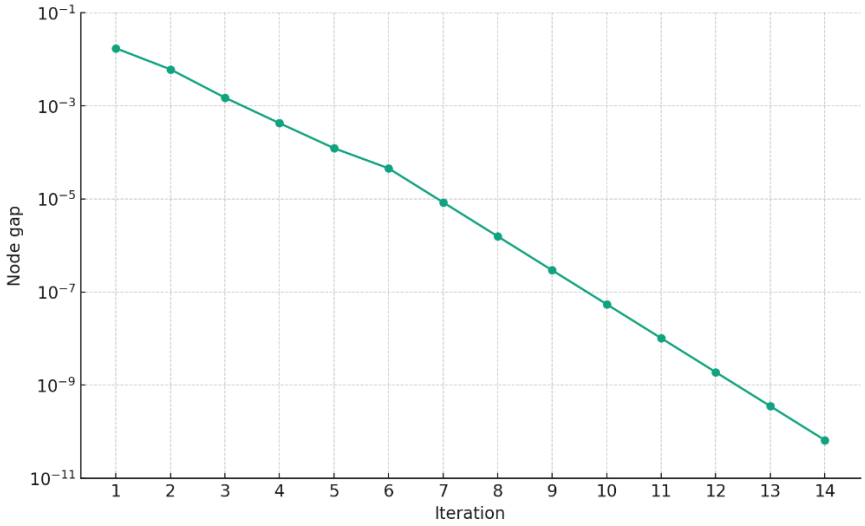


Fig. 10. Gap over iterations for case study (logarithmic vertical scale).

Table 6. The accepted outflow in link- and lane-based node model.

i	*j*	1	2	3	4	1	2	3	4
		outflow in *link-based* node model				outflow in *lane-based* node model			
1		0	100	300	600	0	100	300	600
2		137	0	411	2191	143	0	429	2288
3		200	200	0	1200	185	185	0	1112
4		161	1289	1289	0	200	1600	1271	0

Let us now compare the results to the link-based node model, with its implicit all-lanes-allow-all-movements approach. Table 6 presents the accepted outflows for both models aggregated at the link level. Generally, the presence of lane configurations allows drivers in the lane-based node model to avoid congested movements by switching to other lanes (when present and able). This is most noticeable for movement (4,2), here the link-based node model only allows 1289 veh/h, whereas the lane-based node model can accommodate 1600 veh/h. Practically, this means that the queue length and queuing delay in the link-based node model for movement (4,2) will be overestimated. The lane-based node model also lets more flow exit link 3, which competes for capacity with flow for movement (3,4) and therefore link 4 flows are reduced compared to the link-based node model. For uncongested movements, or situations where there are no options to switch to other approach lanes, results are identical or similar, as one would logically expect. These findings demonstrate distinct differences between the link- and lane-based node model, with the lane-based node model considering the approach lane configuration and accounting for the impact it has on lane choice, as well as relaxing the link-FIFO constraint in the process.

To ensure compatibility with the proposed node model, a link model is required to effectively manage movement-specific queues at the downstream link end and aggregate queues at the upstream link end, as illustrated in Fig. 11(a). However, the existing link model assumes link-based FIFO, providing identical link boundary for all movements, which results in the generation of link-level queues along the link, as schematically depicted in Fig. 11(c).

As shown in Fig. 11(b), the proposed node model adopts movement-based sending flows s_{ij}, while implicitly assuming link-based receiving flows r_i. This inconsistency between sending and receiving flows can lead to issues, particularly during spillback, where a queue from one movement may obstruct the entire link. This may be resolved by formulating a link model that determines movement-based receiving flows r_{ij} based on assumptions of queue propagation per movement, or by formulating a link model that splits the link into an aggregate queue component and a disaggregate queue component as illustrated in Fig. 11(a).

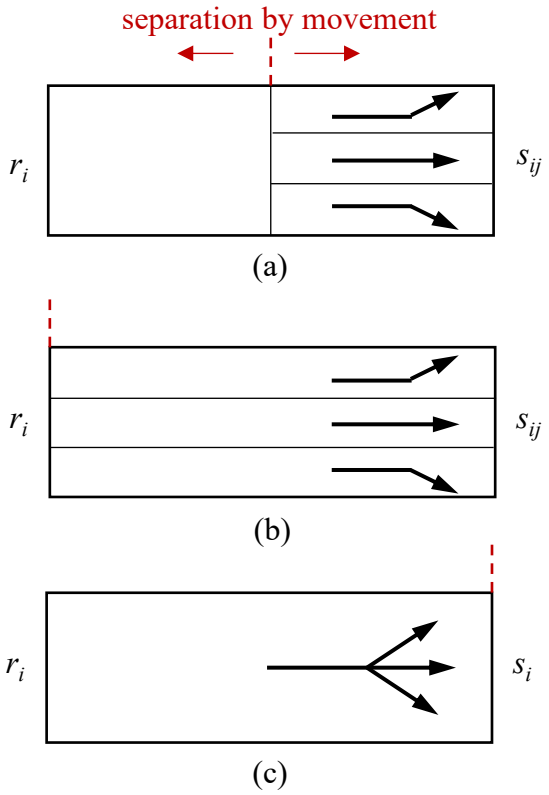


Fig. 11. (a) Desired (b) adopted (c) existing link model implications

3.9 Conclusions and further research

In this paper, a lane-based node model that explicitly considers the approach lane configuration is proposed. For intersections and junctions with multiple lanes on incoming links, it considers potentially separate queues per movement without the need to explicitly model separate approach lanes in the network. Rather, it considers a lane choice equilibrium in an implicit lane-expanded node representation, where vehicles are assigned to a lane when they are exiting the link, thereby relaxing link-based FIFO to movement-based FIFO. In our approach, no explicit approach lanes are created in the network, which avoids exponential growth of routes in the network. The lane choice equilibrium problem was formulated as a variational inequality problem. It was shown that the solution exists and is unique. An iterative solution algorithm to find this lane choice equilibrium was presented. A case study demonstrating general applicability of the algorithm and a comparison with an existing link-based node model are included.

The novel lane-based node model can be applied in existing macroscopic network loading models used in practice, e.g., OmniTRANS, VISUM, and AIMSUN. It provides governments, researchers, and consultants with an effective method that may more realistically describe intersection delays and flows at both traditional intersections as well as freeway-to-freeway or freeway-to-arterial connections. In addition, it may benefit macroscopic (static and dynamic) simulations on the impact of lane closures, lane changes, and other lane-specific phenomena on traffic flow that could not be accurately captured with existing node models.

The proposed node model, while solvable in an elegant equilibrium framework, presents certain limitations. First, the introduction of a lane-based approach brings new challenges related to the compatibility between the existing link model. This is primarily due to the imposition of distinct boundary conditions for each turning movement, an aspect that has not yet been adequately addressed within the current link model. The lane-grouping concept of the node model proposed by Yahyamozdarani (2024) avoids compatibility issues with the existing link model. However, it does not fully relax the link-based FIFO assumption, limiting its ability to capture traffic flow and delay, which remains a notable issue in the node model. Additionally, we adopted the conventional simplifying assumptions in most macroscopic node models, specifically excluding the explicit considerations of traffic signal control, priority rules, and interactions with other road users such as pedestrians. In future research, we aim to develop a lane-based link model that can be effectively integrated with the proposed lane-based node model by appropriately handling distinctive movement boundary conditions. Furthermore, we plan to expand the lane-based node model to incorporate more factors, e.g., by embedding the methodology of Yahyamozdarani and Tampère (2023) to account for traffic signal control. Additionally, we plan to explore the impact of the lane-based node model in a network context as well as by means of obtaining and comparing to empirical data.

4 Chapter 4 Node Model Incorporates Endogenous Exit Capacities

Existing node models often assume exogenous exit capacities, failing to account for variations in capacities with real-time flow dynamics. This chapter proposes a novel approach to incorporate endogenous exit capacities within the node model, accommodating varying movement exit speeds. To achieve this, saturation movement units are introduced to scale up inflows in smu/h during the pre-processing stage and scale back outflows in veh/h during the post-processing stage. Numerical examples validate the capability and effectiveness of the proposed method. This innovative method is the first to consider endogenous exit capacities in a node model by adjusting the flow rates endogenously according to various flow compositions, without requiring modifications to the node model itself. The proposed approach enhances the model ability to capture flow dynamics and improves predictions of persistent delays within the node model.

This chapter includes the following paper:
Gong, X., Bliemer, M.C.J. & Raadsen, M.P.H. (2024) Incorporation of endogenous exit capacities in node models to account for varying movement exit speeds. Submitted to *Transportmetrica A: Transport Science.*

Manuscript: Incorporation of Endogenous Exit Capacities in Node Models with Various Movement Exit Speeds

Abstract

In macroscopic traffic assignment, the node model is fundamental for accurately capturing traffic dynamics at junctions and intersections, which are often the primary sources of travel delays in urban networks. The precision of the node model is important for predicting delays and managing flows, which directly impacts traffic condition simulations, congestion management, and network performance optimisation. However, existing node models typically assume exogenous exit capacities for links, movement, or lanes feeding the node. They do so by either assuming fixed capacities or adjusting externally based on traffic control. This assumption often overlooks variations in exit capacities caused by flow compositions and vehicle turning behaviours, leading to less reliable predictions and reduced applicability in real-world scenarios. This study proposes a novel approach that can achieve endogenous exit capacities within the node model. Rather than dynamically adjusting exit capacities, our approach moderates the flow rates endogenously with flow compositions. This adjustment is made by scaling inflows for each movement inversely proportional to their turn speeds in the pre-processing stage and rescaling outflows in the post-processing stage without requiring modifications to the node model itself. Additionally, this study presents a solution algorithm consistent with our proposed method. Two numerical examples are provided to demonstrate the feasibility and capability of this approach. The results indicate that the node model that does not consider endogenous exit capacities may either overestimate or, in rare cases, underestimate the outflows.

Keywords: Macroscopic traffic assignment, node model, endogenous capacity, dead capacity, movement exit speed

4.1 Introduction

4.1.1 Background

As a key component of traffic assignment, the network loading (NL) model advances the resultant flows along the selected routes. Macroscopic models characterise traffic in terms of flow, whereas microscopic models focus on individual vehicles. Due to their computational efficiency in comparison with microscopic models, this paper focuses on macroscopic NL models with appealing equilibrium convergence properties for application to extensive networks. In macroscopic NL models, transport networks are depicted through *links* and *nodes*. Links characterise homogenous road segments, which are defined by attributes such as length, speed, and the number of lanes. Conversely, nodes represent motorway merges, diverges, or urban intersections, often featuring allowable turn movements with possibly traffic signal controls, and are often the primary source of travel delays in urban networks. Therefore, accurate node models are essential for predicting traffic conditions, optimising signal timing, and improving overall network efficiency.

Node models are integral not only to dynamic NL models, such as the cell transmission model (CTM) (Daganzo, 1994, 1995) and link transmission model (LTM) (Yperman, 2007; Gentile, 2010; Friesz et al., 2013; Han et al., 2016; Himpe et al., 2016; van der Gun et al., 2017; Bliemer & Raadsen, 2020), but also to some capacity-constrained static NL models (Bliemer et al., 2014; Bliemer & Raadsen, 2020; Raadsen & Bliemer, 2023). In the node model, the link exit capacity—the maximum rate at which vehicles can leave a node—can be considered exogenous or endogenous, depending on how it is determined and incorporated into the model. *Exogenous capacity* refers to when link exit capacity is determined independently of the current traffic conditions within the network. It is typically predefined and remains fixed or is adjusted based on external factors such as historical data, standard values, or predetermined rules (e.g., fixed signal timings, and regulatory constraints). The exit capacity can be represented as $c = f(\lambda)$, where λ may be movement-specific but remains time-independent. Conversely, *endogenous capacity* refers to the link exit capacity that is determined based on the current traffic conditions and the flow dynamics within the network. It varies with traffic dynamics, such as vehicle speed, queue lengths, and flow density, and reflects the actual capacity available at any given instant. The endogenous exit capacity that captures real-time traffic dynamics can be presented as $c(t) = f(\lambda(t))$, where $\lambda(t)$ is a dynamic parameter dependent on flow, it varies with the flow q over time instant t.

Endogenous capacity provides a more capable representation of real-time traffic conditions in network modelling, as it dynamically adjusts based on flow dynamics compared to the fixed nature of exogenous capacity. However, in conventional node models, link exit capacities are often considered fixed and are, if adjusted based on exogenous information, independent of the mix of movements in the sending flow on a link. When fixed, they assume exit capacities are determined solely by the physical characteristics of the road and intersection geometry, independent of the current traffic conditions within the network. This simplification can lead to inaccuracies in predicting traffic behaviour, especially under varying traffic loads and dynamic conditions. Since saturation flow rates are maximum rates that are only observed under ideal circumstances, for example, when vehicles drive straight ahead at the critical (saturation) speed without any interruption.

In real-world traffic scenarios, link exit capacities often vary due to the effects of several factors, such as downstream traffic conditions, traffic signal timings, and driver behaviour. A significant factor affecting exit capacities is the action of vehicles slowing down to make turns. In practice, maximum outflow rates are movement specific, as higher outflows are observed for through movements than for left or right turns due to reduced safe vehicle exit speeds for turning movements, where the actual vehicle exit speed depends on the angle of the movement. For example, Fitzpatrick et al. (2006) analysed the impact of a separate right-turn lane on right-turn movement speeds at 19 urban approaches. They found that the 85th-percentile vehicle exit speed in free flow near the middle of the right turn (20.9–33.8 km/h) is lower than that on the entire approach (27.4–46.7 km/h). The actual number of vehicles that can exit a link per hour depends on the speed with which vehicles exit the link, which is influenced by their movement. Therefore, link capacity reduces when vehicles decrease speed to make a turn. We refer to this loss of capacity as *dead capacity*, as it is the unusable part of saturation flow due to reduced safe vehicle exit speeds. In this paper, we focus on introducing a novel methodology to endogenously consider the modelling of dead capacity in combination with existing first-order general node models.

4.1.2 Illustrative example and approach

Consider the diverge node shown in Fig. 1 with a single incoming link $i \in \{1\}$, two outgoing links $j \in \{2,3\}$, and two movements, namely (1, 2) and (1, 3). If all vehicles on link 1 turn left, which involves reducing speed, the exit capacity would be lower than if all vehicles went straight ahead at a higher speed. Therefore, the exit capacity on link 1 will depend on the mix of vehicles turning left and going straight. To illustrate numerically, suppose that the saturation flow rate on link 1 is 1000 veh/h, which gives a saturation time headway of 3.6 seconds. Suppose that half of the vehicles drive straight at critical speed and have a headway of 3.6 seconds, while the other half of the vehicles turn left at lower speed and require 6 seconds of headway. In such a case, the exit capacity on link 1 is $3600/(\frac{1}{2}\cdot 3.6+\frac{1}{2}\cdot 6)=750$ veh/h. Here, a quarter of the capacity on link 1 was lost due to vehicles decelerating to make their turns.

Existing node models typically assume static link exit capacities, overlooking fluctuations attributable to changes in turning fractions. This research introduces a novel approach to changing this assumption by incorporating a scaling factor, denoted as λ_{ij}, termed the *saturation movement unit* (smu). The smu quantifies the relative impact on capacity caused by vehicles making a specific movement (*i, j*). This concept parallels the passenger car unit (pcu), which measures the relative influence of different vehicle types on capacity, as exemplified in the work of Petigny (1967). In the above example, each vehicle on link 1 to link 3 has a saturation time headway of 1/1000 h/veh. This equates to smu of 1, as these vehicles do not need to slow down to exit the link. For a vehicle moving from link 1 to link 2, we assume that the required time headway is $\lambda_{12}/1000$, where $\lambda_{12}=6/3.6=1.67$ for movement (1, 2) on link 1. Thus, a flow of 375 veh/h going straight and 375 veh/h turning left would mean $1\cdot 375+1.67\cdot 375=1000$ smu/h, which means that these 750 vehicles fully consume the link exit capacity. A movement-specific flow of 375 veh/h consumes $1.67\cdot 375=626$ smu/h of the available saturation flow 1000 veh/h on link 1. By expressing flow in smu/h, we avoid explicitly determining turn fraction-dependent link exit capacities. In our approach, only the inputs (inflows) and outputs (outflows) to the node model are scaled based on smu. In this study, we will demonstrate the benefit of not having to make any modifications to the underlying node model itself by using smu, making it preferable over existing methods and more attractive for practical applications.

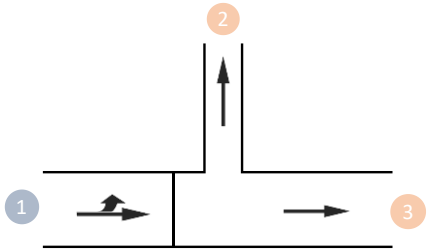


Fig. 1. Diverge node example with incoming/outgoing links and allowable movements.

4.1.3 Contribution and outline

In this study, we make the following principal contributions: (1) introducing the concept of smu based on movement exit speeds; (2) the incorporation of smu into the node model to achieve endogenous exit capacities without changing the node model itself; (3) the development of a solution algorithm that is consistent with the proposed approach; and (4) the introduction of numerical examples that demonstrate the viability and applicability of the proposed approach.

The structure of the remainder of the paper is as follows: Section 4.2 provides a literature review of the state-of-the-art NL models that consider exogenous or endogenous link exit capacities. The methodology of the prevalent existing node model is described in Section 4.3. The demonstration of the fundamental diagram and the modelling of the endogenous exit capacities are provided in Section 4.4. Section 4.3 presents an efficient solution algorithm for the proposed method. Section 4.6 validates the proposed method through two numerical examples. Finally, the conclusions and directions for future research are discussed in Section 4.7.

4.2 Literature review

In NL, link capacity is often considered exogenously, with the majority of NL models assuming constant link exit capacities (Tampère et al., 2011; Flötteröd & Rohde, 2011; Gibb, 2011; Himpe et al., 2016; Jabari, 2016; Raadsen & Bliemer, 2019; Bliemer & Raadsen, 2019; Yahyamozdarani & Tampère, 2023). Some research studies introduced variability through exogenous parameters such as rainfall and incident severity. Adverse weather conditions significantly impact vehicle turning speed at intersections by reducing visibility and altering driver behaviour. Asamer and Reinthaler (2010) estimated urban road capacity and free flow speed under adverse weather conditions using aggregated flow and speed data from local sensors. The results show a significant reduction in road capacity and free flow speed depending on intensity and type of precipitation. Xu et al. (2013) quantified the impact of rainfall on urban traffic using a fundamental diagram, showing noticeable decreases in traffic flow, speed, and network capacity. In these studies, while it is straightforward to change capacity based on various exogenous factors, these approaches neglect the variations in flow dynamics and the real-time adjustments of capacities, leading to less accurate traffic predictions.

Some studies focused on developing NL models that account for endogenous link capacity, thereby more accurately capturing flow dynamics. Several studies achieved this by

implementing dynamic lane management. Van der Gun et al. (2019) extended the LTM with the capabilities of non-empty initial conditions and computing within-link density profiles and validated the proposed model with a dynamic lane management simulation study. Levin and Boyles (2016a) proposed a CTM with a variable number of lanes in space and time consistent with the hydrodynamic theory of traffic flow to model dynamic lane reversal. Some studies advanced the understanding of endogenous link capacities, especially in mixed traffic involving autonomous vehicles (AVs) and human-driven vehicles (HVs). Ngoduy et al. (2021) proposed a two-region transmission model that takes into account the variations in link capacities in response to the dynamics of vehicle class proportions. Similarly, Zhang et al. (2023) adjusted link capacity by dynamically determining the AV share factor based on the stochastic arrival process of AVs and HVs and their interactions within the network. Similar research efforts on developing multiclass CTM and LTM that account for mixed traffic flows of AVs and HVs highlighted how the flow-density relationship varies with the endogenous proportions of AVs (Levin & Boyles, 2016b; Qin & Wang, 2019; Levin et al., 2019; Guo et al., 2021; Jin et al., 2022). Research has also focused on general multiclass NL models. Smits et al. (2011) proposed a multiclass LTM by normalising fundamental diagrams with respect to the capacity for each user class and maintaining consistent shockwave speeds. However, the simplified triangular shapes may not effectively capture the complexities and nuances of real-world traffic behaviours and subtle flow variations. Liu et al. (2015) developed a CTM for mixed flows of buses and passenger cars, dynamically adjusting link capacity by integrating buses as moving bottlenecks, but they do not essentially work on exit capacities. Levin and Kang (2023) developed a multiclass LTM that incorporates a spatially and temporally varying flow-density relationship with the movement of vehicles, which indirectly affects exit capacities.

Numerous studies on macroscopic NL achieved endogenous dynamic capacity by considering variable speed limits (VSLs) on links. Hajiahmadi et al. (2013) extended the LTM to include VSLs, manipulating travel times and delays to adapt to changing traffic conditions. Raadsen and Bliemer (2021) introduced a continuous-time LTM that incorporates VSLs and uses an information propagation method. This approach allows the model to dynamically adjust speed limits on a link, directly influencing the link's capacity by altering flow rates and travel times based on real-time traffic conditions. Liu et al. (2023) integrated VSL and dynamic lane allocations within the CTM, enabling the model to adaptively adjust link capacity based on real-time conditions and traffic demands. Wei et al. (2023) developed an extended LTM that generates VSLs and incorporates turn-level queued transmission to optimise traffic flow at signalised intersections. While effective in generating dynamic link capacities and traffic flow propagation, this model is considered complex for practical applications. Table 1 summarises the research that considers endogenous exit capacities along with the associated characteristics.

Table 1. Research of endogenous exit capacities in network loading.

Endogenous characteristics		Research
Dynamic lane management		Levin & Boyles (2016a), Van Der Gun et al. (2019)
Multiclass	AVs	Levin & Boyles (2016b), Levin et al., (2019), Qin & Wang (2019), Ngoduy et al. (2021), Guo et al. (2021), Jin et al. (2022), Zhang et al. (2023)
	General multiclass	Bliemer (2007), Smits et al. (2011), Levin & Kang (2023)
VSL		Hajiahmadi et al. (2013), Raadsen & Bliemer (2021), Liu et al. (2023), Wei et al. (2023)

Despite advancements in NL models that incorporate endogenous exit capacity, several limitations have persisted, including complexity, high data requirements, and computational intensity. These challenges hinder the feasibility of implementing these models in real-world traffic management systems, particularly in less technologically advanced settings or for real-time operations. To accurately model travel time and travel delay at intersections, this research proposes incorporating endogenous exit capacity within the node model of the NL framework. Most existing node models assume homogeneous flow, typically represented by a single vehicle class and uniform link capacity. However, Bliemer (2007) formulated a multiclass node model that achieves time-varying link exit capacities while making changes to the node model itself and does not maximise flow within each class. No existing node model incorporates endogenous exit capacities based on vehicle turn speeds. To the best of our knowledge, the novel method proposed in the subsequent sections is the first to effectively account for endogenous link capacities based on flow compositions and vehicle turning behaviours. This research focuses on unsignalised junctions and intersections, as traffic signals are the dominant factor in reducing link exit capacity, and therefore, this was deemed of less interest to explore further.

4.3 Conventional node model

The first-order node models proposed by Daganzo (1994, 1995) are among the pioneering macroscopic node models. Building on these foundational models, numerous advanced node models have been developed, incorporating sophisticated merging and diverging schemes and accommodating multiple incoming and outgoing links (Lebacque & Khoshyaran, 2005; Jin, 2010; Flötteröd & Rohde, 2011; Gibb, 2011; Tampère et al., 2011; Jabari, 2016; Wright et al., 2017; Yahyamozdarani & Tampère, 2023; Gong et al., 2024). Tampère et al. (2011) proposed seven requirements of a generic node model: (i) *general applicability*; (ii) *flow maximisation*; (iii) *non-negativity of flows*; *(iv) conservation of vehicles*; (v) *satisfaction of demand and supply constraints*; (vi) *conservation of turning fractions*; and (vii) *satisfaction of the invariance principle*. Node models that satisfy all these requirements are the *generic class of first-order macroscopic node models* (GCNM).

Let $G=(N,A)$ be a directed graph representing a road network, where N is the set of nodes and A is the set of directed links. Each node $n \in N$ has incoming links $i=1,\ldots,I$ and outgoing links $j=1,\ldots,J$, $i,j \subset A$. Let (i,j) denote the movement from incoming link i towards

outgoing link *j*. To ensure the simplicity of notation, we will omit subscript *n* from all variables from now on. The conventional node model considered here does not consider individual lanes; consequently, all variables are either link-based or movement-based. Further, as defined in the general class of macroscopic node models in Tampère et al. (2011), node models consider the following inpus: (i) the matrix of movement-specific sending flows, $\mathbf{s} = [s_{ij}]$, (ii) the vector of link receiving flows, $\mathbf{r} = [r_j]$, and (iii) the vector of link saturation flows $\mathbf{Q} = [Q_i]$ to determine merge priorities. The node model output is given by movement-specific outflow rates $\mathbf{v} = [v_{ij}]$ via an implicit function $\Gamma(\cdot)$ that describes the node model,

$$\mathbf{v} = \Gamma(\mathbf{s}, \mathbf{r}, \mathbf{Q}). \tag{1}$$

Define total link sending flows $s_i = \sum_j s_{ij}$ and total link outflow rates $v_i = \sum_j v_{ij}$ for $\forall i = 1, \ldots, I$. The node model $\Gamma(\cdot)$ implicitly determines a set of supply-constrained (congested) links. When $v_i < s_i$, this indicates a queue build up at the end of link *i*. Conversely, when $v_i = s_i$, the link *i* is demand-constrained, where there is no queue on the link (i.e., in free-flow).

In this paper we adopt the node model proposed by Tampère et al. (2011), while acknowledging that other node models that conform to Eq. (1) are also viable.

4.4 Endogenous exit capacities in node model

In this section, we present a novel approach to incorporating endogenous exit capacities into node models. First, we introduce the general fundamental diagram that characterises the interaction of flows at link boundaries and present the specific shape of the fundamental diagram adopted in this study. Then, we discuss how link exit capacities can be accounted for endogenously in node models by scaling the inflows and outflows based on the movement-specific smu.

While this research does not directly investigate flow propagation along a link, the inclusion of a fundamental diagram (FD) is important for calculating the smu. The FD depicts the relationship among traffic flow *q* (veh), density *k* (veh/km), and speed σ (km/h). The flow and density can be empirically observed from traffic counts and are influenced by factors such as the number of lanes, maximum speed limit, and road type. The speed is determined using the fundamental relationship, $\sigma = q / k$. An FD applies to each cross-section of a homogeneous road segment. Traffic at a cross-section on a link can exist in one of two states (Cascetta, 2009). It is in a *hypocritical* (uncongested) state when traffic is demand-constrained, with flow increasing as density increases, indicating no congestion and no queues. Conversely, it is in a *hypercritical* (congested) state when traffic is supply-constrained, with flow decreasing as density increases, indicating the presence of congestion and queues. A general FD can be represented by function $q = \Phi(k)$, also known as the flux function. We therefore define a two-regime function of FD as follows:

$$\Phi(k) = \begin{cases} \Phi_{\mathrm{I}}(k), & \text{if } 0 \le k \le k^{\mathrm{crit}}, \\ \Phi_{\mathrm{II}}(k), & \text{if } k^{\mathrm{crit}} \le k \le k^{\mathrm{jam}}, \end{cases} \tag{2}$$

where k^{jam} (veh/km) is jam density, $\Phi_{\text{I}}(k)$ present function of FD in hypocritical situation, $\Phi_{\text{II}}(k)$ denote function of FD in hypercritical situation. When the density is lower than the critical density k^{crit} (veh/km), traffic is in a hypocritical state, otherwise traffic transitions to a hypercritical state.

Various shapes of an FD exist for the purposes of computational efficiency, realism, and ease of calibration (Bliemer et al., 2017). The FD in this study can adopt any shape of a hypocritical branch, as the proposed method considers only the critical speed and the back wave speed and excludes the free flow speed. For simplicity, we assume a linear representation in the hypercritical branch.

Consider each link *i* is a homogenous road segment with the following attributes: movement saturation flow Q_{ij} (veh/h), movement exit capacity as c_{ij} (veh/h), critical density k_i^{crit} (veh/km), jam density k_i^{jam} (veh/km), backward wave speed w_i (km/h), movement exit speed σ_{ij} (km/h). When vehicles travel at a speed higher than the critical speed σ_i^{crit} (km/h), a hypocritical condition prevails. In this state, no congestion or queues are present on the link, allowing the capacity at maximum flow (saturation flow) are available to traffic flow. In this case we have

$$c_{ij} = Q_{ij}, \text{ if } \sigma_{ij} \geq \sigma_i^{\text{crit}}, \ \forall i = 1, \ldots, I, \ j = 1, \ldots, J. \tag{3}$$

When vehicles travel at a speed lower than the critical speed, traffic transitions to a hypercritical condition. In this state, congestion and queues build up on the link, consuming the capacity that can be utilised by the traffic flow, leading to a drop in capacity. Fig. 2 presents the hypercritical linear branch of the FD (red line), from which we have

$$k_i^{\text{jam}} - k_i^{\text{crit}} = \frac{Q_{ij}}{-w_i} \tag{4}$$

$$k_i^{\text{crit}} = \frac{Q_{ij}}{\sigma_i^{\text{crit}}} \tag{5}$$

Substituting Eq. (4) with (5) we get

$$k_i^{\text{jam}} = \frac{Q_{ij}(w_i - \sigma_i^{\text{crit}})}{\sigma_i^{\text{crit}} w_i} \tag{6}$$

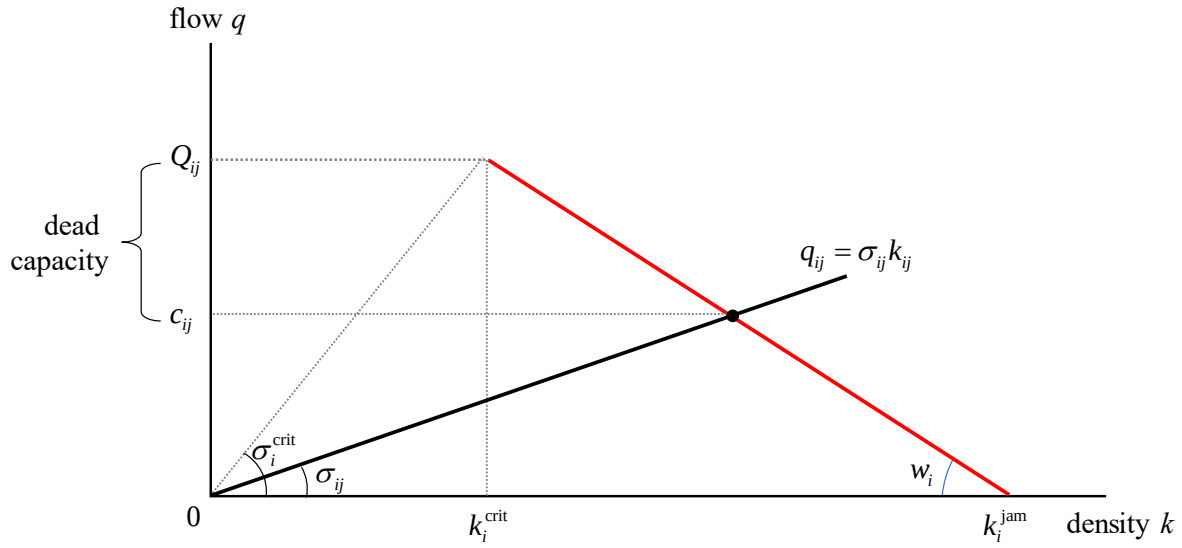


Fig. 2. Movement exit capacity depending on movement exit speed.

As shown in Fig. 2, the functional form of the FD within the hypercritical regime exhibits linearity. Specifically, it is characterised by a linear relationship with a slope w_i over the domain k_i^{crit} to k_i^{jam}. Within this domain, at critical density k_i^{crit} (refer to Eq. (5)) the flow achieves its maximum, denoted as $q_{ij} = Q_{ij}$. At jam density k_i^{jam} (refer to Eq. (6)) the flow reaches its minimum, indicated by $q_{ij} = 0$. Therefore, the formulation of the FD in the hypercritical branch can be succinctly constructed as in Eq. (7).

$$\Phi_{\text{II}}(k_{ij}) = w_i k_{ij} + \left(\frac{\sigma_i^{\text{crit}} - w_i}{\sigma_i^{\text{crit}}} \right) Q_{ij}, \text{ if } \sigma_{ij} \le \sigma_i^{\text{crit}}, \ \forall i = 1, \ldots, I, \ j = 1, \ldots, J. \tag{7}$$

Movement density is the ratio of movement flow rates q_{ij} and movement exit speed σ_{ij}, which is $k_{ij} = \Phi_{\text{II}}(k_{ij}) / \sigma_{ij}$. Substituting k_{ij} in Eq. (7) we get the FD in a function of movement exit speed as

$$\Phi_{\text{II}}(\sigma_{ij}) = \left(\frac{\sigma_i^{\text{crit}} - w_i}{\sigma_{ij} - w_i} \right) \left(\frac{\sigma_{ij}}{\sigma_i^{\text{crit}}} \right) Q_{ij}, \text{ if } \sigma_{ij} \le \sigma_i^{\text{crit}}, \ \forall i = 1, \ldots, I, \ j = 1, \ldots, J. \tag{8}$$

Eq. (8) also represents the maximum flow that can exit per movement, which corresponds to the movement exit capacity, as a function of movement exit speeds,

$$c_{ij}(\sigma_{ij}) = \left(\frac{\sigma_i^{\text{crit}} - w_i}{\sigma_{ij} - w_i} \right) \left(\frac{\sigma_{ij}}{\sigma_i^{c}} \right) Q_{ij}, \text{ if } \sigma_{ij} \le \sigma_i^{\text{crit}}, \ \forall i \in I, \ j \in J. \tag{9}$$

From Eqs. (3) and (9), the movement exit capacity can be formulated as the two-regime function of movement exit speed in both hypocritical and hypercritical situations,

$$c_{ij}(\sigma_{ij}) = \begin{cases} \left(\dfrac{\sigma_i^{\text{crit}} - w_i}{\sigma_{ij} - w_i}\right)\left(\dfrac{\sigma_{ij}}{\sigma_i^{\text{crit}}}\right) Q_{ij}, & \text{if } \sigma_{ij} \le \sigma_i^{\text{crit}}, \\ Q_{ij}, & \text{otherwise.} \end{cases} \tag{10}$$

Eq. (10) presents that if the vehicle exit speed exceeds the critical speed, then movement saturation flow Q_{ij} can be achieved. However, if vehicle exit speeds are lower than the critical speed, then the movement exit capacity is reduced. To illustrate, for a specific movement (*i, j*) assume that $Q_{ij} = 500$ veh/h, $\sigma_i^{\text{crit}} = 40$ km/h, $w_i = -12$ km/h and suppose that the turning vehicle exit speed σ_{ij} is 20 km/h. Then the link exit capacity is $c_{ij} = 406$ veh/h according to Eq. (10). The dead capacity $Q_{ij} - c_{ij} = 94$ veh/h is the unusable part of saturation flow due to reduced safe vehicle exit speeds, in contrast to the usable part c_{ij}.

When applied to a node model, for an inlink with multiple movements, it is necessary to calculate a weighted average link exit capacity based on the sending flow for each movement (i, j). Instead of computing such a weighted average link exit capacity each time the node model is run, we propose scaling the sending flows to account for differences in dead capacity for each movement. This idea is similar to scaling factors used in traffic assignment to account for the different impacts that different vehicle types have on capacity (see Petigny, 1967, who proposed the use of passenger car units to convert, for example, truck flow into car flow). We define a saturation movement unit (smu) λ_{ij} that expresses the relative difference between the headway required for movement (i, j) to exit a link and movement saturation headway. With the smu λ_{ij}, movement sending flow s_{ij} in unit of veh/h is converted to unit of smu/h to reflect the capacity it occupies. The expression of smu can be presented as

$$\lambda_{ij}(\sigma_{ij}) = Q_{ij} / c_{ij}(\sigma_{ij}) \tag{11}$$

Based on Eqs. (10) and (11) we get the two-regime function of saturation movement unit as

$$\lambda_{ij}(\sigma_{ij}) = \begin{cases} \left(\dfrac{\sigma_i^{\text{crit}}}{\sigma_{ij}}\right)\left(\dfrac{\sigma_{ij} - w_i}{\sigma_i^{\text{crit}} - w_i}\right), & \text{if } \sigma_{ij} < \sigma_i^{\text{crit}}, \\ 1, & \text{otherwise.} \end{cases} \tag{12}$$

Note that λ_{ij} is always in the interval $[1, \infty)$. A larger λ_{ij} implies that the flow for movement (*i, j*) occupies a larger portion of the available capacity compared to other flows with critical speed or higher.

Using smu, accounting for various movement exit capacities becomes straightforward in the node model. A movement-specific flow of s_{ij} veh/h consumes $\lambda_{ij} s_{ij}$ smu/h of the available saturation flow. One additional important aspect to consider is that the sending flow in smu/h on link *i* should be constrained by the link saturation flow Q_i. Thus, we get the movement-specific sending flow in smu/h as

$$\tilde{s}_{ij} = \min(\lambda_{ij} s_{ij}, \frac{\lambda_{ij} s_{ij} Q_i}{\sum_j \lambda_{ij} s_{ij}}) \qquad \forall i = 1, \ldots, I, \; j = 1, \ldots, J. \tag{13}$$

This minimisation is more than a simple scaling, it directly accounts for the impeding of turning vehicles on other movements in case they cause the total flow to exceed the space on the incoming link. In this sense, it can be seen as an implicit diverge node accounting for this effect and ensuring the First-In-First-Out (FIFO) principle in the process.

By expressing flow in smu/h, we avoid explicitly determining turn fraction-dependent link exit capacities. The implicit function $\Gamma(\cdot)$ that describes the node model in Eq. (1) is therefore replaced by,

$$\tilde{\mathbf{v}} = \Gamma\left(\tilde{\mathbf{s}}, \mathbf{r}, \mathbf{c}\right). \tag{14}$$

where $\tilde{\mathbf{s}}=[\tilde{s}_{ij}]$ is a matrix of movement-specific sending flow in smu/h, $\tilde{\mathbf{v}}=[\tilde{v}_{ij}]$ is a matrix of movement-specific outflow in smu/h.

Note that $\tilde{v}_{ij}$ needs to be converted from smu/h to flow/h to fit the reduced capacity after being obtained from the node model. Denote $\mathbf{v}^*=[v_{ij}^*]$ as a matrix of corrected movement-specific outflows. The corrected movement outflow v_{ij}^* in veh/h is obtained via Eq. (15).

$$v_{ij}^* = \tilde{v}_{ij} / \lambda_{ij}, \qquad \forall i = 1, \ldots, I,\ j = 1, \ldots, J. \tag{15}$$

4.5 Algorithm

In this section, we introduce an algorithm for solving unsignalised intersections considering endogenous link exit capacities based on the methodology introduced in Section 4.4. Fig. 3 provides a schematic interpretation of how the main model variables relate to the algorithm structure.

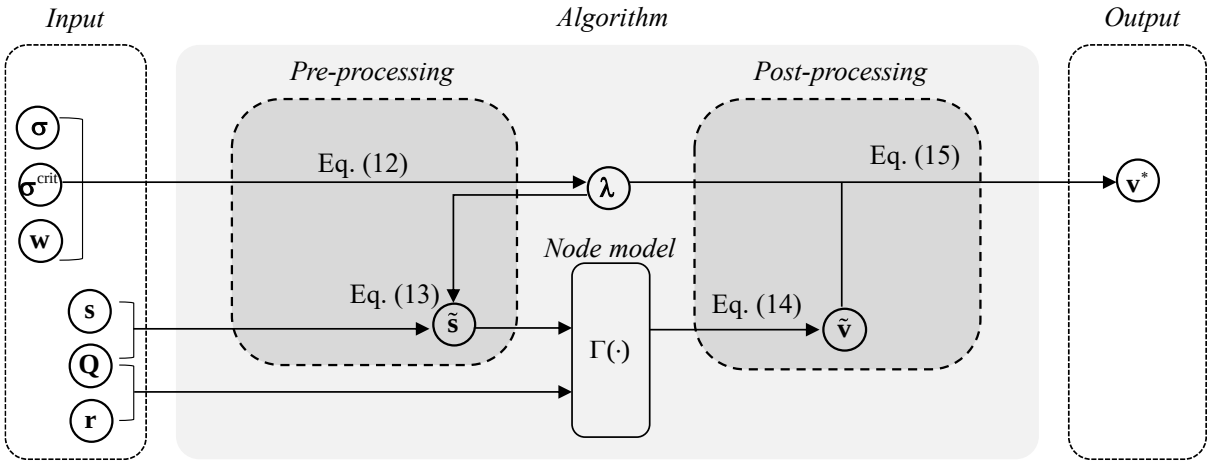


Fig. 3. Schematic diagram of variable relationship in solution algorithm

The solution scheme for solving the node model with endogenous link exit capacities consists of three steps:

Step 1: *Pre-processing.* Sending flows in veh/h are converted to flows expressed in smu/h that account for movement exit speeds.

Step 2: *Node model*. The outflow rates in smu/h are calculated through node model function $\Gamma(\cdot)$,

Step 3: *Post-processing.* The outflows in smu/h are converted to flow/h, which are the corrected outflows.

The full solution algorithm is given in *Algorithm* below.

Algorithm - solution scheme.

Input: Movement exit speed $\boldsymbol{\sigma} = [\sigma_{ij}]$, link critical speed $\boldsymbol{\sigma}^{\text{crit}} = [\sigma_i^{\text{crit}}]$, link back wave speed $\mathbf{w} = [w_i]$, movement sending flows $\mathbf{s}$, link saturation flows $\mathbf{Q}$, link receiving flows $\mathbf{r}$,

Output: Corrected movement-specific link outflows $\mathbf{v}^*$.

Step 1: *Pre-processing*

For each $i = 1, \ldots, I$:

First, compute the smu λ_{ij} using Eq. (12):

$$\lambda_{ij}(\sigma_{ij}) = \begin{cases} \left(\dfrac{\sigma_i^{\text{crit}}}{\sigma_{ij}}\right)\left(\dfrac{\sigma_{ij} - w_i}{\sigma_i^{\text{crit}} - w_i}\right), & \text{if } \sigma_{ij} < \sigma_i^{\text{crit}}, \\ 1, & \text{otherwise.} \end{cases}$$

Second, compute movement-specific sending flows in smu/h, $\tilde{s}_{ij}$, using Eq. (13):

$$\tilde{s}_{ij} = \min(\lambda_{ij} s_{ij}, \frac{\lambda_{ij} s_{ij} Q_i}{\sum_j \lambda_{ij} s_{ij}}) \qquad \forall i = 1, \ldots, I,\ j = 1, \ldots, J.$$

Then we get the matrix of movement sending flows in smu/h, $\tilde{\mathbf{s}} = [\tilde{s}_{ij}]$.

Step 2: *Node model*

Bring the sending flows in smu/h, $\tilde{\mathbf{s}} = [\tilde{s}_{ij}]$, as input into node model function in Eq. (14) to calculate the outflows in smu/h, $\tilde{\mathbf{v}}$.

$$\tilde{\mathbf{v}} = \Gamma(\tilde{\mathbf{s}}, \mathbf{r}, \mathbf{Q}).$$

Step 3: *Post-processing*

Compute corrected movement-specific outflows, v_{ij}^{*}, using Eq. (15).

$$v_{ij}^{*} = \tilde{v}_{ij} / \lambda_{ij}, \qquad \forall i = 1, \ldots, I,\ j = 1, \ldots, J.$$

A numerical example demonstrating the algorithm for solving the node model with endogenous link exit capacities is presented in Section 4.6.1. Note that while the proposed method is applicable to any form of node model, in the following section, we utilise the well-known generic node model proposed by Tampère et al. (2011).

4.6 Numerical examples

This section presents two numerical examples. Example 1 illustrates the computational procedure of the solution algorithm and validates the capability of the proposed method by applying it to the numerical example presented by Tampère et al. (2011). We compare the node model outcomes of with endogenous exit capacity to the outcomes of Tampère's node model assumes exogenous capacity. One might expect that outflows are always constrained when considering endogenous exit capacities. However, Example 2 demonstrates that this does not always hold true, as it presents cases where flows can both decrease and increase with various endogenous exit capacities.

4.6.1 Example 1 – model capability

The example's link saturation flows, sending and receiving flows are given as inputs as per Table 2. Assuming critical speed σ_i^{crit} is 50 km/h and back wave speed w_i is -13 km/h for each link. The summary of exit σ_{ij} of each turn movement is given in Table 3, where we assume through traffic move at a critical speed, and vehicles make turns are at speed of 20 km/h.

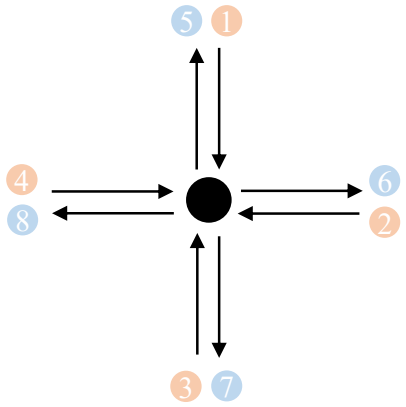


Fig. 4. Intersection example from Tampère et al. (2011)

Table 2. Node model inputs of Tampère et al. (2011)

i \ j	5	6	7	8	Q_i
	s_{ij} (veh/h)				(veh/h)
1	0	50	150	300	1000
2	100	0	300	1600	2000
3	100	100	0	600	1000
4	100	800	800	0	2000
r_j (veh/h)	1000	2000	1000	2000	

Table 3. Movement exit speed.

i \ j	5	6	7	8
	σ_{ij} (km/h)			
1	-	20	50	20
2	20	-	20	50
3	50	20	-	20
4	20	50	20	-

The calculation procedure, following the algorithm described in Section 4.5, is detailed below.

Step 1: *Pre-processing*

We show the calculation on link 3 which triggers the minimisation in Eq. (13). The calculations for links 1, 2, and 4 should follow the same step, but to avoid redundant repetition, we omit these steps in the following calculation.

For $i := 3$,

Compute scale factors using Eq. (12):

$\sigma_{35} = \sigma_3^{\text{crit}}$, $\lambda_{35} = 1.000$;

$\sigma_{36} = 20$ km/h $< \sigma_3^{\text{c}}$, $\lambda_{36} = \left(\frac{\sigma_3^{\text{crit}}}{\sigma_{36}}\right)\left(\frac{\sigma_{36} - w_3}{\sigma_3^{\text{crit}} - w_3}\right) = \left(\frac{50}{20}\right)\left(\frac{20+13}{50+13}\right) = 1.310$;

$\sigma_{38} = 20$ km/h $< \sigma_3^{\text{c}}$, $\lambda_{38} = \left(\frac{\sigma_3^{\text{crit}}}{\sigma_{38}}\right)\left(\frac{\sigma_{38} - w_3}{\sigma_3^{\text{crit}} - w_3}\right) = \left(\frac{50}{20}\right)\left(\frac{20+13}{50+13}\right) = 1.310$;

Compute scaled directional link-based sending flows, $\tilde{s}$, using Eq. (13):

$$\tilde{s}_{35} = \min\left(\lambda_{35} s_{35}, \frac{\lambda_{35} s_{35} Q_3}{\lambda_{35} s_{35} + \lambda_{36} s_{36} + \lambda_{38} s_{38}}\right) = \min(100, 98) = 98 \text{ smu/h};$$

$$\tilde{s}_{36} = \min(\lambda_{36}s_{36}, \frac{\lambda_{36}s_{36}Q_3}{\lambda_{35}s_{35} + \lambda_{36}s_{36} + \lambda_{38}s_{38}}) = \min(131, 129) = 129 \text{ smu/h};$$

$$\tilde{s}_{38} = \min(\lambda_{38}s_{38}, \frac{\lambda_{38}s_{38}Q_3}{\lambda_{35}s_{35} + \lambda_{36}s_{36} + \lambda_{38}s_{38}}) = \min(786, 773) = 773 \text{ smu/h}$$

The calculated smu and scaled-up movement-specific inflows are summarised in Table 4.

Table 4. The saturation movement units and the movement-specific inflow in smu/h.

i	j	5	6	7	8	5	6	7	8
		λ_{ij}				$\tilde{s}_{ij}$ (smu/h)			
1		-	1.310	1.000	1.310	-	66	150	393
2*		1.310	-	1.310	1.000	123	-	370	1507
3*		1.000	1.310	-	1.310	98	129	-	773
4		1.310	1.000	1.310	-	131	800	1048	-

* link that trigger sending flows exceeding capacity after scaling up

Step 2: *Node model*

Compute outflows $\tilde{v}_{ij}$ in smu/h using Eq. (14). The results are shown in Table 5.

Table 5. The accepted outflow in smu/h.

i	j	5	6	7	8
		outflow $\tilde{v}_{ij}$ (smu/h)			
1		0	65	147	385
2		73	0	221	899
3		91	119	0	715
4		79	483	632	0

Step 3: *Post-processing*

Compute corrected movement-specific outflows, v_{ij}^*, using Eq. (15).

$v_{16}^* = \tilde{v}_{16} / \lambda_{16} = 65/1.310 = 50$ veh/h, $v_{17}^* = \tilde{v}_{17} / \lambda_{17} = 147/1.000 = 147$ veh/h,

$v_{18}^* = \tilde{v}_{18} / \lambda_{18} = 385/1.310 = 294$ veh/h, $v_{25}^* = \tilde{v}_{25} / \lambda_{25} = 73/1.310 = 56$ veh/h,

$v_{27}^* = \tilde{v}_{27} / \lambda_{27} = 221/1.310 = 169$ veh/h, $v_{28}^* = \tilde{v}_{28} / \lambda_{28} = 899/1.000 = 899$ veh/h,

$v_{35}^* = \tilde{v}_{35} / \lambda_{35} = 91/1.000 = 91$ veh/h, $v_{36}^* = \tilde{v}_{36} / \lambda_{36} = 119/1.310 = 91$ veh/h,

$v_{38}^* = \tilde{v}_{38} / \lambda_{38} = 715/1.310 = 546$ veh/h, $v_{45}^* = \tilde{v}_{45} / \lambda_{45} = 79/1.310 = 60$ veh/h,

$v_{46}^* = \tilde{v}_{46} / \lambda_{46} = 483/1.000 = 483$ veh/h, $v_{47}^* = \tilde{v}_{47} / \lambda_{47} = 632/1.310 = 483$ veh/h,

Table 6. The accepted outflow from Tampere's case and proposed approach.

i	j	5	6	7	8	5	6	7	8
		v_{ij} following Tampère's node model (veh/h)				corrected outflow v_{ij}^{*} (veh/h)			
1		0	50	150	300	0	49	147	294
2		69	0	206	1096	56	0	169	899
3		100	100	0	600	91	91	0	546
4		81	645	645	0	60	483	483	0

Table 6 presents the accepted outflow as computed within the framework of the node model, considering fixed capacities (Tampère et al., 2011) and varying movement exit capacities, respectively. Upon accounting for the varying capacities, it is observed that the total number of vehicles that can flow out using the proposed method is 3368 veh/h, representing a 16.67% decrease compared to the outflow considered with identical capacities in the original link-based Tampère's node model (4042 veh/h). Notable reductions in flow are also observed in all links. In Tampère's results both inlinks 1 and 3 are demand constrained, while when considering various movement exit capacities, all four links become supply constrained. In addition, outlinks 7 and 8 become more restricted after claiming additional turning flows. Consequently, all movements toward outlinks 7 and 8 with lower speeds experience more severe constraints and receive a less equitable share of the flow. Furthermore, links 2 and 3 claim more flow after converting flow in veh/h to smu/h, which are then both restricted by link saturation flow. All above reasons result in overall flow reductions compared to Tampère's results. This indicates that node models ignore exit capacity reduction due to low turn movement speeds and may overestimate outflow.

Although in this example the outflows for all movements are higher compared to the same situation while considering endogenous exit link capacities, this is not the case in all scenarios. An example demonstrating node models that ignore exit capacity may not only overestimate but also underestimate outflows is discussed in Section 4.6.2.

4.6.2 Example 2 – scenario of flow underestimation

This example adopts a three-legged intersection as shown in Fig. 5. The link saturation flow, sending and receiving flows are given as inputs as per Table 7. Assuming a critical speed σ_i^{crit} of 50 km/h and a back wave speed w_i of -10 km/h for each link. According to Eq. (12), the saturation flow units λ_{ij} are determined based on the given exit speed σ_{ij} on each movement, both are summarised in Table 8.

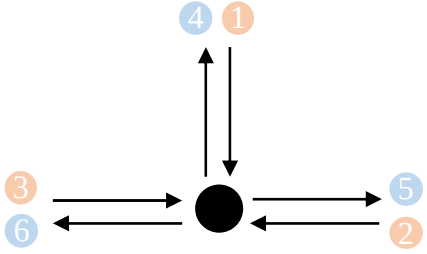


Fig. 5. Three-legged intersection.

Table 7. Node model inputs for case study 2.

i	j	4	5	6	Q_i
		s_{ij}			
1		0	200	0	1000
2		400	0	500	1000
3		400	500	0	1000
r_j		500	500	1000	

Table 8. Exit movement speed and saturation flow unit.

i	j	4	5	6	4	5	6
		σ_{ij} (km/h)			smu λ_{ij}		
1		-	15	-	-	1.389	-
2		30	-	50	1.111	-	1.000
3		30	50	-	1.111	1.000	-

Table 9. The outflow from Tampere's node model and the corrected accepted outflow.

i	j	4	5	6	4	5	6
		v_{ij} following Tampère's node model (veh/h)			corrected outflow v_{ij}^* (veh/h)		
1		0	200	0	0	200	0
2		260	0	325	272	0	340
3		240	300	0	177	222	0

The outflows from the node model, considering both fixed and endogenous exit capacities, are detailed in Table 9. It is observed that the flow on link 2 increases upon incorporating endogenous exit capacities. This adjustment arises because congestion develops on link 5 when more flows are claimed on turn movement (1, 5). Consequently, movements (3, 4) and (3, 5) experience more constraints. This redistribution results in link 2 receiving a larger portion of flow destined for links 4 and 6, leading to respective increases of 12 veh/h and 15 veh/h in flows for movements (2, 4) and (2, 6). This case illustrates that neglecting reductions in exit capacity due to slower turning movement speeds in node models can lead to both overestimations and underestimations of outflows.

4.7 Conclusions

This paper proposes a novel approach of including endogenous exit capacities within macroscopic node models by considering movement exit speeds, thereby enhancing the model's capability to reflect real-world traffic conditions. We introduced smu to the model, which facilitates obtaining node outflows that consider the impact of reduced turn movement speeds. This is achieved by scaling the inflow rate during the pre-processing stage and rescaling the outflows in the post-processing stage of a conventional node model. In addition, a solution algorithm is proposed to solve the proposed method. Two numerical examples are provided to validate the viability and effectiveness of the proposed method. They reveal that ignoring the

impact of turn movement speeds on exit link capacity may lead to both overestimation and underestimation of outflows on links.

Instead of dynamically changing link capacities, which may lead to complications in the underlying link model, or at the very least make them significantly more complex to solve, the proposed method scales the in and outputs of the node model. This has the benefit that no changes are required on either the node model, nor the link model. The proposed method is demonstrated to be capable in capturing the impact of flow dynamics for various turning movements, which may assist in providing more accurate traffic flow and congestion predictions which are important for governments and consultants.

This study primarily addresses homogeneous traffic conditions. In future research, we will focus on augmenting the model to incorporate multiclass vehicle dynamics, possibly by converting multiclass flow into a single class unit, such as a passenger car unit. Additionally, we plan to explore the incorporation of lane-specific exit capacities alongside the lane-based node model proposed by Gong et al. (2024). Further investigation is required to confirm the feasibility of this approach. Furthermore, we aim to examine the implications of the proposed method within a broader network framework. This will involve collecting and analysing empirical data, thereby providing robust validation and a deeper understanding of the model's effectiveness in simulating real-world traffic scenarios.

5 Chapter 5 Non-persistent Delays in LTM

Existing macroscopic dynamic network loading models typically ignore non-persistent delays. This chapter introduces a novel method that directly integrates non-persistent delays into the link transmission model (LTM). To achieve this, a virtual link is implicitly incorporated into the link model representation for each link- or movement-specific traffic signal control, where Webster's uniform delay is reformulated as a hypothetical branch of the associated fundamental diagram. The proposed method is the first rigorous approach to incorporate non-persistent delay in LTM by averaging the effects of traffic control without violating the FIFO principle. Existing time-discretised solution algorithms can be adapted to solve the augmented LTM. Numerical examples demonstrate the feasibility and capability of the proposed model. Furthermore, the augmented LTM can be applied to any form of node model, including those discussed in Chapters 3 and 4.

This chapter includes the following publication:
Bliemer, M.C.J. **Gong, X.** & Raadsen, M.P.H. (2024) Incorporation of non-persistent delays at signalised intersections in the link transmission model. Submitted to *Transportation Research Part B: Methodological*.

Manuscript: Incorporation of Non-persistent Delays at Signalised Intersections in the Link Transmission Model

Abstract

In macroscopic dynamic traffic assignment, traffic flows are allocated to a transport network by means of a dynamic network loading model, with travel time being a primary output. Dynamic network loading not only determines the flows on roads, facilitated by a link model, but also governs flows passing through intersections by means of a node model. Notably, in urban settings, most travel time delays arise due to queue formations at intersections. These intersection delays can be categorised into persistent and non-persistent delays. Contemporary flow-based models – such as link transmission models (LTM) – account for persistent delays through boundary conditions imposed by the node model. Non-persistent delays are absent in most existing LTM formulations unless green and red phases are explicitly simulated at a more detailed microscopic level. In this work, we propose a novel methodology to directly integrate non-persistent delays into the LTM formulation in which the effects of traffic lights are averaged without violating the First-In-First-Out (FIFO) principle this model relies on. To achieve this, one or more virtual links are implicitly considered in an augmented link model representation, wherein Webster's uniform delay is reformulated as a hypocritical branch of an associated fundamental diagram. Augmented algorithms for networks are provided for the cases of link-specific or movement-specific exit capacities and numerical examples are provided to illustrate the new approach. The proposed methodology results in a more realistic depiction of traffic dynamics by directly embedding non-persistent delay in the network loading propagation scheme. In doings so, enhanced accuracy in travel time and traffic flow predictions compared to existing best-practice in macroscopic dynamic network loading can be achieved.

Keywords: Macroscopic dynamic network loading, link transmission model, intersection delay, non-persistent delay, Webster's uniform delay, first-in-first-out principle

5.1 Introduction

Modelling of traffic flows typically occurs as part of dynamic *traffic assignment* (DTA). It distributes predicted travel demand to the infrastructure network, producing traffic flows and travel time forecasts. DTA models generally contain two key components: a route choice model and a dynamic network loading (DNL) model. The route choice model distributes traffic across feasible routes based on generalised costs that include travel time, whereas the DNL model propagates traffic along the chosen routes from origin to destination through a network consisting of links and nodes. Links represent homogeneous road segments with attributes such as length, speed limit, and the number of lanes. Nodes correspond to junctions or intersections with permissible movements (e.g., left turn, through) and are potentially equipped with traffic signal control. This research focuses on *macroscopic DNL models*. Macroscopic models describe traffic in terms of flow, in contrast to microscopic models that consider individual vehicles. Macroscopic DNL models are computationally efficient and allow applications on large networks typically achieve equilibrium convergence with more ease than their microscopic counterparts, i.e., they are more stable. However, they lack the detailed description of traffic flow dynamics on the individual vehicle level present in microscopic models.

In urban networks, travel times consist of time spent traversing a link and any additional intersection delays. The latter is often the primary source of delay and has two principal components, namely persistent and non-persistent delay. *Persistent delay*, also referred to as overflow delay, arises when – on average – traffic demand surpasses road capacity, leading to queue formation whereby the queue persists for some time before it dissipates. In case of a signalised intersection this reflects the situation where vehicles wait more than a single cycle for passage, in other situations it may be caused by lane-drops, or other physical changes to a road. *Non-persistent delays* occur when there is a short temporary disruption in traffic flow without causing any capacity problems. In case of a signalised intersection this reflects the situation where vehicles temporarily wait for a red signal but all vehicles are able to exit during the green phase. Non-persistent delays may also occur in other situations, such as waiting for pedestrians to cross or yielding at an unsignalised intersection.

Persistent delays at signalised intersections are adequately captured in most DNLs by considering capacity constraints, but non-persistent delays are often ignored. In practice, most intersections are not saturated and hence capturing non-persistent delays is important for providing meaningful travel time forecasts and for describing realistic traffic flow dynamics. There exist different types of non-persistent delays, including uniform delay and random delay. *Uniform delay* assumes the average delay that a vehicle experiences at a signalised intersection when vehicles arrive at a uniform rate, whereas random delay is the additional delay due to vehicles arriving randomly rather than uniformly at the intersection. In this study we focus on uniform delay as the main component of non-persistent delay.

Average uniform delay can be computed using Webster's well-known formula, see Section 2.2, and it depends on the cycle length c of the intersection signal control, as well as the effective green time g and flow-to-capacity ratio of the considered movement. Table 1 illustrates the ranges of uniform delay for various cycle lengths and effective green time ratios, whereby the lower bound reflects average delay when flow is negligible and the upper bound reflects average delay when flow is at capacity. For a typical signalised intersection with a cycle length of 90 seconds and four phases, average uniform delay is around 30 seconds but is flow dependent. Ignoring uniform delay in DNLs would significantly underestimate travel times in urban transport networks.

Table 1. Range of average uniform delay in seconds (rounded).

	Cycle length (c)				
Green ratio (g/c)	60 sec	90 sec	120 sec	150 sec	180 sec
0.1	24 – 27	36 – 41	49 – 54	61 – 68	73 – 81
0.2	19 – 24	29 – 36	38 – 48	48 – 60	58 – 72
0.3	15 – 21	20 – 32	29 – 42	37 – 53	44 – 63
0.4	11 – 18	16 – 27	22 – 36	27 – 45	32 – 54
0.5	8 – 15	11 – 23	15 – 30	19 – 38	23 – 45
0.6	5 – 12	7 – 18	10 – 24	12 – 30	14 – 36
0.7	3 – 9	4 – 14	5 – 18	7 – 23	8 – 27
0.8	1 – 6	2 – 9	2 – 12	3 – 15	4 – 18
0.9	0 – 3	0 – 5	1 – 6	1 – 8	1 – 9

In this study we consider approaches to include uniform delay in macroscopic DNL models that consider signalised intersections, and discuss advantages and disadvantages. We propose a novel method that fully integrates average uniform delay during a signal cycle within the *link transmission model*, a well-known first order macroscopic DNL. We do so by considering an augmented network with virtual links governed by an uncongested branch of the fundamental diagram that is constructed such that it is consistent with Webster's uniform delay. The link transmission model is applied to both the physical and virtual links, which results in an augmented time-discretised algorithm that does not require explicit creation of these virtual links in the network, making it straightforward to apply and requiring only a modest increase in computation time.

The remainder of the paper is structured as follows. Section 5.2 provides a current state-of-the-art in macroscopic DNL models, discusses existing intersection delay models as well as existing approaches to include uniform delay in macroscopic DNL models. Section 5.3 presents the general link transmission model without signal control. In Section 5.4 we propose an augmented link representation to account for non-persistent (and persistent) delays for signalised intersections. Section 5.5 formulates the augmented link transmission model for two cases: link-specific and movement-specific exit capacities. In Section 5.6, the proposed methodology is embedded in a time-discretised algorithm that can be applied to networks. Numerical examples that illustrate our main contribution are provided in Section 5.7 and we finish with conclusions in Section 5.8.

5.2 Literature review

5.2.1 Macroscopic DNL models

In static traffic assignment and network loading, non-persistent delays can simply be added to the link and path travel times in post-processing. However, dynamic models incorporate temporal variations in flow, where non-persistent delays should be considered within the network loading process. In this context we distinguish between exit flow-based and exit time-based DNL models.

Exit time-based models were a seemingly natural extension to static network loading. They adopt a dynamic version of a travel time function, for example the Bureau of Public Roads

(BPR) function (Bureau of Public Roads, 1964). These dynamic travel time functions are used directly to govern the propagation of traffic flow by letting flow exit each link after the predicted travel time has elapsed (Janson, 1991; Friesz et al., 1993; Astarita, 1996; Chabini, 1998; Wu et al., 1998; Xu et al., 1999; Bliemer & Bovy, 2003). While this approach would facilitate the integration of non-persistent delays in a relatively straightforward manner, such models have largely been abandoned in practice because they struggle to properly include persistent delays due to the lack of capacity constraints (or a node model). Further, these models often cannot avoid first-in-first-out (FIFO) violations.

Exit flow-based models, such as Merchant and Nemhauser (1978a, 1978b), reflect congestion effects in flow propagation via the use of exit-flow functions. Kinematic wave models based on the seminal work of Lighthill and Whitham (1955) and Richards (1956) propagate traffic flows along links utilising (cumulative) flow rates, including the cell transmission model (CTM), originally proposed by Daganzo (1994, 1995), and the link transmission model (LTM), originally proposed by Yperman et al. (2005) and based on the work of Newell (1993a, 1993b). These models rely on the FIFO principle to be tractable. CTM employs the first-order finite Godunov scheme for its solution (Lebacque, 1996), while LTM solves the kinematic wave model, implicitly or explicitly, using the Lax-Hopf equation (Bliemer & Raadsen, 2019; Han et al., 2016; Jin, 2015). The fundamental diagram (FD) underpins all contemporary first-order macroscopic DNL models. It governs the relationship between traffic density, flow, and speed on each link. While the original LTM considered a simplified triangular FD, extensions were proposed to consider more general concave FDs (Gentile, 2010; Friesz et al., 2013; Van der Gun et al., 2017; Bliemer & Raadsen, 2019; Raadsen & Bliemer, 2019). CTM and LTM are predominantly link models that can operate on a network level after combining them with a node model that regulates flows across intersections, see (Tampère et al., 2011; Flötteröd and Osorio, 2017). Via these node models they account for persistent delays by imposing boundary restrictions combined with link level properties such as capacities and jam density. However, they generally fail to capture non-persistent delays endogenously within the network loading process. The only exception is Yperman and Tampère (2006) who describe a time-discretised algorithm to include Webster's uniform delay via a point queue at the end of a link governed by a triangular fundamental diagram, based on certain assumptions on the construction of cumulative flow curves. Our newly proposed methodology can be seen as an extension of, and a theoretical foundation for, the algorithm proposed by Yperman and Tampère (2006).

5.2.2 Intersection delay models

Intersection delay models can be categorised as either *stochastic* (having a probabilistic aspect) or *deterministic*, distinguished primarily by their underlying assumptions concerning arrival rates. Early stochastic intersection delay models by Beckmann et al. (1955), and Newell (1960, 1965) pioneered the derivation of expected delay at traffic signals assuming a constant service rate coupled with a binomial distribution of arrivals. Poisson distributed pattern of arrivals were pioneered by Miller (1969) and Ohno (1978). More recent work by Viti and Van Zuylen (2010) proposed a probabilistic queuing model to explain dynamic and stochastic behaviours of queues at fixed controlled signals, which has also been used to estimate queues and signal phase durations. In contrast to stochastic models, deterministic models assume a uniform arrival rate. This simplification facilitates the estimation of average delays experienced by vehicles at intersections. In the remainder of this work, we focus on deterministic intersection delay approaches as it is the most widely adopted in practice in a transport planning context.

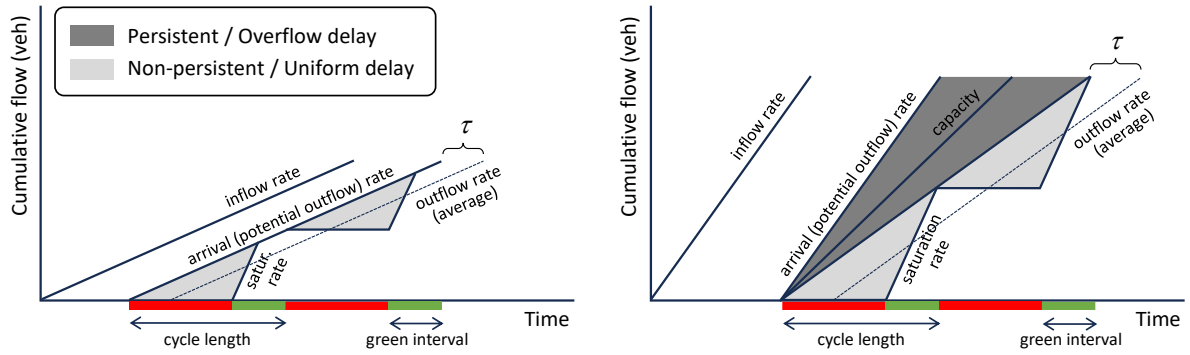


Fig. 1. Cumulative flows and delay accumulation at an approach to a traffic signal for (a) uncongested/unsaturated situation, and (b) congested/saturated situation

Webster's delay model (Webster, 1958) is one of the most widely used deterministic intersection delay models. It is applicable to both saturated flows and non-saturated flows, provided these conditions persist for a duration of the entire cycle. Fig. 1 provides illustrations of persistent and non-persistent delays using cumulative flows in unsaturated and saturated conditions. Webster's uniform delay represents the analytical expression of the average experienced non-persistent delay assuming a uniform arrival pattern, and is defined as:

$$\tau(q) = \frac{c\left(1-\frac{g}{c}\right)^2}{2\left(1-\frac{q}{Q}\right)} = \frac{c\left(1-\frac{g}{c}\right)^2}{2\left(1-\frac{g}{c}\cdot\frac{q}{C}\right)}, \tag{1}$$

where τ is the average uniform delay, g represents the effective green time, c is the cycle length, q denotes the arrival flow rate, Q is the link saturation flow, and $C = \frac{g}{c}Q$ represents the link exit capacity. The minimum and maximum delay are experienced when $q = 0$ and $q = C$, respectively, such that

$$\tau(q) \in \left[\frac{(c-g)^2}{2c}, \frac{c-g}{2}\right], \tag{2}$$

where $c-g$ indicates the red time per cycle, i.e., non-persistent delay is half the red time at most. This range is numerically illustrated in Table 1. Note the degree of saturation, q/C, in Eq. (1) should never exceed 1, therefore if $q > C$ then it is simply set to 1. Further note that the (average) outflow rate may be lower than the exit capacity due to inflow restrictions on downstream links. We revisit Webster's uniform delay in Section 5.4.

Webster's uniform delay is just one of many deterministic non-persistent delay formulations in existence (Beckmann et al., 1956; McNeil, 1968; Miller, 1968; Akçelik, 1980). However, as mentioned, these non-persistent delays are typically calculated separately from the DNL algorithm, potentially leading to inconsistencies and/or incorrect traffic flows. Knowing that persistent delay calculation has been successfully integrated in network loading models through

the use of a node model (Flötteröd & Rohde, 2011; Tampère et al., 2011; Bliemer et al., 2014; Jabari, 2016), or more recently node models considering signals (Yahyamozdarani & Tampère, 2023), it is only natural to try and seek the same integration for non-persistent delays.

5.2.3 Modelling of signalised intersections in DNL

According to Yperman and Tampère (2006), two main options for modelling traffic signal controls in macroscopic DNL models exist. The first involves the modelling of red and green phases explicitly, similar to microscopic traffic simulation models, rather than applying an average reduction in capacity (Shirke et al., 2019). The explicit modelling of traffic control mechanisms, despite its precise representation of traffic signals, is hardly ever employed in practice in a macroscopic model because information about a phase design is required for each signalised intersection, which for strategic planning purposes on large scale networks is typically considered too much detail. Further, in macroscopic modelling one is usually mainly interested in average flow rates and travel times across cycles rather than fluctuations within a cycle.

The second option considers average effects of traffic lights without explicitly simulating them directly. This second option is recommended by Durlin and Henn (2005) and Yperman and Tampère (2006). Yperman and Tampère (2006) and Yperman (2007) implicitly consider uniform delays in LTM by converting the delay into a number of vehicles in temporary point queues and shifting the cumulative flow curves. Their method is proposed in the form of an algorithm without providing a theoretical foundation. To the best of our knowledge, the novel DNL model presented in the next sections is the first to embed non-persistent delays directly into a first order DNL model in a consistent and theoretically rigorous manner, and accounts for fanning behaviour when flow increases.

5.3 General continuous-time link transmission model

In this study we adapt the continuous-time LTM formulation to describe the solution to the kinematic wave model with a general concave fundamental diagram. This is an exact formulation without the need to discretise time or space, but we would like to point out that other solutions to the kinematic wave model, such as CTM, could also be used. LTM assuming general concave fundamental diagrams can be solved using the discrete time-based algorithm proposed by Gentile (2010) or the event-based algorithm proposed by Raadsen and Bliemer (2019). In the coming sections we focus on the link model, which is where our main contribution lies. Multicommodity flows on networks are supported when we embed our methodology in the algorithm presented in Section 5.6.

Let us now introduce the mathematical notation – based on Bliemer and Raadsen (2019) – which does not yet consider traffic signal control, but is required as the starting point for our methodology. In subsequent sections we extend this formulation to account for signalised intersections and non-persistent delays.

5.3.1 Model input and notation

Consider a road network represented by a directed graph of links and nodes. Consider a single node with incoming links $i \in I$ and outgoing links $j \in J$. Each link is considered a homogenous road segment with the following attributes: free-flow speed $\sigma_i^{\max}$, length L_i, saturation flow Q_i, critical density κ_i, and jam density K_i. An overview of the variables used in this section is provided in Table 2.

Table 2. List of variables in link transmission model.

Variable	*Unit*	*Description*
Input		
L_i	km	Physical length of link *i*
$\sigma_i^{\max}$	km/h	Maximum vehicle speed of link *i*
$\gamma_i^{\max}$	km/h	Maximum kinematic wave speed on link *i*
$\gamma_i^{\min}$	km/h	Minimum kinematic wave speed on link *i*
Q_i	veh/h	Saturation flow of link *i*
K_i	veh/km	Jam density of link *i*
κ_i	veh/km	Critical density of link *i*
Speeds		
σ_i	km/h	Vehicle speed on link *i* (hypocritical $\sigma_{\mathrm{I},i}$, hypercritical $\sigma_{\mathrm{II},i}$)
γ_i	km/h	Kinematic wave speed on link *i* (hypocritical $\gamma_{\mathrm{I},i}$, hypercritical $\gamma_{\mathrm{II},i}$)
Cumulative flows		
$U_i(t)$	veh	Cumulative realised link inflow on link *i* at time *t*
$\tilde{U}_i(t)$	veh	Cumulative potential link outflow on link *i* at time *t*
$V_i(t)$	veh	Cumulative realised link outflow on link *i* at time *t*
$\tilde{V}_i(t)$	veh	Cumulative potential link inflow on link *i* at time *t*
Flow rates		
$u_i(t)$	veh/h	Realised link inflow rate on link *i* at time *t*
$\tilde{u}_i(t)$	veh/h	Potential link outflow rate on link *i* at time *t*
$v_i(t)$	veh/h	Realised link outflow rate on link *i* at time *t*
$\tilde{v}_i(t)$	veh/h	Potential link inflow rate on link *i* at time *t*
$s_i(t)$	veh/h	Sending flow on link *i* at time *t*
$r_i(t)$	veh/h	Receiving flow on link *i* at time *t*
$d_{ij}(t)$	veh/h	Realised transition flow from link *i* to link *j* at time *t*
Other		
$\varphi_{ij}(t)$	--	Flow splitting rate at the end of link *i* in the direction of link *j* at time *t*

LTM propagates traffic states forwards and backwards on each link by tracking cumulative inflows and outflows at the link boundaries. Adopting the notation[1] in Bliemer and Raadsen (2019), in the next subsections we first introduce how potential cumulative link outflow $\vec{U}_i(t)$ is determined based on realised cumulative link inflow $U_i(\cdot)$ at an earlier time instant, and how potential cumulative link inflow $\vec{V}_i(t)$ is determined based on realised cumulative link outflow $V_i(\cdot)$ at an earlier time instant. Then we describe how link sending flows $s_i(t)$ and receiving flows $r_i(t)$ can be computed based on $\vec{U}_i(t)$ and $\vec{V}_i(t)$, respectively. Finally, we describe how realised inflow rates $u_i(t)$ and outflow rates $v_i(t)$ are determined via a node model. Fig. 2 presents a visualisation of the variables on a single link.

The node model requires information about the splitting rates at the end of the link. Splitting rate $\varphi_{ij}(t)$ prescribes the proportion of flow exiting link i at time instant t that uses link j next on their path, where it holds that $\sum_{j\in J}\varphi_{ij}(t)=1$. In this section we assume that splitting rates are given at each time instant, but when we presented an augmented time-discretised algorithm to solve LTM in Section 5.6 we let splitting rates depend on network path flows.

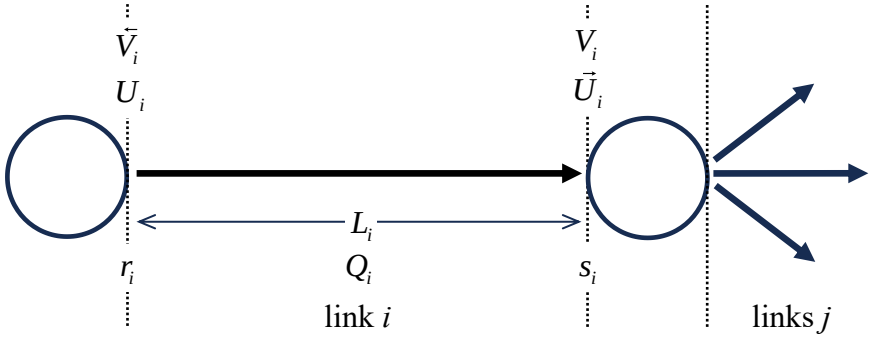


Fig. 2. Visualisation of LTM variables

5.3.2 Fundamental diagram

Propagation of traffic flow states on a link is governed by a fundamental diagram (FD), also referred to as a flux function, which describes the relationship between flow and density. For each link, consider a general concave FD as illustrated in Fig. 3 defined by a two-regime function $\Phi_i:[0,K_i]\rightarrow[0,Q_i]$, such that $q=\Phi_i(k)$ with flow rate q (veh/h) and density k. Traffic states can be classified as hypocritical (uncongested) and hypercritical (congested), see Cascetta (2009) or Bliemer et al. (2017). A traffic state is hypocritical when traffic is constrained by demand and flow increases with increased density. Further, it is hypercritical when traffic is constrained by supply and flow decreases with increased density. A general two-regime FD can be defined via

[1] Arrows are used on top of cumulative flow variables to indicate that the corresponding cumulative curves are constructed using shifts in a figure that shows cumulative vehicles over time, e.g., a shifted cumulative inflow curve represents a directly translated, *potential* cumulative outflow, rather than a *realised* outflow curve, see Bliemer and Raadsen (2019) for a visualisation and more information.

$$\Phi_i(k) = \begin{cases} \Phi_{\mathrm{I},i}(k), & \text{if } 0 \le k \le \kappa_i, \\ \Phi_{\mathrm{II},i}(k), & \text{if } \kappa_i \le k \le K_i, \end{cases} \tag{3}$$

where κ_i (veh/km) is the critical density, $\Phi_{\mathrm{I},i}(k)$ reflects the hypocritical branch, and $\Phi_{\mathrm{II},i}(k)$ represents the hypercritical branch. Assuming continuity, which is not strictly necessary (e.g., see Van der Gun et al. (2017)), it holds that $\Phi_{\mathrm{I},i}(\kappa_i) = \Phi_{\mathrm{II},i}(\kappa_i) = Q_i$.

Since LTM commonly uses flow rates instead of densities to describe traffic states at the link boundaries, we will consider the inverse of each branch, i.e., $k = \Phi_{\mathrm{I},i}^{-1}(q)$ and $k = \Phi_{\mathrm{II},i}^{-1}(q)$. Forward and backward wave speeds, $\gamma_{\mathrm{I},i}$ and $\gamma_{\mathrm{II},i}$ respectively, can be determined as the slope of the flux function. The maximum wave speed $\gamma_i^{\max} > 0$ is obtained at zero density and the minimum wave speed $\gamma_i^{\min} < 0$ is obtained at jam density. Assuming that these branches are strictly increasing and decreasing, respectively, these functions are invertible. Wave and vehicle speeds can then be expressed as a function of flow rates:

$$\gamma_{\mathrm{I},i}(q) = \frac{1}{\dfrac{d\Phi_{\mathrm{I},i}^{-1}(q)}{dq}}, \quad \text{and} \quad \gamma_{\mathrm{II},i}(q) = \frac{1}{\dfrac{d\Phi_{\mathrm{II},i}^{-1}(q)}{dq}}. \tag{4}$$

Further, corresponding vehicle speeds $\sigma_{\mathrm{I},i}$ and $\sigma_{\mathrm{II},i}$ can be computed as the ratio of flow and density,

$$\sigma_{\mathrm{I},i}(q) = \frac{q}{\Phi_{\mathrm{I},i}^{-1}(q)}, \quad \text{and} \quad \sigma_{\mathrm{II},i}(q) = \frac{q}{\Phi_{\mathrm{II},i}^{-1}(q)}. \tag{5}$$

noting that the maximum speed is obtained at zero density and is equal to the maximum wave speed, i.e., $\sigma_i^{\max} = \gamma_i^{\max}$. The various variables are graphically illustrated in Fig. 3.

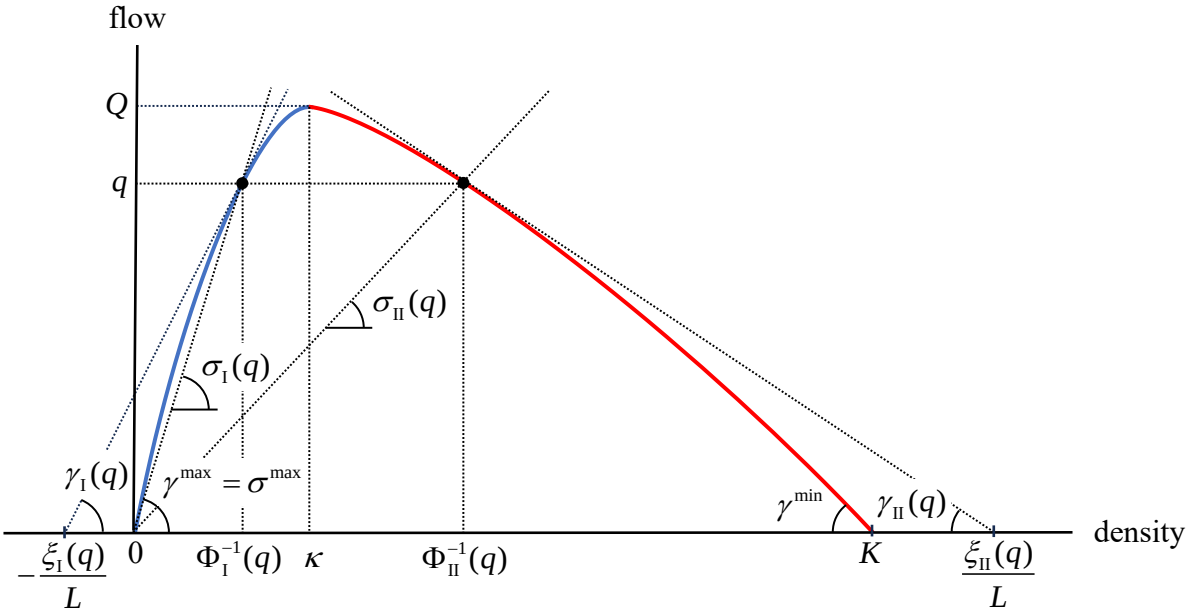


Fig. 3. Concave continuous two-regime fundamental diagram

5.3.3 Realised and potential cumulative link inflow and outflow

Based on realised link inflow and outflow rates $u_i(t)$ and $v_i(t)$, the realised cumulative link inflow and outflow are defined as

$$U_i(t) = \int_{x=-\infty}^{t} u_i(x)dx, \quad \text{and} \quad V_i(t) = \int_{x=-\infty}^{t} v_i(x)dx. \tag{6}$$

The FD dictates how traffic flow propagates from the upstream to the downstream link boundary and how queues propagate from the downstream to the upstream link boundary. During uncongested conditions, based on the realised cumulative link inflow U_i and the inverse hypocritical branch of the FD, $\Phi_{\mathrm{I},i}^{-1}$, potential cumulative link outflow $\vec{U}_i$ can be determined. Similarly, during spillback conditions, based on realised cumulative link outflow V_i and the hypercritical branch of the FD, $\Phi_{\mathrm{II},i}^{-1}$, potential cumulative link inflow $\overleftarrow{V}_i$ can be computed. Let $\vec{u}_i(t)$ (veh/h) and $\overleftarrow{v}_i(t)$ (veh/h) denote the potential link outflow and inflow rates at time instant *t*, respectively. Potential outflows would be realised only if there are no outflow (supply) constraints, whereas potential inflows would be realised only if there are no inflow (demand) constraints. The potential cumulative link outflows and inflows can respectively be defined as

$$\vec{U}_i(t) = \int_{x=-\infty}^{t} \vec{u}_i(x)dx, \quad \text{and} \quad \overleftarrow{V}_i(t) = \int_{x=-\infty}^{t} \overleftarrow{v}_i(x)dx. \tag{7}$$

As shown in Bliemer & Raadsen (2019), these potential cumulative outflows and inflows can be determined based on the Lax-Hopf formula, and are obtained as follows:

$$\vec{U}_i(t) = \min_{q\in[0,Q_i]} \left\{ U_i\left(t - \frac{L_i}{\gamma_{\mathrm{I},i}(q)} \right) + \xi_{\mathrm{I},i}(q) \right\}, \quad \text{with} \quad \xi_{\mathrm{I},i}(q) = L_i q \left(\frac{1}{\gamma_{\mathrm{I},i}(q)} - \frac{1}{\sigma_{\mathrm{I},i}(q)} \right), \tag{8}$$

$$\overleftarrow{V}_i(t) = \min_{q\in[0,Q_i]} \left\{ V_i\left(t + \frac{L_i}{\gamma_{\mathrm{II},i}(q)} \right) + \xi_{\mathrm{II},i}(q) \right\}, \quad \text{with} \quad \xi_{\mathrm{II},i}(q) = L_i q \left(\frac{1}{\sigma_{\mathrm{II},i}(q)} - \frac{1}{\gamma_{\mathrm{II},i}(q)} \right). \tag{9}$$

According to Eq. (8), the cumulative potential outflow is influenced by the cumulative inflow occurring exactly $L_i / \gamma_{\mathrm{I},i}(q)$ earlier, following the characteristic line of traffic state *q*. An adjustment of $\xi_{\mathrm{I},i}(q)$ is necessary to incorporate vehicles that have surpassed the characteristic wave during this interval, recognising that these vehicles have been moving at a speed greater than the wave speed $\gamma_{\mathrm{I},i}(q)$ in the same direction. In this context, $\xi_{\mathrm{I},i}(q)$ represents the number of passed vehicles while traversing the link downstream at kinematic wave speed $\gamma_{\mathrm{I},i}(q)$, see also Fig. 3. The minimisation over all possible traffic states is essential in case of a nonlinear FD as it accounts for expansion fans. A similar explanation can be provided for Eq. (9). We refer to Bliemer & Raadsen (2019) for more details.

In case of an asymmetric triangular FD (Newell, 1993a, 1993b), in which the hypocritical and hypercritical branches are both linear, it holds that $\gamma_{\mathrm{I},i}(q) = \gamma_i^{\max}$, $\gamma_{\mathrm{II},i}(q) = \gamma^{\min}$, $\xi_{\mathrm{I},i}(q) = 0$,

and $\xi_{\Pi,i}(q) = LK$, such that Eqs. (8) and (9) simplify to the formulation consistent with Yperman et al. (2005) and Yperman (2007), resulting in

$$\vec{U}_i(t) = U_i\left(t - \frac{L_i}{\gamma_i^{\max}}\right), \quad \text{and} \quad \overleftarrow{V}_i(t) = V_i\left(t + \frac{L_i}{\gamma_i^{\max}}\right) + L_i K_i. \tag{10}$$

5.3.4 Sending and receiving flows

Sending flow rate $s_i(t)$ (veh/h) represents the demand for vehicles wanting to exit the link, and is equal to the potential outflow rate $\vec{u}_i(t)$ if the link is entirely uncongested, and is equal to the saturation flow rate otherwise. A link is uncongested if the potential cumulative outflow is equal to the realised cumulative outflow, i.e., if $\vec{U}(t) = V(t)$. Receiving flow rate $r_i(t)$ (veh/h) represents the supply for vehicles wanting to enter the link, and is equal to the potential inflow rate $\overleftarrow{v}_i(t)$ if the link is entirely congested (i.e., in spillback), and is equal to the saturation flow rate otherwise. A link is in spillback if the potential cumulative inflow is equal to the realised cumulative inflow, i.e., if $\overleftarrow{V}_i(t) = U_i(t)$. In other words, sending and receiving flow rates can be computed as

$$s_i(t) = \begin{cases} \vec{u}_i(t), & \text{if } \vec{U}_i(t) = V_i(t), \\ Q_i, & \text{otherwise,} \end{cases} \quad \text{and} \quad r_i(t) = \begin{cases} \overleftarrow{v}_i(t), & \text{if } \overleftarrow{V}_i(t) = U_i(t), \\ Q_i, & \text{otherwise.} \end{cases} \tag{11}$$

Note that potential outflow rates $\vec{u}_i(t)$ and potential inflow rates $\overleftarrow{v}_i(t)$ are simply obtained as subgradients of $\vec{U}_i(t)$ and $\overleftarrow{V}_i(t)$, respectively, i.e., $\vec{u}_i(t) \in \partial\vec{U}_i(t)$ and $\overleftarrow{v}_i(t) \in \partial\overleftarrow{V}_i(t)$.

5.3.5 Realised link inflow and outflow rates

The node model considers for each node the demand and supply for all incoming and outgoing links and determines the realised link outflow rate rates for incoming links $i \in I$ and the realised link inflow rates for outgoing links $j \in J$.

Node models, as defined in the general class of macroscopic node models in Tampère et al. (2011), consider the following variables as input for each time instant *t*: (i) sending flows on incoming links, $\mathbf{s}(t) = [s_i(t)]$, (ii) receiving flows on outgoing links, $\mathbf{r}(t) = [r_j(t)]$, (iii) split proportions, $\boldsymbol{\varphi}(t) = \varphi_{ij}(t)$, and (iv) link saturation flows $\mathbf{Q} = [Q_i]$ to reflect merge priorities. The first two inputs come from the link model, in our case LTM but could also be CTM or another model based on kinematic wave theory.

Output of the node model are realised transition flow rates, $\mathbf{d}(t) = [d_{ij}(t)]$,

$$\mathbf{d}(t) = \Gamma\left(\mathbf{s}(t), \mathbf{r}(t), \boldsymbol{\varphi}(t), \mathbf{Q}\right), \tag{12}$$

where $\Gamma(\cdot)$ denotes an implicit node model function. In this paper we adopt the node model proposed by Tampère et al. (2011), but several other node models could be used that also satisfy the requirements for a first order node model, such as Flötteröd and Rohde (2011), Gibb

(2011), and Smits et al. (2015). One of these requirements is the "conservation of turn fractions", which relates to FIFO, such that the transition flow rates are proportional to the split proportions, i.e., $d_{ij}(t) / d_{ij'}(t) = \varphi_{ij}(t) / \varphi_{ij'}(t)$ for all $j' \in J$. Several numerical examples of the node model proposed by Tampère et al. (2011) are provided in Bliemer et al. (2014). Realised link outflow rates for links $i \in I$ and realised link inflow rates into downstream links $j \in J$ can be computed as

$$v_i(t) = \sum_{j \in J} d_{ij}(t), \qquad \text{and} \qquad u_j(t) = \sum_{i \in I} d_{ij}(t). \tag{13}$$

5.4 Augmented link representation

In this section we present a novel augmented link representation for signalised intersections to account for non-persistent delays (as well as persistent delays). First, we consider different cases of exit capacities on incoming links $i \in I$ based on the type of signal control and the lane configuration. Then we discuss how non-persistent delay can be accounted for in LTM by adding one or more virtual links at the end of each physical incoming link. Finally, we derive the fundamental diagram for these virtual links, consistent with Webster's uniform delay.

5.4.1 Exit capacities

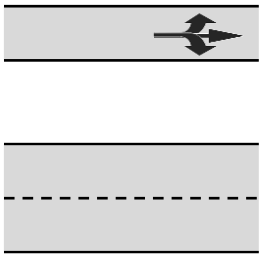
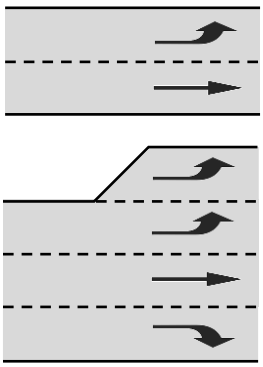
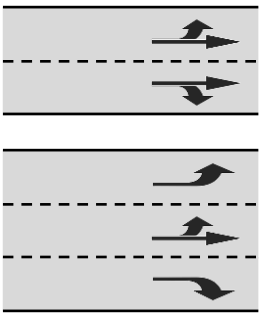
Table 3 provides an overview of three cases of signal control, whereby Case I is simplest and Case III is most general. Case I considers link-specific signal control, resulting in link-specific exit capacities and link-specific non-persistent delays. Case II considers movement-specific signal control, exit capacities and non-persistent delays, while Case III considers lane-specific signal control, exit capacities and non-persistent delays.

Case I assumes that all allowable movements on a link are controlled by the same traffic signal indicator, i.e., belong to the same signal group and phase. This specification is common when the link only has a single lane, or when it has multiple lanes whereby no explicit approach lane configuration is available. The resulting link-specific exit capacity, C_i, is given by

$$C_i = \left(\frac{g_i}{c} \right) Q_i, \tag{14}$$

where g_i is the effective green time for all movements on link *i* and *c* is the cycle length. This link exit capacity C_i is shared across all movements, whereby sending flow of movement *ij* at time *t* can claim its 'fair share' of capacity based on splitting proportions.

Table 3. Signal control and lane configurations

	Case I	*Case II*	*Case III*
Signal control	Link-specific	Movement-specific	Lane-specific
Lanes	Single, or multiple with unknown configuration	Multiple with known configuration, dedicated lanes only	Multiple with known configuration, support shared lanes
Example applications			

Case II assumes that a link has multiple dedicated lanes with a known configuration and each movement *ij* is controlled by a separate traffic signal with an effective green time of g_{ij}. Then the movement-specific exit capacity C_{ij} equals

$$C_{ij} = \left(\frac{g_{ij}}{c} \right) Q_{ij}, \quad \text{with} \quad Q_{ij} = \beta_{ij} Q_i, \tag{15}$$

where β_{ij} is a factor that describes the claim on exit capacity depending on – for example – the number of available lanes for movement *ij*. This formulation is consistent with the capacity formulation for signalised intersections proposed by Tampère et al. (2011), although without explicit consideration of a phase design. In the case of dedicated lanes, determining factors β_{ij} is straightforward; for example, in the first example of Case II in Table 3 one could set $\beta_{ij} = \frac{1}{2}$ for both the left turning movement as well as through traffic. While one would typically expect that $\sum_{j \in J} \beta_{ij} = 1$, it is possible that this sum is smaller or greater than one. In the same example, one could set $\beta_{ij} < \frac{1}{2}$ for the left turning movement to account for the fact that saturation flow is not achievable due to a necessary speed reduction to make the turn. In the second example of Case II in Table 3 an extra approach lane appears at the end of the link. In this case, Q_i refers to the link saturation flow of three lanes such that β_{ij} is equal to $\frac{1}{2}$ for left turns, and $\frac{1}{3}$ for through traffic as well as for right turns, resulting in $\sum_{j \in J} \beta_{ij} > 1$.

Case III assumes that a link has multiple lanes with a known configuration in which some, or all lanes, lanes are shared for multiple movements. Each lane *m*, $m = 1, \ldots, M_i$, on link *i* can be

controlled by a separate signal indicator with an effective green time of g_{im}. Then the lane-specific exit capacity C_{im} can be computed as

$$C_{im} = \left(\frac{g_{im}}{c} \right) Q_{im}, \quad \text{with} \quad Q_{im} = \tfrac{1}{M_i} Q_i. \tag{16}$$

Using a lane-based exit capacity requires distributing the sending flows for each movement across multiple lanes. Gong et al. (2024) proposed to distribute sending flows according to an equilibrium principle in which each driver chooses the exit lane that maximises their outflow potential and thereby minimises their delay. They did not consider non-persistent delay in their methodology. Extending their methodology to include non-persistent delays is likely possible but non-trivial and beyond the scope of this paper. Therefore, in this paper we will not consider Case III but rather assume that shared lanes are accounted for in Case II by assuming a fixed distribution of capacity to each movement. For example, in the first example of Case III we could conjecture a priori values for β_{ij} as $\frac{1}{2}$ for through traffic and $\frac{1}{4}$ for both left and right turns, but one could also base these factors on historical outflow patterns. Note that using split proportions to distribute capacities, i.e., setting $\beta_{ij} = \varphi_{ij}$, should be avoided, as it would make capacities time-dependent and demand proportional, which would violate the invariance principle (Lebacque & Khoshyaran, 2005) in the node model and result in unexpected outcomes.

5.4.2 Virtual links to account for non-persistent delays

To account for any non-persistent delays, one or more virtual links are added at the end of each incoming link. Fig. 4. illustrates the augmented link representation for Case I (link-specific exit capacities), while Fig. 5 illustrates the augmented link representation for Case II (movement-specific exit capacities). In each case, the representation of the physical link is the same as before, see Fig. 2. The additional variables used for the virtual links are listed in Table 4.

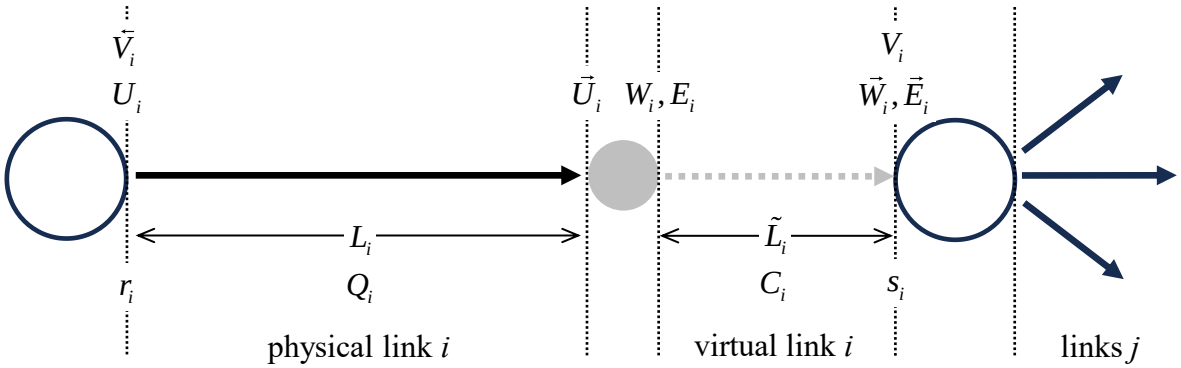


Fig. 4. Augmented link representation for Case I (link-specific exit capacities)

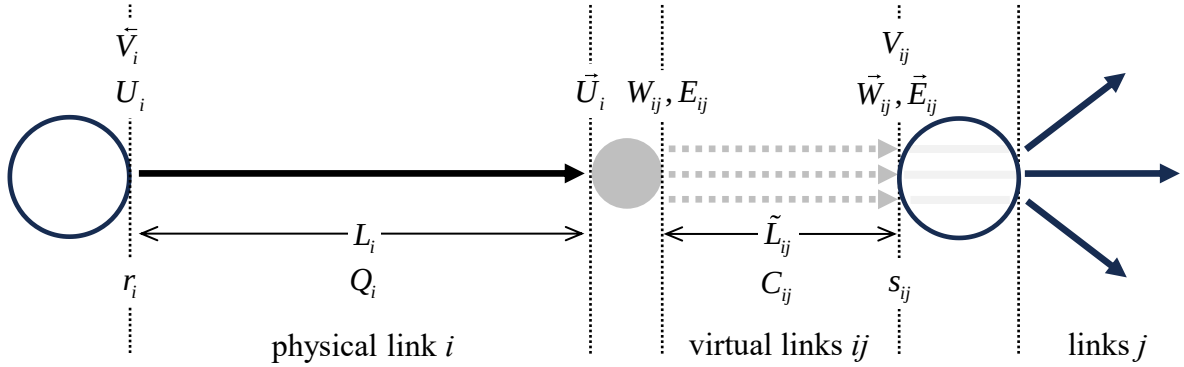


Fig. 5. Augmented link representation for Case II (movement-specific exit capacities)

Table 4. Additional variables in augmented link transmission model.

Variable	*Unit*	*Description*
Input		
$\tilde{L}_i, \tilde{L}_{ij}$	km	Length of virtual link *i* (Case I) or *ij* (Case II)
C_i, C_{ij}	veh/h	Exit capacity of virtual link *i* or *ij*
Q_i, Q_{ij}	veh/h	Saturation flow of virtual link *i* or *ij*
Speeds		
$\tilde{\sigma}_i, \tilde{\sigma}_{ij}$	km/h	Hypocritical vehicle speed on virtual link *i* or *ij*
$\tilde{\gamma}_i, \tilde{\gamma}_{ij}$	km/h	Hypocritical kinematic wave speed on virtual link *i* or *ij*
Cumulative flows		
$W_i(t), W_{ij}(t)$	veh	Cumulative realised link inflow on virtual link *i* or *ij* at time *t*
$\vec{W}_i(t), \vec{W}_{ij}(t)$	veh	Cumulative potential link outflow on virtual link *i* or *ij* at time *t*
$E_i(t), E_{ij}(t)$	veh	Cumulative realised auxiliary link inflow on virtual link *i* or *ij* at time *t*
$\vec{E}_i(t), \vec{E}_{ij}(t)$	veh	Cumulative potential auxiliary link outflow on virtual link *i* or *ij* at time *t*
Flow rates		
$w_i(t), w_{ij}(t)$	veh/h	Realised link inflow rate on virtual link *i* or *ij* at time *t*
$\vec{w}_i(t), \vec{w}_{ij}(t)$	veh/h	Potential link outflow rate on virtual link *i* or *ij* at time *t*
$e_i(t), e_{ij}(t)$	veh/h	Realised auxiliary link inflow rate on virtual link *i* or *ij* at time *t*

In Case I, the single virtual link has a capacity of C_i, calculated via Eq. (14), and a length $\tilde{L}_i$ (km) that is determined such that traversing this link at maximum link speed $\sigma_i^{\max}$ matches the minimum uniform delay, see Section 5.4.3. If $\tilde{L}_i \to 0$ then the model collapses to a typical LTM with signal control without accounting for non-persistent delays. Two new cumulative flow variables for the virtual link are introduced: realised cumulative inflow into the virtual link, denoted by $W_i(t)$, and the potential cumulative outflow out of the virtual link, denoted by $\vec{W}_i(t)$. It holds that

$$W_i(t) = \int_{x=-\infty}^{t} w_i(x)dx, \quad \text{and} \quad \vec{W}_i(t) = \int_{x=-\infty}^{t} \vec{w}_i(x)dx, \tag{17}$$

where $w_i(t)$ and $\vec{w}_i(t)$ are the corresponding realised inflow rate and potential outflow rate, respectively. Considering the same cumulative flows as in Fig. 1, Fig. 6 illustrates how our methodology captures uniform delay that varies across the signal cycle into an average uniform delay via the virtual link, where the horizontal difference between W and $\vec{W}$ indicates the average uniform delay. Note that in unsaturated conditions it holds that $\vec{U}_i(t) = W_i(t)$ and $\vec{W}_i(t) = V_i(t)$, see Fig. 6(a). In the same figure, a vehicle that enters the (physical) link at time instant t_1 will exit the link at time instant t_2. In general, a vehicle that exits the link at time *t* experienced a travel time of $t - U_i^{-1}(V_i(t))$, which includes non-persistent delay as well as persistent delay in saturated conditions. At any time *t*, the number of vehicles on the link is $U_i(t) - V_i(t)$.

In Case II, multiple virtual links are added at the end of the link to account for any non-persistent delay. For each movement *ij*, exit capacities C_{ij} are calculated via Eq. (15) and the length of the corresponding virtual link, $\tilde{L}_{ij}$, has a length such that $\tilde{L}_{ij} / \sigma_i^{\max}$ matches the minimum uniform delay for this movement. In contrast to Case I, we now need to explicitly track realised movement-specific cumulative link outflows $V_{ij}(t)$ since uniform delay can vary across movements. All other virtual link variables now also have indices *ij* to indicate that they now relate to movements. The movement-specific realised cumulative flow into the virtual link and cumulative potential link flow out of the virtual link are defined as

$$W_{ij}(t) = \int_{x=-\infty}^{t} w_{ij}(x)dx, \quad \text{and} \quad \vec{W}_{ij}(t) = \int_{x=-\infty}^{t} \vec{w}_{ij}(x)dx, \tag{18}$$

where $w_{ij}(t)$ and $\vec{w}_{ij}(t)$ are the corresponding realised movement-specific inflow rate and potential outflow rate, respectively, and $s_{ij}(t)$ is the movement-specific sending flow rate.

How to determine potential cumulative outflows, $\vec{W}_i(t)$ and $\vec{W}_{ii}(t)$, and corresponding sending flows $s_i(t)$ and $s_{ij}(t)$, will be discussed in Section 5.5. We first discuss the need for auxiliary flows in Section 5.4.3 before deriving the hypocritical branch of the fundamental diagram that governs the virtual link, consistent with average uniform delay, in Section 5.4.4.

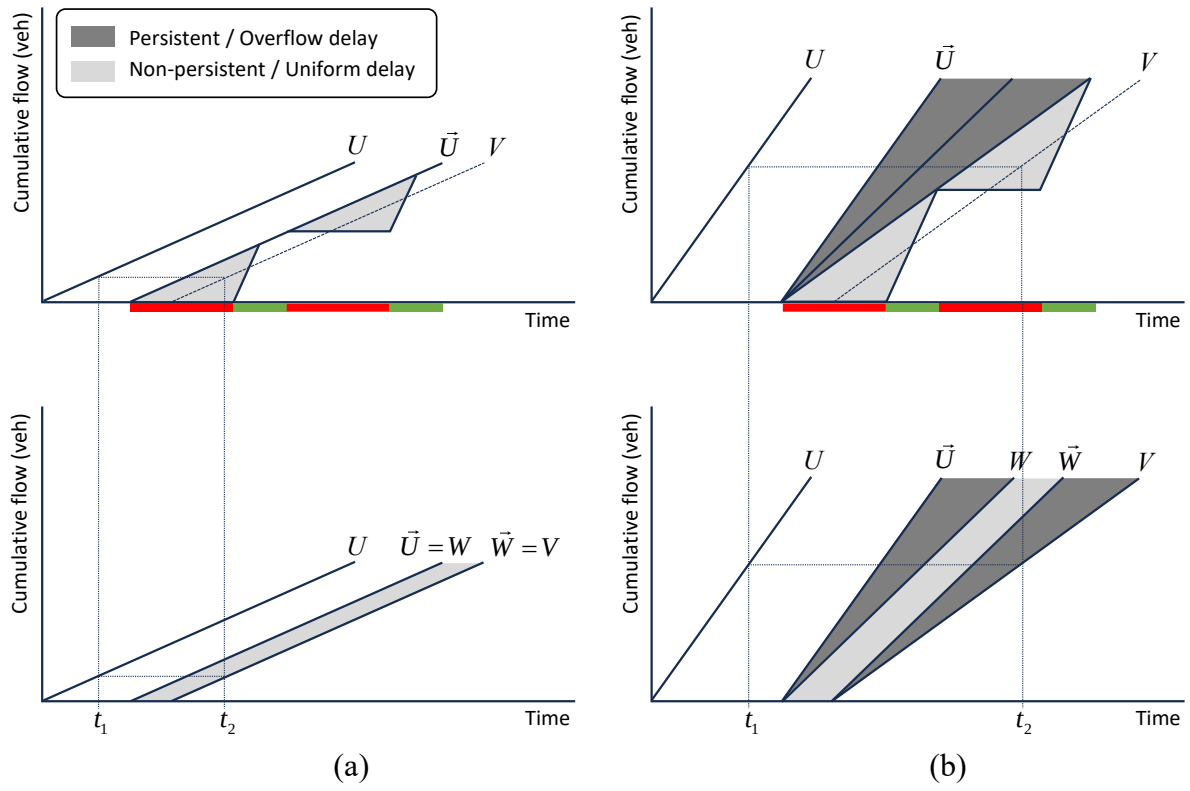


Fig. 6. Cumulative flows, delays, and virtual link variables in (a) uncongested/unsaturated situation, and (b) congested/saturated situation. Using explicit red and green intervals (top) versus proposed averaged approach (bottom).

5.4.3 Auxiliary flow

First, we consider Case I. It is important that $w_i(t) = C_i$ whenever there is any persistent queue and delay, either on the physical link or on the virtual link, to ensure that the uniform delay becomes half the red time. If the physical link is congested, the sending flow equals the link saturation flow so that $w_i(t) = C_i$ occurs naturally. However, in the special case where the virtual link is congested, but the physical link is uncongested, then $w_i(t) = C_i$ in general does not automatically happen because the potential outflow out of the physical link may be small. This situation occurs when the potential outflow of the virtual link exceeds the accepted outflow due to downstream receiving flow constraints, e.g., due to spillback. In this special case, to avoid uniform delay from being underestimated, we set $w_i(t)$ to capacity C_i on the virtual link by including extra (auxiliary) flow. This auxiliary flow does not represent actual vehicles, but rather it 'tricks' the virtual link to propagate its flows based on capacity, which yields the correct non-persistent delay. This auxiliary flow simply disappears when it reaches the end of the link. Let $e_i(t)$ be the auxiliary flow rate into the virtual link at time t, such that $w_i(t) - e_i(t)$ is the actual vehicle inflow rate on the virtual link. Let $E_i(t)$ and $\tilde{E}_i(t)$ denote the cumulative realised auxiliary inflow and cumulative potential auxiliary outflow, respectively. The realised cumulative auxiliary inflow into the virtual link is defined as

$$E_i(t) = \int_{x=-\infty}^{t} e_i(x)dx. \tag{19}$$

Since $W_i(t)$ is the total cumulative inflow into the virtual link, this means that $W_i(t) - E_i(t)$ is the actual cumulative vehicle inflow and $\vec{W}_i(t) - \vec{E}_i(t)$ the actual cumulative potential vehicle outflow. Let t' denote the time of link entrance of vehicle flow that exits the virtual link at time instant t. Because of FIFO this link entrance time instant can be determined by solving $W_i(t') = \vec{W}_i(t)$, which yields $t' = W_i^{-1}\left(\vec{W}_i(t)\right)$. Since the auxiliary flow is included in these cumulative flows, the potential cumulative auxiliary outflow can be computed as

$$\vec{E}_i(t) = E_i(t') = E_i\left(W_i^{-1}\left(\vec{W}_i(t)\right)\right). \tag{20}$$

If $\vec{W}_i(t) - \vec{E}_i(t) = V_i(t)$, then the virtual link has no excess flow, and if $\vec{W}_i(t) - \vec{E}_i(t) > V_i(t)$ then there are vehicles in a queue on the virtual link.

For Case II we consider auxiliary flow separately for each movement *ij*, whereby all relevant variables are defined in the same way but have subscript *ij*, namely $e_{ij}(t)$, $E_{ij}(t)$ and $\vec{E}_{ij}(t)$.

5.4.4 Fundamental diagram for virtual links

Without the need to make a distinction between link-specific or movement-specific non-persistent virtual links or delays, assuming cycle time *c*, effective green time *g*, link saturation flow *Q*, and arrival rate *q*, Webster's uniform delay τ is given in Eq. (1) as[2]

$$\tau(q) = \frac{\frac{c}{2}\left(1-\left(\frac{g}{c}\right)\right)^2}{1-\frac{q}{Q}}. \tag{21}$$

To derive a fundamental diagram for the virtual links consistent with uniform delay, we will rewrite Eq. (21) as

$$\tau(q) = \frac{\tilde{L}}{\tilde{\sigma}(q)}, \tag{22}$$

where $\tilde{L}$ is the length of the virtual link, and $\tilde{\sigma}(q)$ is the vehicle speed on this virtual link given flow rate *q*. Once we have this speed-flow relationship, it is easy to derive the corresponding flow-density relationship.

[2] To be consistent with units of the flow and speed variables – which are expressed in veh/h and km/h – we assume that *c*, *g*, and τ are expressed in hours instead of seconds. To avoid confusion, in our numerical examples in Section 5.7 we express *y* seconds as $y/3600$ (h).

To be able to derive a unique speed function $\tilde{\sigma}(q)$, we first need to normalise virtual link length $\tilde{L}$. Without loss of generality[3], we choose to normalise $\tilde{L}$ such that traversing this link with maximum link speed $\sigma^{\max}$ yields minimum uniform delay, i.e., $\tau(0) = \tilde{L} / \sigma^{\max}$, such that $\tilde{L}$ can be computed as

$$\tilde{L} = \frac{c}{2}\left(1-\left(\frac{g}{c}\right)\right)^2 \sigma^{\max}. \tag{23}$$

This means that we can now rewrite Webster's uniform delay as

$$\tau(q) = \frac{\frac{c}{2}\left(1-\left(\frac{g}{c}\right)\right)^2}{1-\frac{q}{Q}} = \left(\frac{\sigma^{\max}}{\sigma^{\max}}\right) \cdot \frac{\frac{c}{2}\left(1-\left(\frac{g}{c}\right)\right)^2}{1-\frac{q}{Q}} = \frac{\tilde{L}}{\tilde{\sigma}(q)}, \tag{24}$$

where, by definition, it then must hold that

$$\tilde{\sigma}(q) = \sigma^{\max}\left(1-\frac{q}{Q}\right). \tag{25}$$

Since density is flow divided by speed, it follows that the inverse fundamental diagram for the virtual link is

$$\tilde{\Phi}^{-1}(q) = \frac{q}{\tilde{\sigma}(q)}. \tag{26}$$

Note that this only represents the hypocritical branch of the fundamental diagram. The hypercritical branch is not needed since all backward propagation of traffic states will take place on the physical link as only the physical link length L is relevant when determining spillback. The above fundamental diagram is consistent with Webster's uniform delay and is illustrated in Fig. 7, shown together with a general concave fundamental diagram and a triangular fundamental diagram that could be used for the physical link. The corresponding kinematic wave speeds $\tilde{\gamma}$ are

$$\tilde{\gamma}(q) = \left(\frac{d\tilde{\Phi}^{-1}(q)}{dq}\right)^{-1} = \left(\frac{\sigma^{\max}}{\tilde{\sigma}^2(q)}\left(1-\frac{q}{Q}+\frac{q}{Q}\right)\right)^{-1} = \frac{\tilde{\sigma}^2(q)}{\sigma^{\max}}. \tag{27}$$

[3] In Section 5.5.1 it is shown that the model outcomes are independent of the chosen normalisation.

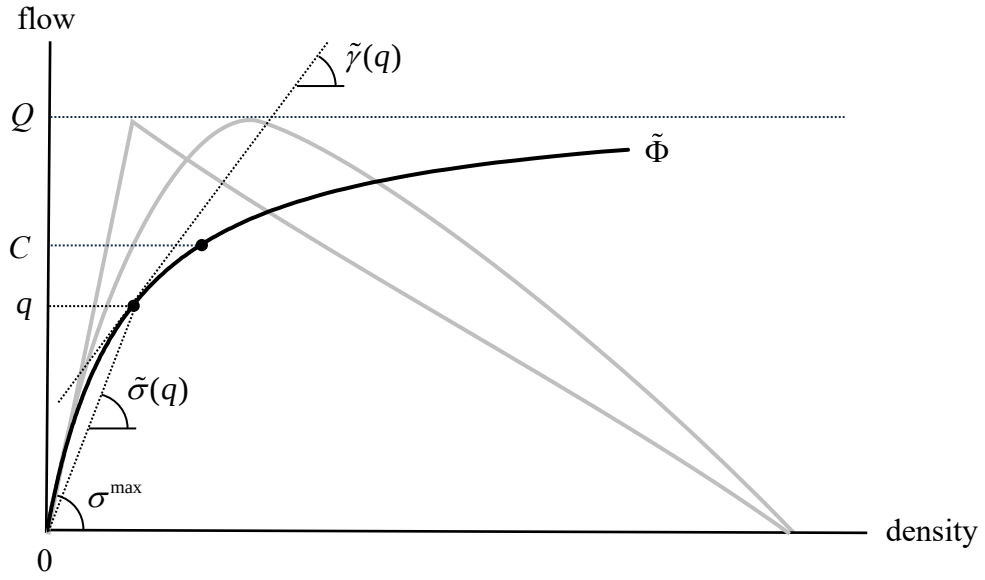


Fig. 7. Hypocritical part of the fundamental diagram for the virtual link

In Case I, the speed-flow relationship of each virtual link *i* can be written as

$$\tilde{\sigma}_i(q_i) = \sigma_i^{\max}\left(1-\frac{q_i}{Q_i}\right), \quad \text{for} \quad 0 \le q \le \left(\frac{g_i}{c}\right)Q_i, \tag{28}$$

and this relationship is applicable across the entire virtual link with length

$$\tilde{L}_i = \frac{c}{2}\left(1-\left(\frac{g_i}{c}\right)\right)^2 \sigma_i^{\max}. \tag{29}$$

In Case II, the speed-flow function and length of each virtual link *ij* can be written as

$$\tilde{\sigma}_{ij}(q_{ij}) = \sigma_i^{\max}\left(1-\frac{q_{ij}}{Q_{ij}}\right), \quad \text{for} \quad 0 \le q \le \left(\frac{g_{ij}}{c}\right)Q_{ij} = \beta_{ij}\left(\frac{g_{ij}}{c}\right)Q_i, \tag{30}$$

$$\tilde{L}_{ij} = \frac{c}{2}\left(1-\left(\frac{g_{ij}}{c}\right)\right)^2 \sigma_i^{\max}. \tag{31}$$

5.5 Augmented link transmission model

Now that the augmented link representation and the shape of the fundamental diagram of the virtual links has been described, the augmented link transmission model can be formulated.

Regarding forward propagation of traffic states and vehicle flow in both Cases I and II, Eq. (8) determines how potential cumulative link outflow $\vec{U}_i(t)$ is determined based on realised cumulative link inflow $U_i(t)$. Further, with respect to the backward propagation of traffic states and queues, Eq. (9) determines how potential cumulative link inflow $\overleftarrow{V}_i(t)$ is determined based on realistic cumulative link outflow $V_i(t)$. It is important to note that, even with our added virtual links, the backward propagation only needs to consider physical link length L_i since spillback only depends on the physical amount of space on the link to store the queue. In other words, any vehicle in a queue on the virtual link is implicitly stored on the physical link when determining whether the link is in spillback. In Case II, the link aggregated cumulative realised flows for the virtual link at time *t* are defined as

$$W_i(t) = \sum_{j \in J} W_{ij}(t), \qquad E_i(t) = \sum_{j \in J} E_{ij}(t), \quad \text{and} \quad V_i(t) = \sum_{j \in J} V_{ij}(t). \tag{32}$$

For notational convenience we define $\tilde{S}_i^+(t)$ (veh) as the excess sending flow on the physical link, which is defined as the difference between the potential outflow and the actual outflow on the physical link,

$$\tilde{S}_i^+(t) = \vec{U}_i(t) - W_i(t) + E_i(t). \tag{33}$$

In other words, $\tilde{S}_i^+(t) = 0$ if the physical link has no excess vehicles waiting to exit at time *t*, and $\tilde{S}_i^+(t) > 0$ if there are vehicles waiting in a queue on the physical link. This happens when the sending flow on the physical link exceeds the exit capacity.

Similarly, we define $S_i^+(t)$ (for Case I) and $S_{ij}^+(t)$ (for Case II) as the excess sending flow on the virtual link, which is defined as the difference between the potential actual (not auxiliary) outflow and the realised actual outflow on the virtual link,

$$S_i^+(t) = \vec{W}_i(t) - \vec{E}_i(t) - V_i(t), \quad \text{or} \quad S_{ij}^+(t) = \vec{W}_{ij}(t) - \vec{E}_{ij}(t) - V_{ij}(t). \tag{34}$$

If $S_i^+(t) = 0$ ($S_{ij}^+(t) = 0$) then the virtual link has no excess vehicles waiting to exit at time *t*, whereas if $S_i^+(t) > 0$ ($S_{ij}^+(t) > 0$) then there are vehicles waiting in a queue on the virtual link. This happens when the sending flow on the virtual link exceeds the downstream accepted outflow (e.g., due to spillback).

While receiving flows $r_i(t)$ are calculated as before by Eq. (11), sending flows and the application of the node model are modified in our augmented LTM. We formulate the augmented LTM for Case I and Case II in the following subsections.

5.5.1 Case I: Link-specific exit capacities

The realised virtual link inflow rate in uncongested conditions is equal to the potential outflow out of the physical link (but no more than the capacity), and is equal to capacity in congested conditions. This results in (see also Appendix D)

$$w_i(t) = \begin{cases} \min\{\vec{u}_i(t), C_i\}, & \text{if } \tilde{S}_i^+(t) = 0 \text{ and } S_i^+(t) = 0, \\ C_i, & \text{otherwise.} \end{cases} \tag{35}$$

In the special case that congestion only exists on the virtual link, the auxiliary flow is positive and equal to the difference between capacity and the accepted inflow from the physical link. In this situation, the actual vehicle inflow into the virtual link is $\min\{\vec{u}_i(t), C_i\}$ while the auxiliary flow supplements this to C_i. This means that we can formulate the auxiliary flow rate on the virtual link as (see Appendix D)

$$e_i(t) = \begin{cases} C_i - \min\{\vec{u}_i(t), C_i\}, & \text{if } \tilde{S}_i^+(t) = 0 \text{ and } S_i^+(t) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

If the virtual link has no vehicles queuing on the link, its sending flow is equal to the potential outflow rate, and is otherwise equal to the link exit capacity,

$$s_i(t) = \begin{cases} \vec{w}_i(t), & \text{if } S_i^+(t) = 0, \\ C_i, & \text{otherwise,} \end{cases} \tag{37}$$

where $\vec{w}_i(t) \in \partial \vec{W}_i(t)$. Based on realised cumulative inflow into the virtual link, $W_i(t)$, which can be calculated using Eq. (17), we can determine the potential cumulative outflow $\vec{W}_i(t)$ out of the virtual link analogous to Eq. (8),

$$\vec{W}_i(t) = \min_{q \in [0, C_i]} \left\{ W_i\left(t - \frac{\tilde{L}_i}{\tilde{\gamma}_i(q)} \right) + \tilde{\xi}_i(q) \right\}, \tag{38}$$

where $\tilde{L}_i$ is given in Eq. (23), $\tilde{\gamma}_i(q_i)$ is given in Eq. (27), and $\tilde{\xi}_i(q_i)$ is given by

$$\tilde{\xi}_i(q) = \tilde{L}_i q \left(\frac{1}{\tilde{\gamma}_i(q)} - \frac{1}{\tilde{\sigma}_i(q)} \right) = \tilde{L}_i q \left(\frac{\sigma_i^{\max} - \tilde{\sigma}_i(q)}{\tilde{\sigma}_i^2(q)} \right). \tag{39}$$

Substituting Eqs. (23), (25), and (27) into (38) and (39) yields

$$\vec{W}_i(t) = \min_{q \in [0, C_i]} \left\{ W_i\left(t - \frac{\tau_i(q) Q_i}{Q_i - q} \right) + \frac{\tau_i(q) q^2}{Q_i - q} \right\}, \tag{40}$$

where $\tau_i(q_i)$ is Webster's uniform delay function in Eq. (21). Note that the minimisation in Eq. (40) is over the range $[0, C_i]$, and since $C_i < Q_i$ there is no risk of division by zero. Also note that $\vec{W}_i(t)$ does not depend on virtual link length $\tilde{L}_i$, therefore any other (speed) normalisation for the virtual link length could be used. Eq. (40) constitutes one of our main results.

The node model formulated in Eq. (12) can be used to compute transition flow rates out of the incoming virtual links where vector $\mathbf{Q}$ is replaced by vector $\mathbf{C} = [C_i]$,

$$\mathbf{d}(t) = \Gamma\left(\mathbf{s}(t), \mathbf{r}(t), \boldsymbol{\varphi}(t), \mathbf{C}\right). \tag{41}$$

Resulting realised link outflow rates, $v_i(t)$, and realised inflow rates into downstream links, $u_j(t)$, can be determined using Eq. (13).

5.5.2 Case II: Movement-specific exit capacities

Now we consider virtual links for each movement as shown in Fig. 5. The virtual node that connects the physical and virtual links is considered a diverge node such that flow transfers are based on the most restrictive virtual link capacity. Any vehicles that are not able to exit the virtual link are stored in a queue on the physical link. Similar to Case I, the physical link is uncongested if $\vec{U}_i(t) = W_i(t)$ and is congested if $\vec{U}_i(t) > W_i(t)$, while for each virtual link *ij* it holds that it has no excess flow if $\vec{W}_{ij}(t) - \vec{E}_{ij}(t) = V_{ij}(t)$ and has excess flow if $\vec{W}_{ij}(t) - \vec{E}_{ij}(t) > V_{ij}(t)$, where $W_i(t)$ and $V_i(t)$ are defined in Eq. (32). As shown in Appendix D, the realised inflow rate into virtual link *ij* at time instant *t* can be determined as

$$w_{ij}(t) = \begin{cases} \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ \vec{u}_i(t), C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^+(t) = 0 \text{ and } S_{ij}^+(t) = 0, \\ \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ Q_i, C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^+(t) > 0 \text{ and } S_{ij}^+(t) = 0, \\ C_{ij}, & \text{otherwise,} \end{cases} \tag{42}$$

where $\tilde{\varphi}_{ij}(t)$ are the split proportions at the end of the physical link, noting that these are generally different from $\varphi_{ij}(t)$ since split proportions vary over time. The corresponding auxiliary flow rates are given as (see Appendix D)

$$e_{ij}(t) = \begin{cases} C_{ij} - \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ \vec{u}_i(t), C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^+(t) = 0 \text{ and } S_{ij}^+(t) > 0, \\ C_{ij} - \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ Q_i, C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^+(t) > 0 \text{ and } S_{ij}^+(t) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{43}$$

The movement-specific sending flow of the virtual link is equal to this potential outflow rate,

$$s_{ij}(t) = \begin{cases} \vec{w}_{ij}(t), & \text{if } S_{ij}^+(t) = 0, \\ C_{ij}, & \text{otherwise,} \end{cases} \tag{44}$$

where $\vec{w}_{ij}(t) \in \partial \vec{W}_{ij}(t)$. Based on movement-specific realised cumulative inflow into the virtual link, $W_{ij}(t)$, which can be calculated using Eq. (18), we can determine the movement-specific potential cumulative outflow out of the virtual link analogous to Eq. (40),

$$\vec{W}_{ij}(t) = \min_{q \in [0, C_{ij}]} \left\{ W_{ij}\left(t - \frac{\tau_{ij}(q) Q_{ij}}{Q_{ij} - q} \right) + \frac{\tau_{ij}(q) q^2}{Q_{ij} - q} \right\}. \tag{45}$$

Application of the node model at the end of the link now requires considering each virtual link as a separate incoming link, which relaxes FIFO at the link level to FIFO at the movement level. This is achieved by Eq. (41) with

$$\text{stack}[\text{diag}(\mathbf{d}(t))] = \Gamma\left(\text{vec}[\mathbf{s}(t)], \mathbf{r}(t), \text{stack}[\mathbf{1}_{|J|}], \text{vec}[\mathbf{C}]\right). \tag{46}$$

where $\text{stack}[\text{diag}(\mathbf{d}(t))]$ is a matrix of size $|I| \cdot |J| \times |J|$ consisting of stacked diagonal matrices containing the transition flow rates, $\text{vec}[\mathbf{s}(t)]$ and $\text{vec}[\mathbf{C}]$ are column vectors of length $|I| \cdot |J|$ in which all elements related to movements *ij* are stacked. The matrix with split proportions is now a binary matrix of size $|I| \cdot |J| \times |J|$ consisting of stacked identity matrices, which we denote as $\text{stack}[\mathbf{1}_{|J|}]$, such that all flow on virtual link *ij* is directed to outgoing link *j*. For example, in case of a node with two incoming links (numbered 1 and 2) and three outgoing links (numbered 3, 4, and 5) we would have for $\text{stack}[\text{diag}(\mathbf{d}(t))]$, $\text{vec}[\mathbf{s}(t)]$, $\text{stack}[\mathbf{1}_{|J|}]$, and $\text{vec}[\mathbf{C}]$ respectively

$$\begin{pmatrix} d_{13}(t) & 0 & 0 \\ 0 & d_{14}(t) & 0 \\ 0 & 0 & d_{15}(t) \\ d_{23}(t) & 0 & 0 \\ 0 & d_{24}(t) & 0 \\ 0 & 0 & d_{25}(t) \end{pmatrix}, \quad \begin{pmatrix} s_{13}(t) \\ s_{14}(t) \\ s_{15}(t) \\ s_{23}(t) \\ s_{24}(t) \\ s_{25}(t) \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} C_{13} \\ C_{14} \\ C_{15} \\ C_{23} \\ C_{24} \\ C_{25} \end{pmatrix}. \tag{47}$$

Resulting realised movement-specific link outflow rates, $v_{ij}(t)$, are simply $v_{ij}(t) = d_{ij}(t)$, whereas $v_i(t)$ and $u_j(t)$ can again be determined using Eq. (13).

5.6 Augmented LTM algorithm for networks

While in this paper we focus on our main contribution at the link level, in this section we present how to integrate our novel methodology into existing time-discretised network algorithms. Let the transport network be represented by graph $G(N, A)$ with nodes N and directed links A. Let $N^* \subseteq N$ is the subset of nodes with traffic signal control, and $A^* \subseteq A$ is the subset of links with downstream traffic signal control. For each node $n \in N$, let $I_n \subseteq A$ denote its incoming links and $J_n \subseteq A$ its outgoing links. Let P be the set of paths and let δ_{ip} be a link-path incidence indicator that equals 1 if link $i \in A$ is on path $p \in P$. Let $F_p(t)$ denote the given cumulative flow on path $p \in P$ at time t. Assume that time is discretised into time steps with duration $\Delta t \leq \min_{i \in A}\{L_i / \sigma_i^{\max}, \tilde{L}_i / \sigma_i^{\max}\}$, and consider simulation time period $[0, T]$.

Algorithm 1 lists the steps to solve the augmented LTM that accounts for non-persistent delays in Case I (link-specific exit capacities). Steps 2, 5, and 7e are analogous to existing LTM approaches, where path-specific cumulative link inflows are used to determine split proportions

and where path-specific flows are aggregated and disaggregated under FIFO assumptions. We refer to Yperman (2007) for more details. Steps 3d, 5 and 7e require linear interpolation of cumulative flows.

Compared to existing LTM approaches, the new steps are indicated with an asterisk (*), namely Steps 1b, 3c-d, 4c-d, 6b-c and 7c-d. In Step 4, instead of using sending and receiving flow rates $s_i(t)$ and $r_i(t)$, we instead use $S_i(t) = s_i(t) \cdot \Delta t$ and $R_i(t) = r_i(t) \cdot \Delta t$, expressed in vehicles. In Step 6b, transition flows $\mathbf{D}_n(t) = [D_{ij}(t)]_{i \in I_n,\, j \in J_n}$ at node $n \in N$ during time period $[t, t + \Delta t]$ are determined using the node model, where $\mathbf{S}_n(t) = [S_i]_{i \in I_n}$, $\mathbf{R}_n(t) = [R_j]_{j \in J_n}$, and $\mathbf{C}_n(t) = [C_i]_{i \in I_n}$. While previously we considered flow *rates* as input and output, since node model function $\Gamma(\cdot)$ is homogenous of degree 1 with respect to flows it holds that

$$\begin{aligned} \mathbf{D}_n(t) &= \Delta t \cdot \mathbf{d}_n(t) \\ &= \Delta t \cdot \Gamma\left(\mathbf{s}_n(t), \mathbf{r}_n(t), \boldsymbol{\varphi}_n(t), \mathbf{C}_n\right) \\ &= \Gamma\left(\mathbf{s}_n(t) \cdot \Delta t, \mathbf{r}_n(t) \cdot \Delta t, \boldsymbol{\varphi}_n(t), \mathbf{C}_n \cdot \Delta t\right) \\ &= \Gamma\left(\mathbf{S}_n(t), \mathbf{R}_n(t), \boldsymbol{\varphi}_n(t), \mathbf{C}_n \cdot \Delta t\right). \end{aligned} \tag{48}$$

In Algorithm 1, most computation time is spent in Steps 3a-c (applying the Lax-Hopf formula in the link model), Step 5 (determining split proportions from path flows), and Step 6a (applying the node model). In our augmented algorithm one additional application of the Lax-Hopf formula is required in Step 3, namely for $\vec{W}_i(t)$, while Steps 5 and 6a remain the same. This means that our augmented LTM algorithm requires only a modest increase in computation time.

Algorithm 2 in Appendix E lists the steps to solve the augmented LTM in Case II (movement-specific exit capacities). Given that flows need to be split across multiple virtual links, determining split proportions will now be different for links with and without downstream signal control as can be observed in Step 5. Also, movement-specific cumulative flows and sending flows are now required, which means increased computation time compared to Case I.

Algorithm 1. Pseudo-algorithm for augmented LTM for Case I (link-specific exit capacities)

Step 1: Initialisation.

a. Set $U_i(0) = U_{ip}(0) = V_i(0) = V_{ip}(0) = 0, \forall i \in A$.

b. * Set $W_i(0) = E_i(0) = 0, \forall i \in A^*$.

c. Set $t := 0$.

Step 2: Realised cumulative inflows.

a. $U_{ip}(t) = \begin{cases} F_p(t), & \text{if link } i \text{ is the first link on path } p, \\ V_{i'p}(t), & \text{if link } i' \text{ is the previous link on path } p, \end{cases} \quad \forall i \in A, \forall p \in P.$

b. $U_i(t) = \sum_{p \in P} U_{ip}(t), \quad \forall i \in A.$

Step 3: Potential cumulative flows.

a. $\vec{U}_i(t+\Delta t) = \min\limits_{q\in[0,Q_i]}\left\{U_i\left(t+\Delta t-\frac{L_i}{\gamma_{\mathrm{I},i}(q)}\right)+\frac{L_i q}{\gamma_{\mathrm{I},i}(q)}-\frac{L_i q}{\sigma_{\mathrm{I},i}(q)}\right\}, \quad \forall i\in A.$ [Eq. (8)]

b. $\overleftarrow{V}_i(t+\Delta t) = \min\limits_{q\in[0,Q_i]}\left\{V_i\left(t+\Delta t+\frac{L_i}{\gamma_{\mathrm{II},i}(q)}\right)+\frac{L_i q}{\sigma_{\mathrm{II},i}(q)}-\frac{L_i q}{\gamma_{\mathrm{II},i}(q)}\right\}, \quad \forall i\in A.$ [Eq. (9)]

c. * $\vec{W}_i(t+\Delta t) = \min\limits_{q\in[0,C_i]}\left\{W_i\left(t+\Delta t-\frac{\tau_i(q)Q_i}{Q_i-q}\right)+\frac{\tau_i(q)q^2}{Q_i-q}\right\}, \quad \forall i\in A^*.$ [Eq. (40)]

d. * $\vec{E}_i(t+\Delta t) = E_i\left(W_i^{-1}\left(\vec{W}_i(t+\Delta t)\right)\right), \quad \forall i\in A^*.$ [Eq. (20)]

Step 4: Sending and receiving flows during time interval $[t, t+\Delta t]$.

a. $S_i(t) = \min\left\{\vec{U}_i(t+\Delta t)-V_i(t), C_i\cdot\Delta t\right\}, \quad \forall i\in A\setminus A^*.$ [Eq. (11)]

b. $R_i(t) = \min\left\{\overleftarrow{V}_i(t+\Delta t)-U_i(t), Q_i\cdot\Delta t\right\}, \quad \forall i\in A.$ [Eq. (11)]

c. * $\tilde{S}_i(t) = \min\left\{\vec{U}_i(t+\Delta t)-W_i(t)+E_i(t), Q_i\cdot\Delta t\right\}, \quad \forall i\in A^*.$ [Eq. (50)]

d. * $S_i(t) = \min\left\{\vec{W}_i(t+\Delta t)-\vec{E}_i(t+\Delta t)-V_i(t), C_i\cdot\Delta t\right\}, \quad \forall i\in A^*.$ [Eq. (37)]

Step 5: Split proportions.

a. $t_i' = U_i^{-1}\left(V_i(t)+S_i(t)\right), \quad \forall i\in A.$

b. $\varphi_{ij}(t) = \frac{\sum_{p\in P}\delta_{jp}\left(U_{ip}(t_i')-V_{ip}(t)\right)}{\sum_{j'\in J_n}\sum_{p\in P}\delta_{j'p}\left(U_{ip}(t_i')-V_{ip}(t)\right)}, \quad \forall i\in I_n, \forall j\in J_n, \forall n\in N.$

Step 6: Transition and auxiliary flows.

a. $\mathbf{D}_n(t) = \Gamma\left(\mathbf{S}_n(t), \mathbf{R}_n(t), \boldsymbol{\varphi}_n(t), \mathbf{C}_n\cdot\Delta t\right), \quad \forall n\in N.$ [Eq. (41)]

b. * $\tilde{D}_i(t) = \min\left\{\tilde{S}_i(t), C_i\cdot\Delta t\right\}, \quad \forall i\in A^*.$ [Eq. (49)]

c. * $\tilde{E}_i(t) = \min\left\{\vec{W}_i(t)-\vec{E}_i(t)-V_i(t), C_i\cdot\Delta t-\tilde{D}_i(t)\right\}, \quad \forall i\in A^*.$ [Eq. (52)]

Step 7: Update realised cumulative flows.

a. $V_i(t+\Delta t) = V_i(t)+\sum_{j\in J}D_{ij}(t), \quad \forall i\in A.$ [Eqs. (6) and (13)]

b. $V_{ip}(t+\Delta t) = U_{ip}\left(U_i^{-1}\left(V_i(t+\Delta t)\right)\right), \quad \forall i\in A, \forall p\in P.$

c. * $E_i(t+\Delta t) = E_i(t)+\tilde{E}_i(t), \quad \forall i\in A^*.$ [Eq. (19)]

d. * $W_i(t+\Delta t) = W_i(t)+\tilde{D}_i(t)+\tilde{E}_i(t), \quad \forall i\in A^*.$ [Eq. (17) and Eq. (51)]

Step 8: Terminate or repeat.

a. Set $t := t+\Delta t$.

b. If $t = T$ then stop. Otherwise, return to Step 2.

5.7 Numerical examples

We provide two examples that illustrate the results of our augmented LTM. The first example considers link-specific traffic control (Case I), while the second example considers movement-specific traffic control (Case II). Given that our main contribution lies in the augmented link model, we focus in each case on a single link. Further, since in congested conditions Webster's uniform delay is fixed at half the red time, we focus on the more interesting case of uncongested conditions.

5.7.1 Example 1 (Case I)

Consider a single link with a saturation flow rate of $Q = 2000$ (veh/h), and link-specific traffic signal control characterised by $c = 100/3600$ (h) and $g = 50/3600$ (h). This means that the exit capacity is $C = 1000$ (veh/h). We assume the following arrival rates: 0 veh/h for 10 seconds, then 200 veh/h for 30 seconds, then 800 veh/h for 40 seconds, after which the flow decreases to 400 veh/h.

We use Eq. (40) for Case I (link-specific non-persistent delay) to determine $\vec{W}(t)$ at a discretised time resolution of $\Delta t = 1/3600$ (h) to obtain cumulative outflow curve as shown in Fig. 8. Fig. 9 shows the corresponding Webster's uniform delay. These graphs clearly show the fanning effect starting at 40 seconds where both the flow rate and the delay gradually increase. After 80 seconds there is a gradual decline in delay.

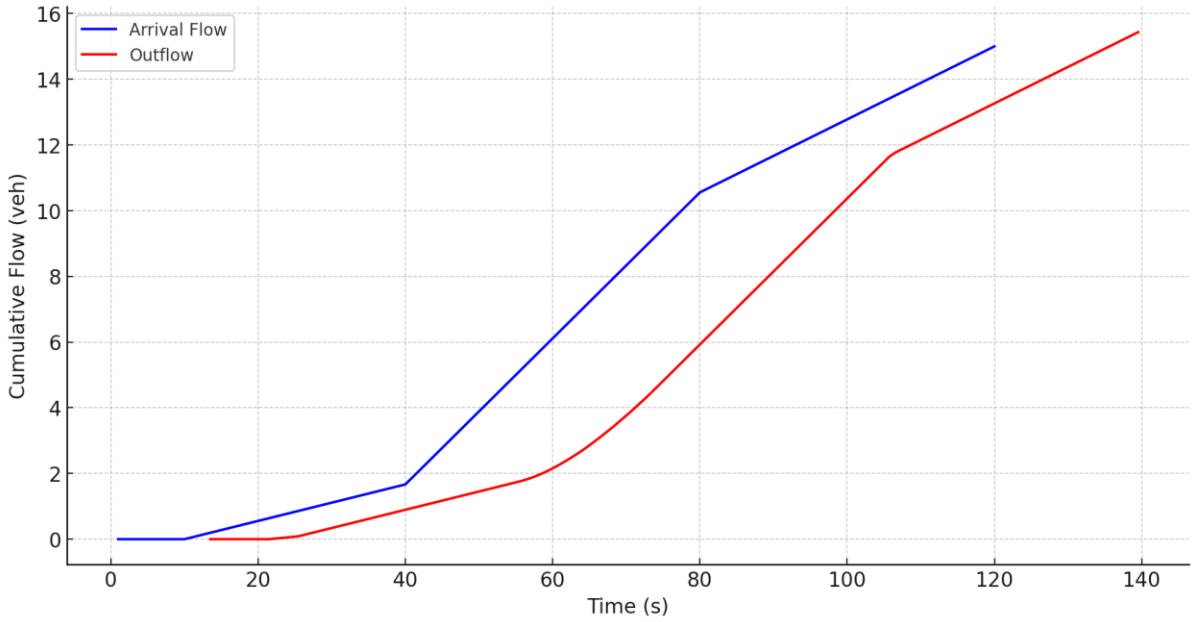


Fig. 8. Cumulative arrival flow and outflow for Example 1 in augmented LTM

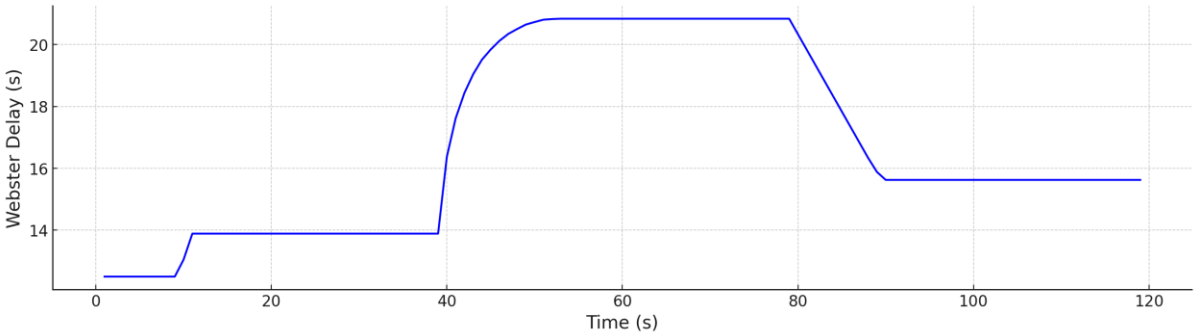


Fig. 9. Webster's uniform delay over time for Example 1 in augmented LTM

In our proposed method, Eq. (40) solves the kinematic wave model to determine exit flows, after which non-persistent delays can be computed as $\tau(t) = \vec{W}^{-1}(W(t)) - t$. An alternative approach is to first compute non-persistent delay $\tau(t)$ using Webster's formula and then apply an exit time-based model to propagate, i.e., $\vec{W}(t + \tau(t)) = W(t)$. However, such an approach is inconsistent with the exit-flow based approach in LTM and results in FIFO violations when flow rates decrease. To illustrate, Fig. 10 depicts the resulting cumulative flows and Fig. 11 shows the resulting delays. The cumulative outflow does not exhibit a similar fanning effect as in Fig. 8; instead, the increase in arrival rate results in flow rate of zero for a period of time before jumping up. Further, the decrease in arrival rate results in an outflow rate that suddenly jumps up due to exiting flows from multiple arrival times. This is caused by FIFO violations whereby the first vehicles arriving at a rate of 400 veh/h exit the link earlier than the last vehicles arriving at a rate of 800 veh/h. As a result, for a short period of time the outflow rate is $800 + 400 = 1200$ veh/h, which is unrealistic as it exceeds the link exit capacity. Fig. 12 shows the change rate of Webster's uniform delay, $d\tau(t)/dt$. As discussed in Appendix F, to ensure non-negative flows and to satisfy FIFO it should hold that $d\tau(t)/dt > -1$. This FIFO condition is clearly violated in the exit-time based approach as can be observed in Fig. 12(b), whereas FIFO is always satisfied in our augmented LTM.

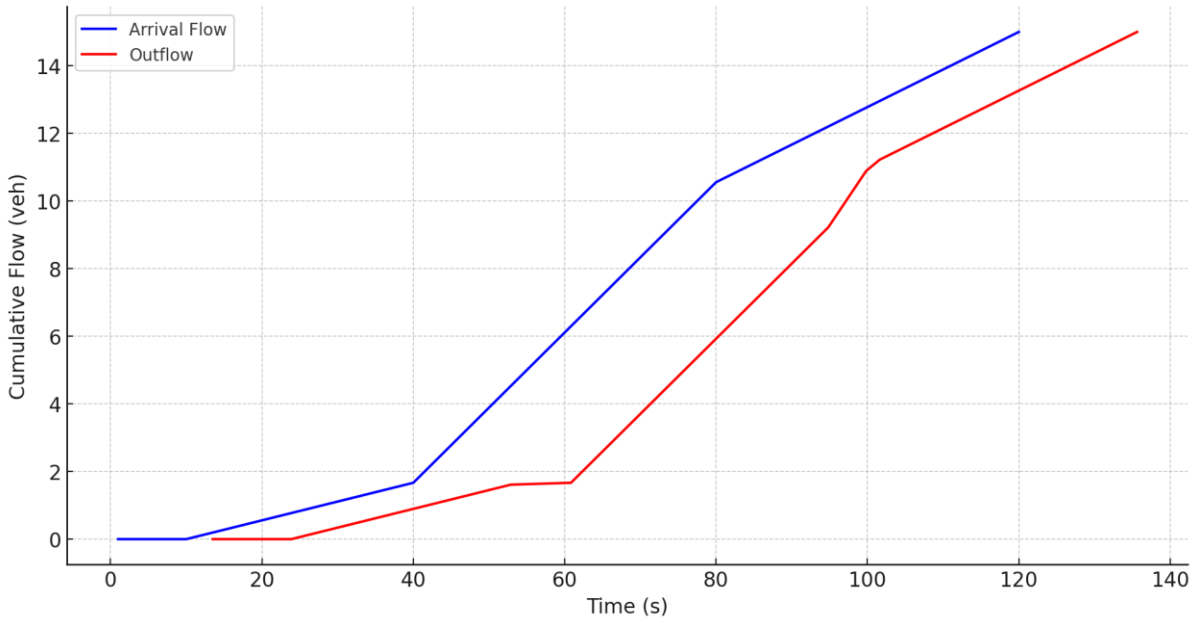


Fig. 10. Cumulative arrival flow and outflow for Example 1 in exit time-based approach

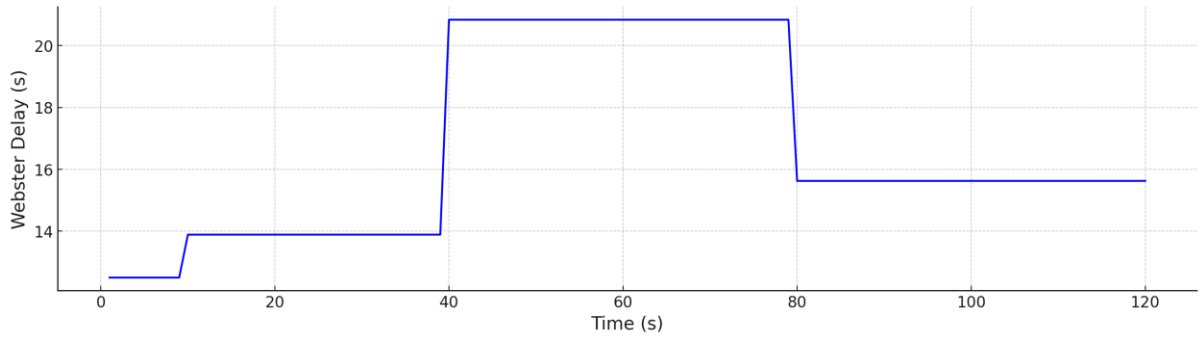
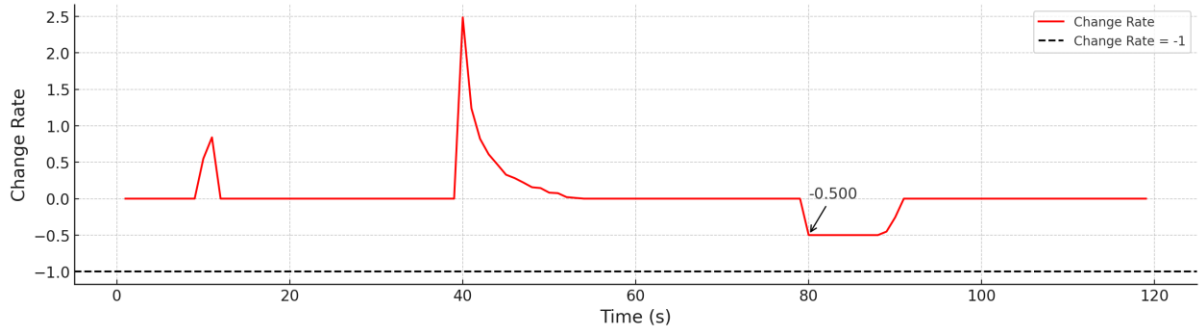
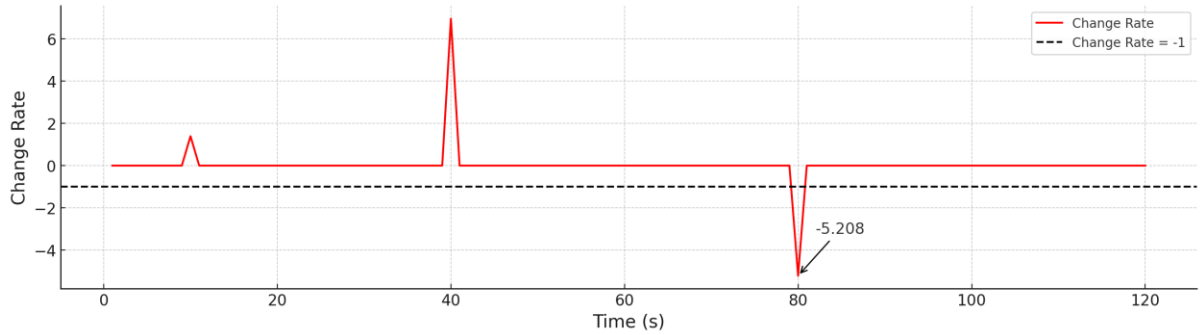


Fig. 11. Webster's uniform delay over time for Example 1 in exit time-based approach



(a) Augmented LTM



(b) Exit time-based approach

Fig. 12. The change rate of Webster's uniform delay

5.7.2 Example 2 (Case II)

Consider the three-legged signalised intersection depicted in Fig. 13, where we focus on link 1 that has a dedicated left and right turn. We assume that the saturation flow for the left and right are $Q_{14} = 800$ (veh/h) and $Q_{15} = 1200$ (veh/h), respectively. Further, we assume that the signal control has a cycle length of $c = 120/3600$ (h) and that the effective green time per cycle for left and right turns is $g_{14} = 60/3600$ (h) and $g_{15} = 70/3600$ (h), respectively. This yields movement-specific exit capacities $C_{14} = \frac{60}{120} \cdot 800 = 400$ (veh/h) and $C_{15} = \frac{70}{120} \cdot 1200 = 700$

(veh/h). For link 1 we assume an arrival rate of 800 (veh/h), which drops to 400 (veh/h) after 140 seconds. Further, we assume fixed split proportions of $\varphi_{14} = 0.4$ (for left turns) and $\varphi_{15} = 0.6$ (for right turns). This means that $w_{14}(t) = 0.4 \cdot 800 = 320$ (veh/h) and $w_{15}(t) = 0.4 \cdot 800 = 480$ (veh/h) for the first 140 seconds, and $w_{14}(t) = 160$ (veh/h) and $w_{15}(t) = 240$ (veh/h) after that. Fig. 14 depicts the corresponding cumulative arrival and realised outflow curves as well as the non-persistent delay.

We now determine the non-persistent delays for the left and right turns separately (Case II). In congested conditions, Webster's uniform delay is half the red time, so for left turns this is 30 seconds, namely $\frac{1}{2}(c - g_{14}) = 30/3600$ (h), and for right turns this is 25 seconds, namely $\frac{1}{2}(c - g_{15}) = 25/3600$. In this example we again focus on uncongested situations in which Webster's uniform delay varies. Using Webster's uniform delay formula, we would expect the following non-persistent delays. For left turning vehicles, 25 seconds delay will be experienced initially, and after 140 seconds this becomes 18.75 seconds. Right turning vehicles should initially experience 17.36 seconds of delay, and 13.02 seconds delay after 140 seconds.

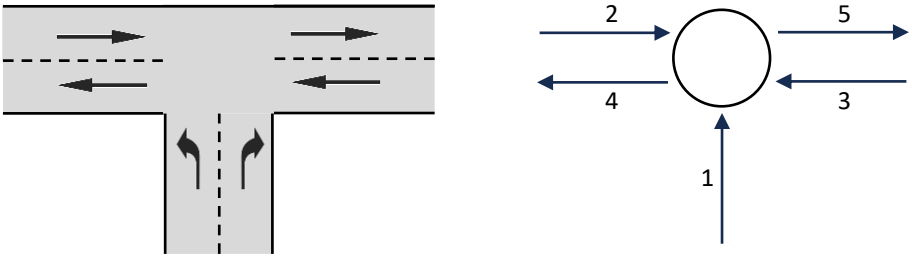
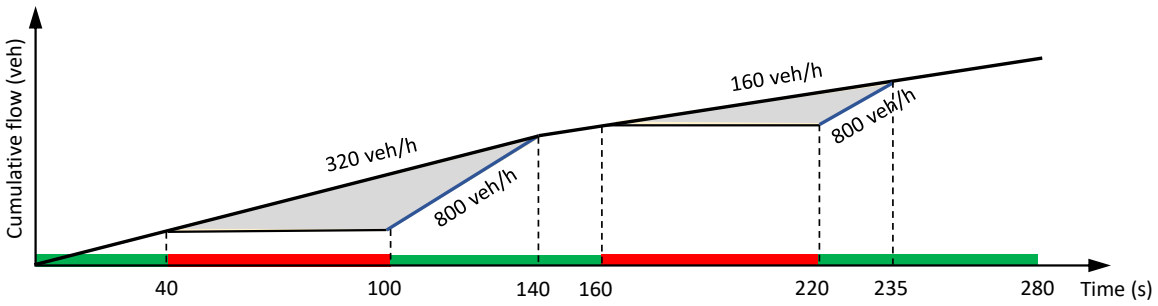
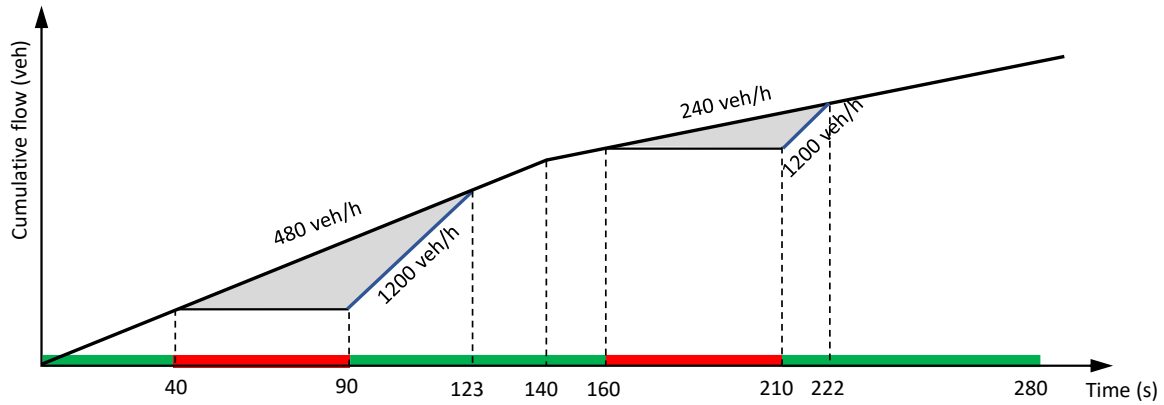


Fig. 13. Three-legged signalised intersection for Example 2



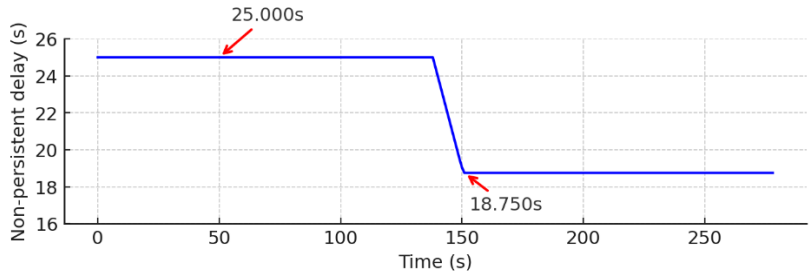
(a) Left turn



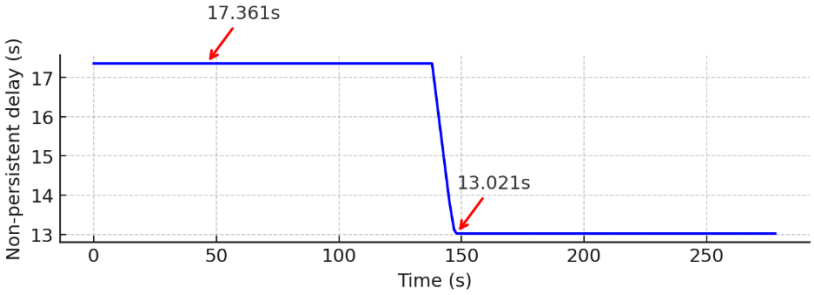
(b) Right turn

Fig. 14. Cumulative arrival and outflow curves for Example 2 with explicit red and green intervals

We use Eq. (45) in our augmented LTM to determine $\vec{W}_{ij}(t)$ for movements $ij=14$ and 15 at a discretised time resolution of $\Delta t = 1/3600$ (h). Webster's uniform delay can be calculated as $\tau_{ij}(t) = \vec{W}_{ij}^{-1}\left(W_{ij}(t)\right) - t$ for movements $ij=14$ and 15. Fig. 15 shows this non-persistent delay, which confirms consistency with the delays calculated using Webster's formula above. These graphs also again show is that our method gradually decreases non-persistent delay after 140 seconds, which ensures that FIFO is satisfied.



(a) Left turn



(b) Right turn

Fig. 15. Non-persistent delay determined by augmented LTM for Example 2

5.8 Conclusions

In this paper, we have proposed an innovative method that integrates non-persistent delay into the link transmission model. This is achieved via an augmented link representation for signalised intersections whereby Webster's uniform delay is represented on a virtual link governed by a specially derived fundamental diagram. We presented two alternative formulations, namely Case I where the traffic signal controls all movements on the link, and Case II whereby each movement is controlled by a separate traffic signal. We further indicated where adaptations of existing time-discretised algorithms are required to solve our augmented LTM on general transport networks without the need to explicitly modify the network. Using two numerical examples we demonstrated how resulting traffic flows are consistent with Webster's uniform delay and comply with FIFO. The adoption of this method leads to a more accurate representation of traffic flow dynamics in the case of signalised intersections, thereby improving the precision of traffic flow and travel time forecasts that will benefit transport planning and traffic management.

Our proposed methodology requires basic information regarding traffic signal control, namely cycle lengths and effective green times per link or movement to determine exit capacities, and possible information with respect to the number of approach lanes and lane layout. Lane-specific exit capacities as discussed in Case III in Section 5.4 could possibly be considered in conjunction with the lane-based node model proposed in Gong et al. (2024) but further research is needed to confirm feasibility. If the phase design is also known, one may be able to further refine the methodology, although if a more detailed simulation of signalised intersections is required then one might consider micro-simulation instead.

This study assumed only a single vehicle class. In the presence of multiple vehicle types, such as cars and trucks, a multiclass extension of LTM would be required. With respect to average uniform delay, a likely reasonable approximation would be to convert multiclass flow into single class flow by converting trucks and other vehicle types into passenger car units (pcu's) based on the amount of exit capacity they consume.

6 Chapter 6 Conclusions

This chapter summarises the findings of the thesis (Section 6.1) and the overall contributions of the study (Section 6.2). It also discusses the research limitations and presents ideas for future studies (Section 6.3).

6.1 Summary

This thesis focuses on improving the prediction of both persistent and non-persistent delays in macroscopic intersection modelling. Chapters 2, 3, and 4 examine persistent delay forecasting. These three chapters cover the advancements and emerging trends in node model development and extended modelling frameworks and solution algorithms for existing node models. Non-persistent delay forecasting is explored in Chapter 5, where a novel mathematical method for incorporating non-persistent delays within the LTM framework is developed. The following paragraphs provide a summary of these four main chapters (Chapters 2 to 5).

Chapter 2

Chapter 2 presents an in-depth systematic review on the advancements and emerging trends in node model research. The chapter proposes six general principles and five extension categories that characterise additional features. The findings are based on the review of 48 papers on node models. Building on these findings, Chapter 2 classifies the models according to the identified characteristics and presents a comprehensive classification table. The table serves as the principal output of this chapter.

Chapter 2 also presents several other key findings. It highlights that recent research has often contributed by exploring one or more of the extension categories, with an emphasis on integrating signal control and assessing the impact of intersection layout. These areas are particularly promising for further investigation. Additionally, the incorporation of stochastic elements within node models has become a notable theme, offering insights into the variability and unpredictability of traffic flows, even within a macroscopic framework.

Finally, this chapter identifies future research opportunities in node modelling. The first recommendation is to investigate methods that accurately, efficiently, and robustly address conflicts within node models. The second suggestion is to explore a more integrated approach that incorporates multiple extension aspects simultaneously. The third is to define flow maximisation in terms of the presumed underlying behaviour in specific studies, particularly how flow maximisation is achieved under specific conditions. The last recommendation is to consider at what point the inclusion of micro-level details would provide more benefits than the convenience and broad applicability of macroscopic modelling.

Chapter 3

Chapter 3 introduces a novel lane-based node model. The model explicitly considers the configuration of approach lanes while relaxing link-based FIFO to movement-based FIFO in scenarios where nodes have multiple approach lanes. Rather than explicitly modelling each approach lane in the network, the proposed model presents a novel lane choice equilibrium principle, within an implicitly lane-expanded node representation, based on which vehicles are allocated to lanes as they exit the link. The problem is formalised as a variational inequality

problem. Theoretical proofs confirm the existence and uniqueness of the solution. Furthermore, this chapter develops an iterative solution algorithm to address the extended node model. An analysis of the algorithm's performance indicates that the proposed algorithm solves small systems of linear equations with minimal computation time. Additionally, in real-life scenarios, most nodes can be resolved within limited iterations. Only when multiple shared movements become congested are more iterations required to achieve convergence, and such scenarios are uncommon in real-life networks.

Chapter 3 also provides and compares a numerical example of the link-based node model adapted from Tampère et al. (2011) to demonstrate the proposed node model's viability and utility. The results illustrate the algorithm's applicability. The findings demonstrate that the lane configurations in the lane-based node model help drivers to avoid congested movements by switching lanes when possible. Specifically, the queue length and queuing delay for specific movements on congested links in the link-based node model tend to be both overestimated and underestimated. For uncongested movements or scenarios where lane-switching options are unavailable, the results are either identical or similar, which was expected based on the theory.

Chapter 4

Chapter 4 proposes a novel modelling approach for macroscopic node models. This new approach accounts for exit capacities by endogenously adjusting the flow rates, without explicitly changing the node model. The proposed generic solution scheme consists of three stages: pre-processing, node modelling, and post-processing. The chapter introduces the term 'saturation movement unit' (smu), which is a play on 'passenger car unit' (pcu), a measure that quantifies the impact on capacity caused by vehicles making a specific movement. Converting the flow into smu helps to adjust the flow by scaling up inflow rates based on the specific turn speed of each movement. This method accounts for so-called 'dead capacity' during the pre-processing stage and scales down outflow rates in the post-processing stage. A numerical example adapted from Tampère et al. (2011) is presented to illustrate the calculation process of the proposed algorithm and validate the method's effectiveness and reliability. The findings indicate that when different movement capacities are considered, the total number of vehicles that can exit the node model drops by almost 16%, compared to the original Tampère et al. (2011) model with fixed capacities. This finding indicates that node models with exogenous capacities tend to overestimate the outflow on a link. A second numerical example illustrates that the node model with exogenous capacity can also underestimate the outflow on a link.

Chapter 5

Chapter 5 develops an augmented LTM that successfully integrates non-persistent delays at intersections controlled by signals. The chapter proposes that one or more virtual links should be incorporated implicitly into the representation of the link model. This involves reformulating Webster's uniform delay as a hypothetical branch of the associated fundamental diagram. Existing algorithms that solve the general LTM can be applied to this augmented network. Two numerical examples are provided to validate the feasibility and effectiveness of the proposed model. The first example compares the proposed model with a method called the naïve method, which appends Webster's delay directly via a flow shift. The results show that the naïve method violates the FIFO principle when flow decreases, whereas the proposed method adheres to the FIFO principle when flow rates change. Additionally, the results reveal that when flow increases, the naïve method exhibits an immediate surge in Webster's delay, indicating an unrealistic instantaneous impact. By contrast, this impact does not occur in the proposed model.

The second example compares the proposed LTM-based DNL model with a method that explicitly simulates signal control. The results show that there is consistency and congruence between the two approaches. The proposed model is more realistic in simulating real-world traffic dynamics, as evidenced by these findings.

6.2 Contributions

This thesis offers numerous theoretical and practical contributions that are based on rigorous research, analysis, and discussion. A summary of these contributions is presented in this section. The theoretical contributions fall within three main areas: a literature review, forecasting persistent delays, and forecasting non-persistent delays. They can be summarised as follows.

1) The thesis presents a systematic review of the literature on node models in the context of macroscopic NL. It updates the general principles proposed by Tampère et al. (2011), which are commonly adopted in contemporary node model research. In addition, the thesis proposes five categories of extensions to the node model and consolidates current knowledge on the development and application of node models within DNL frameworks. Specifically, it (a) identifies the fundamental characteristics of generic node models, (b) examines the methodologies and technologies employed in these models, (c) evaluates the ability of various node models to accurately replicate actual traffic flows, and (d) highlights emerging trends and prospective areas for future research in the field. This work offers valuable insights to researchers, policymakers, and traffic engineers engaged in urban traffic network planning and management. To the authors' knowledge, this study is the first to provide a comprehensive systematic review of macroscopic node models.

2) The thesis presents a formulation of a novel lane-based node model designed to relax link-based FIFO for nodes with multiple lanes. This method introduces an innovative principle for achieving equilibrium in the distribution of sending flows across lanes. The principle considers the approach lane configuration and ensures movement-based FIFO, formalised through a variational inequality problem formulation. Additionally, a general solution algorithm to solve the extended node model is developed. Numerical examples and a case study are provided to demonstrate the viability and utility of the proposed model. In the novel node model, no explicit approach lanes are created in the network, which avoids the exponential growth of routes. The proposed lane-based node model effectively captures the impacts of lane-specific phenomena on traffic flow. Prior link-based node models do not fully capture these impacts, which means the new model enhances the forecasting of persistent delays. To the authors' knowledge, the lane-based node model is the first in the literature to consider approach lane configuration while adhering to the movement-based FIFO principle.

3) This thesis introduces an approach to the node model that accounts for endogenous exit capacities, which vary according to the composition of traffic flows. An efficient algorithm to address this issue without the need to dynamically alter the exit capacity or the node model is presented. Additionally, the thesis provides numerical examples to demonstrate the feasibility and effectiveness of the proposed model's outcomes. This approach underscores the significant impact of endogenous exit capacities on traffic flow. The thesis also highlights the need to consider dead capacity in traffic flow models, as the model imposes constraints on the link exit flow. The proposed method improves the accuracy of

predicting persistent delays in node models by considering exit capacities for different turning movements; this offers a more refined and precise technique for selecting routes within the context of the transportation network. To the authors' knowledge, this work is the first to address endogenous exit capacity in node models by adjusting flow rates without requiring modifications to the node model itself.

4) The thesis proposes a novel approach that integrates non-persistent delays within the LTM framework. It augments the LTM with virtual links and constructs a hypothetical fundamental diagram that aligns with Webster's uniform delay. Existing algorithms that solve the general LTM can be applied to this augmented network, and adaptations to existing time-discretised algorithms are discussed. To demonstrate the feasibility and effectiveness of the proposed model, two numerical examples are presented. The results indicate that accuracy in capturing variations in non-persistent delays due to changing flow rates is enhanced. The proposed model thus offers a more refined and precise technique for route selection in the context of DNL.

This thesis focuses mainly on methodology rather than applications. Nonetheless, it makes several practical contributions, which are summarised as follows.

1) The presented approach and models can be incorporated into existing macroscopic NL models that are currently utilised in practice, such as OmniTRANS, VISUM, and AIMSUN. This integration enhances the capacity of the models to accurately describe intersection delays and flows, both at traditional intersections and at freeway-to-freeway or freeway-to-arterial connections. This thesis offers an efficient and reliable method for improving traffic projections, which is crucial for making informed decisions in domains such as infrastructure investment appraisal and traffic management. The work is thus highly relevant for governments, researchers, and consultants.

2) The methodology presented in this thesis can also be utilised to optimise traffic signal control throughout a network. Through using refined intersection models, traffic signal timings can be adjusted dynamically to minimise congestion and improve the traffic flow efficiency. This optimisation is valuable for urban planners and traffic engineers who seek to implement adaptive signal control systems. Real-time traffic systems respond quickly to changing traffic conditions, thus reducing travel time, fuel consumption, and emissions. Overall, the advancements proposed in this thesis can improve the performance and sustainability of traffic management systems.

6.3 Limitations and future research directions

The thesis has several limitations, and these issues point to directions for potential future research. They are summarised below.

1) The systematic review in Chapter 2 highlights avenues for future research on node models. These include (a) investigating an approach in which conflicts in node models can be considered efficiently; (b) exploring a more integrated approach that would simultaneously incorporate multiple extension aspects, such as multiclass vehicle, approach lane configuration, and signal control; (c) defining a flow maximisation principle in terms of the assumed behaviour in each specific study; (d) identifying the threshold at which the benefits

of incorporating micro-level details surpass the convenience and broad applicability of macroscopic modelling.

2) The proposed lane-based node model in Chapter 3 excludes the explicit consideration of traffic signal control, priority rules, and interactions with other road users (such as pedestrians). Future research could expand the lane-based node model to incorporate these factors. For example, the model could embed the methodology of Yahyamozdarani and Tampère (2023) to account for traffic signal control.

3) In order to further develop the lane-based node model, studies could improve lane-based node model by incorporating the method introduced in Chapter 4 to include the endogenous exit capacities.

4) The proposed methods and models in Chapters 3, 4, and 5 all focus on a single link or node. This approach is insufficient for demonstrating the effectiveness and feasibility of the model at the network level. Future research could investigate the implications of the proposed method or models within a broader network framework. Such investigations would require obtaining and analysing empirical data, with the aim of providing robust validation and a deeper understanding of the model's effectiveness in simulating real-world traffic scenarios.

Appendix A: Lane-based node model algorithm

The algorithm employed for resolving the link-based node model was first introduced in Tampère et al. (2011). A compact version of this algorithm is included in Bliemer et al. (2014). This latter version is adapted to this work's notation and further streamlined. In this form it is presented here for the benefit of the reader in the context of the lane-based node model, recall Eq. (2). Let X and $\bar{A}^{\text{out}}$ denote the sets of unprocessed lanes and outgoing links, respectively. The subsequent algorithm elucidates the iterative methodology utilised for determining the movement-specific lane outflows and the set of supply constrained lanes of the node.

Algorithm 2. Lane-based node model.

Input: Movement-specific lane sending flows $\mathbf{s}$, receiving flows $\mathbf{r}^{\bullet}$, lane exit capacities $\mathbf{c}$.

Output: Movement-specific lane outflows $\mathbf{v}$, set of supply constrained lanes E.

Step 1: *Initialisation*

Set remaining receiving flows $r'_j := r^{\bullet}_j, \forall j = 1, \dots J.$

Set unprocessed lane set $X := \{1, \dots, M\},$ unprocessed outlink set $\bar{A}^{out} := \{1, \dots, J\},$ and supply constrained lanes $Z := \varnothing$.

Then, per lane $m = 1, \dots, M,$ do

Determine lane sending flows s_m using Eq. (3).

Determine the capacity scaling factors $\alpha_m = c_m / s_m.$

Set initial lane reduction factor $\gamma_m = 1.$

Step 2: *Determine current most restricting outgoing link.*

Find $\bar{j} = \arg\min_{j \in \bar{A}^{\text{out}}} \left\{ r'_j \Big/ \sum_{m \in X} \alpha_m s_{mj} \right\}, \quad \forall m = 1, \dots, M.$

Step 3: *Identify demand/capacity constrained affected lanes.*

Determine capacity scaled reduction factor $\beta := r'_{\bar{j}} \Big/ \sum_{m \in X} \alpha_m s_{m\bar{j}}, \ \forall m = 1, \dots, M.$

Determine demand constrained affected lane set $Y := \{m \in X \mid s_{m\bar{j}} > 0, \beta\alpha_m \geq 1\}.$

If $Y = \varnothing,$ then capacity constrained lane $Z := \{m \in X \mid s_{m\bar{j}} > 0\},$ otherwise $Z := \varnothing.$

Step 4: *Determine realised lane sending flows*

Determine lane reduction factors for supply constrained lanes via $\gamma_m = \beta\alpha_m, \quad m \in Z.$

Determine realised lane-movement outflows when $\eta_{mj} = 1$ via $v_{mj} = \gamma_m s_{mj},$ $j = 1, \dots, J.$

Step 5: *Prepare for next iteration or terminate*

Mark iteration affected lanes as processed $X := X \setminus \{Y \cup Z\}.$

Track any found supply constructed lanes for final output $E := E \cup Z,$

If $X = \varnothing$, then terminate and return $\mathbf{v}$ and E, otherwise,

Set remaining receiving flows $r' := r' - \sum_{m \in \{Y \cup Z\}} \gamma_m s_{mj}, \quad \forall j \in \bar{A}^{\text{out}}.$

Mark current most restricting outgoing link as processed via $\bar{A}^{\text{out}} := \bar{A}^{\text{out}} \setminus \{\bar{j}\}.$

Return to Step 2.

Appendix B: Proof of Proposition 1

First, we prove "necessity", namely that equilibrium conditions (12) imply VI problem (13). These conditions can be written as

$$\delta_{im}(h_m - \bar{h}_{ij})\bar{s}_{mj} = 0, \qquad \forall i = 1,\ldots,I; \forall j = 1,\ldots,J; \forall m = 1,\ldots,M. \tag{B.1}$$

By definition it holds that $h_m - \bar{h}_{ij} \geq 0$ and $s_{mj} \geq 0,$ hence the following inequality should hold:

$$\delta_{im}(h_m - \bar{h}_{ij})s_{mj} \geq 0, \qquad \forall i = 1,\ldots,I; \forall j = 1,\ldots,J; \forall m = 1,\ldots,M. \tag{B.2}$$

Subtracting (B.1) from (B.2) yields

$$\delta_{im}(h_m - \bar{h}_{ij})(s_{mj} - \bar{s}_{mj}) \geq 0, \qquad \forall i = 1,\ldots,I; \forall j = 1,\ldots,J; \forall m = 1,\ldots,M. \tag{B.3}$$

Summing this inequality over all indices yields

$$\sum_i \sum_j \sum_m \delta_{im} h_m (s_{mj} - \bar{s}_{mj}) - \sum_i \sum_j \bar{h}_{ij} \sum_m \delta_{im}(s_{mj} - \bar{s}_{mj}) \geq 0. \tag{B.4}$$

Since $\sum_m \delta_{im} s_{mj} = \sum_m \delta_{im} \bar{s}_{mj} = s_{ij}^{\bullet}$ for each movement $(i, j),$ the second term vanishes. Since this inequality is satisfied for any $\mathbf{s} \in S,$ VI problem (13) results.

Next we prove "sufficiency", namely that a solution to VI problem (13) implies equilibrium conditions (12). Consider a specific movement (i, j) and all available lanes $m \in L_{ij}$ for this movement. Consider equilibrium sending flows $\bar{\mathbf{s}}$ and consider another feasible solution $\mathbf{s}$ that is identical to the equilibrium solution except for lanes m_1 and m_2 on incoming link i (i.e., $\delta_{im_1} = \delta_{im_2} = 1$) that allow movement in direction j. Without loss of generality, we consider two situations that could arise while at equilibrium.

In situation 1, $\bar{s}_{m_1 j} > 0$ and $\bar{s}_{m_2 j} > 0,$ i.e., both lanes are used for movement (i, j). Switching a small percentage of flow Δ_1 from lane m_1 to m_2 with $0 < \Delta_1 \leq \bar{s}_{m_1 j}$ results in $s_{m_1 j} = \bar{s}_{m_1 j} + \Delta_1$ and $s_{m_2 j} = \bar{s}_{m_2 j} + \Delta_1.$ Substituting this new feasible solution in inequality (13) yields

$$\bar{h}_{m_1}(s_{m_1 j} - \bar{s}_{m_1 j}) + \bar{h}_{m_2}(s_{m_2 j} - \bar{s}_{m_2 j}) = \Delta_1\left(\bar{h}_{m_2} - \bar{h}_{m_1}\right) \geq 0. \tag{B.5}$$

Similarly, by switching a small percentage of flow Δ_2 with $0 < \Delta_2 \leq \bar{s}_{m_2 j}$ from lane m_2 to m_1 , we obtain:

$$\Delta_2\left(\bar{h}_{m_1} - \bar{h}_{m_2}\right) \geq 0. \tag{B.6}$$

Equations (B.5) and (B.6) together imply that $\bar{h}_{m_1} = \bar{h}_{m_2}$. We can repeat this procedure for each lane and direction, and as a result all used lanes (with positive sending flow) for the same movement have the same travel time. This confirms the first part of equilibrium conditions (12).

In situation 2, $\bar{s}_{m_1 j} > 0$ and $\bar{s}_{m_2 j} = 0.$ Switching a small percentage of flow Δ_1 from m_1 to m_2 with $0 < \Delta_1 \leq \bar{s}_{m_1 j}$ results in $\Delta_1(\bar{h}_{m_2} - \bar{h}_{m_1}) \geq 0.$ Repeating this procedure to verify that for each

incoming and outgoing link pair, all unused lanes (with zero sending flow) have a travel time not lower than the minimal travel time. This confirms the second part of equilibrium conditions (12).

This completes the proof.

Appendix C: Iteration process in case study

This section details the iterative process for the case study in Section 3.8, with $\varepsilon = 10^{-10}$. To reduce repetitive content in the presentation of the results, we only present the initial, second, and the final (14th) iteration on unconverged links.

The link-based sending and receiving flow inputs are given in Table 4, while approach lane configuration can be derived from Fig. 9(b) and is shown in Table C.1, together with the lane exit capacities that are all considered to be 1000 veh/h. The gap is set to threshold $\varepsilon = 10^{-10}$.

Iteration $k = 1$

Step 1
The set of supply constrained lanes is initialised as empty in Step 1, i.e., $E = \varnothing$.

Step 2
Gradients **g** are calculated using Eq. (20) as $g_1 = g_2 = \ldots = g_{12} = \frac{1}{1000}$. The adjusted approach lane configuration is initialised as $\tilde{\boldsymbol{\eta}} := \boldsymbol{\eta}$. For link 1, we solve the system of linear equations (where all other movement-specific lane sending flows are set to zero), yielding

$$\begin{pmatrix} s_{1,7}^{(1)} \\ s_{1,8}^{(1)} \\ s_{2,6}^{(1)} \\ s_{2,7}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{1000} & \frac{1}{1000} & -\frac{1}{1000} & -\frac{1}{1000} \end{pmatrix}^{-1} \begin{pmatrix} 300 \\ 600 \\ 100 \\ 0 \end{pmatrix} = \begin{pmatrix} -100 \\ 600 \\ 100 \\ 400 \end{pmatrix}.$$

Table C.1. Approach lane configuration indicators and inlink lane capacities.

i	j	1	2	3	4	
	m	η_{mj}				c_m
1	1	0	0	1	1	1000
	2	0	1	1	0	1000
2	3	1	0	0	1	1000
	4	0	0	0	1	1000
	5	0	0	0	1	1000
	6	0	0	1	1	1000
3	7	1	1	0	1	1000
	8	0	0	0	1	1000
4	9	0	0	1	0	1000
	10	0	1	1	0	1000
	11	0	1	0	0	1000
	12	1	1	0	0	1000

Since $\min(\mathbf{s}_1^{(1)}) = s_{17}^{(1)} < 0$, we set $\tilde{\eta}_{17} = 0$ such $s_{17}^{(1)} = 0$ and we update and solve the system of linear equations again, which results in all non-negative flows,

$$\begin{pmatrix} s_{18}^{(1)} \\ s_{26}^{(1)} \\ s_{27}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 600 \\ 100 \\ 300 \end{pmatrix} = \begin{pmatrix} 600 \\ 100 \\ 300 \end{pmatrix}.$$

For links 2, 3 and 4 we also solve the system of linear equations, which do not result in any negative flows,

$$\begin{pmatrix} s_{3,5}^{(1)} \\ s_{3,8}^{(1)} \\ s_{4,8}^{(1)} \\ s_{5,8}^{(1)} \\ s_{6,7}^{(1)} \\ s_{6,8}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ \frac{1}{1000} & \frac{1}{1000} & -\frac{1}{1000} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{1000} & -\frac{1}{1000} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{1000} & -\frac{1}{1000} & -\frac{1}{1000} \end{pmatrix}^{-1} \begin{pmatrix} 200 \\ 600 \\ 3200 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 200 \\ 800 \\ 1000 \\ 1000 \\ 600 \\ 400 \end{pmatrix},$$

$$\begin{pmatrix} s_{7,5}^{(1)} \\ s_{7,6}^{(1)} \\ s_{7,8}^{(1)} \\ s_{8,8}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \frac{1}{1000} & \frac{1}{1000} & \frac{1}{1000} & -\frac{1}{1000} \end{pmatrix}^{-1} \begin{pmatrix} 200 \\ 200 \\ 1200 \\ 0 \end{pmatrix} = \begin{pmatrix} 200 \\ 200 \\ 400 \\ 800 \end{pmatrix},$$

$$\begin{pmatrix} s_{9,7}^{(1)} \\ s_{10,6}^{(1)} \\ s_{10,7}^{(1)} \\ s_{11,6}^{(1)} \\ s_{12,5}^{(1)} \\ s_{12,6}^{(1)} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{1000} & -\frac{1}{1000} & -\frac{1}{1000} & 0 & 0 & 0 \\ 0 & \frac{1}{1000} & \frac{1}{1000} & -\frac{1}{1000} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{1000} & -\frac{1}{1000} & -\frac{1}{1000} \end{pmatrix}^{-1} \begin{pmatrix} 200 \\ 1600 \\ 1600 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 850 \\ 100 \\ 750 \\ 850 \\ 200 \\ 650 \end{pmatrix}.$$

The resulting movement-specific lane sending flows **s** after one iteration are shown in Table C.2.

Step 3

By applying the lane-based node model, we obtain the movement-specific lane outflows **v** as summarised in Table C.2, with an updated set of supply constrained lanes, $E = \{3,4,5,6,7,8,9,10\}$. The generalised lane exit costs **h** are computed using Eq. (9) and are shown in Table C.2.

Table C.2. Sending flow distribution after iteration 1.

	j	1	2	3	4		1	2	3	4	
i	m	$s_{mj}^{(1)}$				c_m	$v_{mj}^{(1)}$				$h_m^{(1)}$
1	1	0	0	0	600	1000	0	0	0	600	0.600
1	2	0	100	300	0	1000	0	100	300	0	0.400
2	3	200	0	0	800	1000	145	0	0	582	1.376
2	4	0	0	0	1000	1000	0	0	0	727	1.376
2	5	0	0	0	1000	1000	0	0	0	727	1.376
2	6	0	0	600	400	1000	0	0	411	274	1.460
3	7	200	200	0	400	1000	182	182	0	363	1.100
3	8	0	0	0	800	1000	0	0	0	727	1.100
4	9	0	0	850	0	1000	0	0	685	0	1.241
4	10	0	100	750	0	1000	0	81	604	0	1.241
4	11	0	850	0	0	1000	0	850	0	0	0.850
4	12	200	650	0	0	1000	200	650	0	0	0.850
	$r_j^{\bullet}$	2000	4000	2000	4000						

Step 4

The gap at the end of the first iteration is $G^{(1)} \approx 1.722e-2 > \varepsilon$, hence we return to Step 2. Table C.2 indicates that the sending flow distribution on links 1 and 3 are in equilibrium after the first iteration. Since the movement-specific lane sending flows for these two incoming links will not change in consecutive iterations, we only present the further calculations for incoming links 2 and 4 where sending flows are not yet in equilibrium.

Iteration $k=2$

Step 2

Gradients $\mathbf{g}$ are updated using Eq. (20) as $g_3 = g_4 = g_5 = g_7 = g_8 = \frac{1}{727}$, $g_6 = g_9 = g_{10} = \frac{1}{685}$, while all other gradients are equal to $\frac{1}{1000}$. The adjusted approach lane configuration is initialised again as $\tilde{\boldsymbol{\eta}} := \boldsymbol{\eta}$. For link 2, we solve the following system of linear equations,

$$\begin{pmatrix} s_{3,5}^{(2)} \\ s_{3,8}^{(2)} \\ s_{4,8}^{(2)} \\ s_{5,8}^{(2)} \\ s_{6,7}^{(2)} \\ s_{6,8}^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ \frac{1}{727} & \frac{1}{727} & -\frac{1}{727} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{727} & -\frac{1}{727} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{727} & -\frac{1}{685} & -\frac{1}{685} \end{pmatrix}^{-1} \begin{pmatrix} 200 \\ 600 \\ 3200 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 200 \\ 815 \\ 1015 \\ 1015 \\ 600 \\ 355 \end{pmatrix}.$$

For link 4, we solve

$$\begin{pmatrix} s_{9,7}^{(2)} \\ s_{10,6}^{(2)} \\ s_{10,7}^{(2)} \\ s_{11,6}^{(2)} \\ s_{12,5}^{(2)} \\ s_{12,6}^{(2)} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{685} & -\frac{1}{685} & -\frac{1}{685} & 0 & 0 & 0 \\ 0 & \frac{1}{685} & \frac{1}{685} & -\frac{1}{1000} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{1000} & -\frac{1}{1000} & -\frac{1}{1000} \end{pmatrix}^{-1} \begin{pmatrix} 200 \\ 1600 \\ 1600 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 759 \\ -83 \\ 841 \\ 941 \\ 200 \\ 741 \end{pmatrix}.$$

Since $\min(\mathbf{s}_4^{(2)}) = s_{10,6}^{(2)} < 0$, we set $\tilde{\eta}_{10,6} = 0$, and solve a new system of linear equations,

$$\begin{pmatrix} s_{9,7}^{(2)} \\ s_{10,7}^{(2)} \\ s_{11,6}^{(2)} \\ s_{12,5}^{(2)} \\ s_{12,6}^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{685} & -\frac{1}{685} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{1000} & -\frac{1}{1000} & -\frac{1}{1000} \end{pmatrix}^{-1} \begin{pmatrix} 1600 \\ 1600 \\ 200 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 800 \\ 800 \\ 900 \\ 700 \\ 200 \end{pmatrix}.$$

The resulting movement-specific lane sending flows **s** after two iterations are shown in Table C.3.

Step 3

Applying the lane-based node model results in updated movement-specific lane outflows **v** as summarised in Table C.3, with the same set of supply constrained lanes. The updated generalised lane exit costs **h** are computed using Eq. (9) and are shown in Table C.3.

Table C.3. Sending flow distribution after iteration 2.

	j	1	2	3	4		1	2	3	4	
i	m	$s_{mj}^{(2)}$				c_m	$v_{mj}^{(2)}$				$h_m^{(2)}$
1	1	0	0	0	600	1000	0	0	0	600	0.600
1	2	0	100	300	0	1000	0	100	300	0	0.400
2	3	200	0	0	815	1000	145	0	0	589	1.382
2	4	0	0	0	1015	1000	0	0	0	734	1.382
2	5	0	0	0	1015	1000	0	0	0	734	1.382
2	6	0	0	600	355	1000	0	0	406	241	1.478
3	7	200	200	0	400	1000	184	184	0	367	1.090
3	8	0	0	0	800	1000	0	0	0	734	1.090
4	9	0	0	800	0	1000	0	0	647	0	1.237
4	10	0	0	800	0	1000	0	0	647	0	1.237
4	11	0	900	0	0	1000	0	900	0	0	0.900
4	12	200	700	0	0	1000	200	700	0	0	0.900
	$r_j^{\bullet}$	2000	4000	2000	4000						

Step 4

The gap at the end of the second iteration is $G^{(2)} \approx 6.076e-3 > \varepsilon$, hence we return to Step 2. Table C.3 indicates that flows on link 4 are now also in equilibrium, but sending flows for link 2 not yet.

Iteration $k=14$ (iterations 3 to 13 are skipped for brevity reasons)

Step 2

Gradients **g** are updated using Eq. (20) as $g_3 = g_4 = g_5 = g_7 = g_8 = \frac{1}{741}$, $g_6 = g_9 = g_{10} = \frac{1}{635}$, while all other gradients are equal to $\frac{1}{1000}$. The adjusted approach lane configuration is initialised again as $\tilde{\boldsymbol{\eta}} := \boldsymbol{\eta}$. For link 2, we solve the following system of linear equations,

$$\begin{pmatrix} s_{3,5}^{(14)} \\ s_{3,8}^{(14)} \\ s_{4,8}^{(14)} \\ s_{5,8}^{(14)} \\ s_{6,7}^{(14)} \\ s_{6,8}^{(14)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ \frac{1}{741} & \frac{1}{741} & -\frac{1}{741} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{741} & -\frac{1}{741} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{741} & -\frac{1}{635} & -\frac{1}{635} \end{pmatrix}^{-1} \begin{pmatrix} 200 \\ 600 \\ 3200 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 200 \\ 837 \\ 1037 \\ 1037 \\ 600 \\ 289 \end{pmatrix}.$$

The resulting movement-specific lane sending flows **s** after 14 iterations are shown in Table C.4.

Table C.4. Sending flow distribution after iteration 14.

	j	1	2	3	4		1	2	3	4	
i	m	$s_{mj}^{(14)}$				c_m	$v_{mj}^{(14)}$				$h_m^{(14)}$
1	1	0	0	0	600	1000	0	0	0	600	0.600
1	2	0	100	300	0	1000	0	100	300	0	0.400
2	3	200	0	0	837	1000	143	0	0	598	1.399
2	4	0	0	0	1037	1000	0	0	0	741	1.399
2	5	0	0	0	1037	1000	0	0	0	741	1.399
2	6	0	0	600	289	1000	0	0	429	207	1.399
3	7	200	200	0	400	1000	185	185	0	371	1.079
3	8	0	0	0	800	1000	0	0	0	741	1.079
4	9	0	0	800	0	1000	0	0	636	0	1.259
4	10	0	0	800	0	1000	0	0	636	0	1.259
4	11	0	900	0	0	1000	0	900	0	0	0.900
4	12	200	700	0	0	1000	200	700	0	0	0.900
	$r_j^{\bullet}$	2000	4000	2000	4000						

Step 3

Applying the lane-based node model results in updated movement-specific lane outflows **v** as summarised in Table C.4, with the same set of supply constrained lanes. The updated generalised lane exit costs **h** are computed using Eq. (9) and are shown in Table C.4.

Step 4

The gap at the end of the 14th iteration is $G^{(14)} \approx 6.573e-11 < \varepsilon$, hence we have found an equilibrium lane sending flow distribution, hence we terminate the algorithm and set $\bar{\mathbf{s}} := \mathbf{s}^{(14)}$.

Appendix D: Derivation of realised flow rates into virtual links

Consider the virtual node as part of the augmented link representation as shown in Fig. 4 and Fig. 5. Let $\tilde{s}_i(t)$ and $\tilde{r}_i(t)$ denote the sending and receiving flows associated with the virtual node in link *i* at time instant *t*, see Fig. 16. Further, let $\tilde{V}_i(t)$ denote the realised cumulative outflow out of the physical link. Since $W_i(t)$ denotes the realised cumulative inflow into the virtual link including any auxiliary flow, due to flow conservation it holds that $\tilde{V}_i(t) = W_i(t) - E_i(t)$, where $W_i(t) = \sum_{j \in J} W_{ij}(t)$ and $E_i(t) = \sum_{j \in J} E_{ij}(t)$ for Case II. Therefore, we can replace $\tilde{V}_i(t)$ with the respective variables on the virtual link in our formulation.

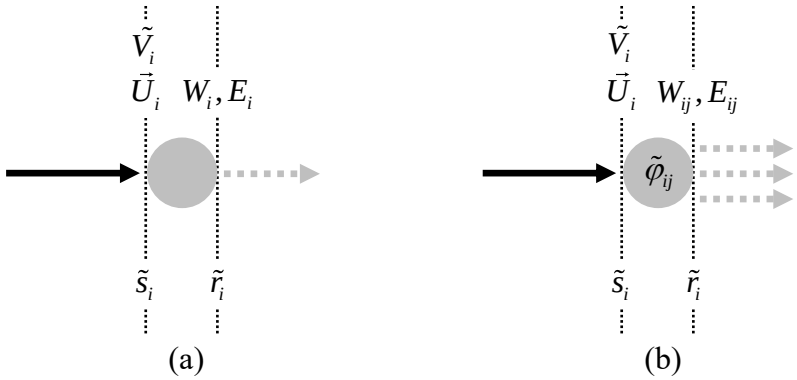


Fig. 16. Virtual node representation in (a) Case I, and (b) Case II

A.1 Case I: Link-specific exit capacities

Consider the virtual node in Case I as depicted in Fig. 16(a). Since the virtual node is a simple cross node, it holds that its transition flow rate $\tilde{d}_i(t)$ is equal to

$$\tilde{d}_i = \min\{\tilde{s}_i(t), \tilde{r}_i(t)\} = \min\{\tilde{s}_i(t), C_i\}, \tag{49}$$

noting that the receiving flow of the virtual link is always equal to the exit capacity, i.e., $\tilde{r}_i(t) = C_i$. If $\tilde{S}_i^+(t) = 0$ then the physical link has no excess sending flow, hence the sending flow rate on the physical link is equal to the potential outflow rate, $\vec{u}_i(t)$. If $\tilde{S}_i^+(t) > 0$ then the physical link has excess sending flow and hence the sending flow rate equals saturation flow. Therefore, we can write

$$\tilde{s}_i(t) = \begin{cases} \vec{u}_i(t), & \text{if } \tilde{S}_i^+(t) = 0, \\ Q_i, & \text{otherwise.} \end{cases} \tag{50}$$

If $S_i^+(t) > 0$, we need to ensure that $w_i(t) = C_i$ to correctly obtain half the red time as non-persistent delay. We ensure this by defining

$$w_i(t) = \tilde{d}_i(t) + e_i(t), \tag{51}$$

where auxiliary flow rate $e_i(t)$ increases the total inflow to capacity if there is excess flow on the virtual link,

$$e_i(t) = \begin{cases} C_i - \tilde{d}_i(t), & \text{if } S_i^+(t) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{52}$$

Combining Eqs. (49)–(52) yields

$$w_i(t) = \begin{cases} \min\{\vec{u}_i(t), C_i\}, & \text{if } \tilde{S}_i^+(t) = 0 \text{ and } S_i^+(t) = 0, \\ C_i, & \text{otherwise,} \end{cases} \tag{53}$$

and

$$e_i(t) = \begin{cases} C_i - \min\{\vec{u}_i(t), C_i\}, & \text{if } \tilde{S}_i^+(t) = 0 \text{ and } S_i^+(t) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{54}$$

A.2 Case II: Movement-specific exit capacities

Now consider the virtual node in Case II as depicted in Fig. 16(b). The sending flow is given by Eq. (50) and the movement-specific receiving flows are $\tilde{r}_{ij}(t) = C_{ij}$, where C_{ij} is given by Eq. (15). The virtual node is now a diverge node, which means that the most restrictive movement *ij* is the one that minimises $\tilde{r}_{ij}(t) / \left(\tilde{\varphi}_{ij}(t)\tilde{s}_i(t)\right)$, where $\tilde{\varphi}_{ij}(t)$ are split proportions at the time of virtual link entrance. Due to the FIFO property on the physical link, the proportion of sending flow that can enter virtual link *i* at time instant *t* is equal to

$$\lambda_i(t) = \min_{j' \in J} \left\{ 1, \frac{\tilde{r}_{ij'}(t)}{\tilde{\varphi}_{ij'}(t)\tilde{s}_i(t)} \right\} = \min_{j' \in J} \left\{ 1, \frac{C_{ij'}}{\tilde{\varphi}_{ij'}(t)\tilde{s}_i(t)} \right\}. \tag{55}$$

Hence, movement-specific transition flows are

$$\tilde{d}_{ij}(t) = \tilde{\varphi}_{ij}(t)\tilde{s}_i(t)\lambda_i(t) = \tilde{\varphi}_{ij}(t) \cdot \min_{j' \in J} \left\{ \tilde{s}_i(t), \frac{C_{ij'}}{\tilde{\varphi}_{ij'}(t)} \right\}. \tag{56}$$

Note that this guarantees that $w_{ij}(t) \le C_{ij}$.

To ensure correct uniform delays in congested conditions, we again define the inflow rate into each virtual link as the sum of the transition flow rate and a movement-specific auxiliary flow rate $e_{ij}(t)$,

$$w_{ij}(t) = \tilde{d}_{ij}(t) + e_{ij}(t), \tag{57}$$

where excess flow on virtual link *ij* at time *t* is defined as

$$e_{ij}(t) = \begin{cases} C_{ij} - \tilde{d}_{ij}(t), & \text{if } S_{ij}^{+}(t) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{58}$$

This guarantees that $w_{ij}(t) = C_{ij}$ if $S_{ij}^{+}(t) > 0$. Combining Eq. (50) with Eqs. (56)–(58) yields

$$w_{ij}(t) = \begin{cases} \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ \vec{u}_i(t), C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^{+}(t) = 0 \text{ and } S_{ij}^{+}(t) = 0, \\ \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ Q_i, C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^{+}(t) > 0 \text{ and } S_{ij}^{+}(t) = 0, \\ C_{ij}, & \text{otherwise,} \end{cases} \tag{59}$$

and

$$e_{ij}(t) = \begin{cases} C_{ij} - \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ \vec{u}_i(t), C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^{+}(t) = 0 \text{ and } S_{ij}^{+}(t) > 0, \\ C_{ij} - \tilde{\varphi}_{ij}(t) \cdot \min\limits_{j' \in J} \left\{ Q_i, C_{ij'} / \tilde{\varphi}_{ij'}(t) \right\}, & \text{if } \tilde{S}_i^{+}(t) > 0 \text{ and } S_{ij}^{+}(t) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{60}$$

Using Eq. (13) we can compute the outflow rate out of the physical link,

$$\tilde{v}_i(t) = \sum_{j \in J} \tilde{d}_{ij}(t). \tag{61}$$

Appendix E: Augmented LTM algorithm for Case II

Algorithm 2 below describes the steps for applying the augmented LTM that accounts for movement-specific non-persistent delay in Case II. An asterisk (*) indicates a new step compared to traditional LTM.

Algorithm 2. Pseudo-algorithm for augmented LTM for Case II (movement-specific exit capacities)

Step 1: Initialisation.

a. Set $U_i(0) = U_{ip}(0) = V_i(0) = V_{ip}(0) = 0, \forall i \in A.$

b. * Set $W_{ij}(0) = W_{ip}(0) = E_{ij}(0) = V_{ij}(0) = \tilde{V}_i(0) = 0, \forall i \in I_n, \forall j \in J_n, \forall n \in N^*.$

c. Set $t := 0.$

Step 2: Realised cumulative inflows.

a. $U_{ip}(t) = \begin{cases} F_p(t), & \text{if link } i \text{ is the first link on path } p, \\ V_{i'p}(t), & \text{if link } i' \text{ is the previous link on path } p, \end{cases} \quad \forall i \in A, \forall p \in P.$

b. $U_i(t) = \sum_{p \in P} U_{ip}(t), \quad \forall i \in A.$

Step 3: Potential cumulative flows.

a. $\vec{U}_i(t+\Delta t) = \min_{q \in [0, Q_i]} \left\{ U_i\left(t + \Delta t - \frac{L_i}{\gamma_{\mathrm{I},i}(q)} \right) + \frac{L_i q}{\gamma_{\mathrm{I},i}(q)} - \frac{L_i q}{\sigma_{\mathrm{I},i}(q)} \right\}, \quad \forall i \in A.$ [Eq. (8)]

b. $\overleftarrow{V}_i(t+\Delta t) = \min_{q \in [0, Q_i]} \left\{ V_i\left(t + \Delta t + \frac{L_i}{\gamma_{\mathrm{II},i}(q)} \right) + \frac{L_i q}{\sigma_{\mathrm{II},i}(q)} - \frac{L_i q}{\gamma_{\mathrm{II},i}(q)} \right\}, \quad \forall i \in A.$ [Eq. (9)]

c. * $\vec{W}_{ij}(t+\Delta t) = \min_{q \in [0, C_{ij}]} \left\{ W_{ij}\left(t + \Delta t - \frac{\tau_{ij}(q) Q_{ij}}{Q_{ij} - q} \right) + \frac{\tau_{ij}(q) q^2}{Q_{ij} - q} \right\}, \quad \forall i \in A^*.$ [Eq. (45)]

e. * $\vec{E}_{ij}(t+\Delta t) = E_{ij}\left(W_{ij}^{-1}\left(\vec{W}_{ij}(t+\Delta t) \right) \right), \quad \forall i \in A^*.$ [Eq. (20)]

Step 4: Sending and receiving flows during time interval $[t, t+\Delta t]$.

a. $S_i(t) = \min\left\{ \vec{U}_i(t+\Delta t) - V_i(t), C_i \cdot \Delta t \right\}, \quad \forall i \in A \setminus A^*.$ [Eq. (11)]

b. $R_i(t) = \min\left\{ \overleftarrow{V}_i(t+\Delta t) - U_i(t), Q_i \cdot \Delta t \right\}, \quad \forall i \in A.$ [Eq. (11)]

c. * $\tilde{S}_i(t) = \min\left\{ \vec{U}_i(t+\Delta t) - W_i(t) + E_i(t), Q_i \cdot \Delta t \right\}, \quad \forall i \in A^*.$ [Eq. (50)]

d. * $S_{ij}(t) = \min\left\{ \vec{W}_{ij}(t+\Delta t) - \vec{E}_{ij}(t+\Delta t) - V_{ij}(t), C_i \cdot \Delta t \right\}, \quad \forall i \in A^*.$ [Eq. (44)]

Step 5: Split proportions.

a. $t_i' = U_i^{-1}\left(V_i(t) + S_i(t) \right), \quad \forall i \in A \setminus A^*.$

b. $\varphi_{ij}(t) = \dfrac{\sum_{p\in P} \delta_{jp}\left(U_{ip}(t_i') - V_{ip}(t)\right)}{\sum_{j'\in J_n}\sum_{p\in P} \delta_{j'p}\left(U_{ip}(t_i') - V_{ip}(t)\right)}, \quad \forall i \in I_n, \forall j \in J_n, \forall n \in N \setminus N^*.$

c. * $t_i'' = U_i^{-1}\left(\tilde{V}_i(t) + \tilde{S}_i(t)\right), \quad \forall i \in A^*.$

d. * $\tilde{\varphi}_{ij}(t) = \dfrac{\sum_{p\in P} \delta_{jp}\left(U_{ip}(t_i'') - \tilde{V}_{ip}(t)\right)}{\sum_{j'\in J_n}\sum_{p\in P} \delta_{j'p}\left(U_{ip}(t_i'') - \tilde{V}_{ip}(t)\right)}, \quad \forall i \in I_n, \forall j \in J_n, \forall n \in N^*.$

Step 6: Transition and auxiliary flows.

a. $\mathbf{D}_n(t) = \Gamma\left(\mathbf{S}_n(t), \mathbf{R}_n(t), \boldsymbol{\varphi}_n(t), \mathbf{C}_n \cdot \Delta t\right), \quad \forall n \in N \setminus N^*.$ [Eq. (41)]

b. * $\text{stack}[\text{diag}(\mathbf{D}_n(t))] = \Gamma\left(\text{vec}[\mathbf{S}_n(t)], \mathbf{R}_n(t), \text{stack}[\mathbf{1}_{|J_n|}], \text{vec}[\mathbf{C}_n] \cdot \Delta t\right), \forall n \in N^*.$ [Eq.(46)]

c. * $\tilde{D}_{ij}(t) = \tilde{\varphi}_{ij}(t) \cdot \min_{j'\in J}\left\{\tilde{S}_i(t), C_{ij'} \cdot \Delta t / \tilde{\varphi}_{ij'}(t)\right\}, \quad \forall i \in A^*.$ [Eq. (56)]

e. * $\tilde{E}_{ij}(t) = \min\left\{\vec{W}_{ij}(t) - \vec{E}_{ij}(t) - V_{ij}(t), C_{ij} \cdot \Delta t - \tilde{D}_{ij}(t)\right\}, \quad \forall i \in A^*.$ [Eq. (58)]

Step 7: Realised cumulative flows.

a. $V_i(t+\Delta t) = V_i(t) + \sum_{j\in J} D_{ij}(t), \quad \forall i \in A \setminus A^*.$ [Eqs. (6) and (13)]

b. $V_{ip}(t+\Delta t) = U_{ip}\left(U_i^{-1}\left(V_i(t+\Delta t)\right)\right), \quad \forall i \in A, \forall p \in P.$

c. * $E_{ij}(t+\Delta t) = E_{ij}(t) + \tilde{E}_{ij}(t), \quad \forall i \in A^*.$ [Eq. (19)]

d. * $E_i(t+\Delta t) = \sum_{j\in J} E_{ij}(t), \quad \forall i \in A^*.$ [Eq. (32)]

e. * $W_{ij}(t+\Delta t) = W_{ij}(t) + \tilde{D}_{ij}(t) + \tilde{E}_{ij}(t), \quad \forall i \in A^*.$ [Eq. (18) and (57)]

f. * $W_i(t+\Delta t) = \sum_{j\in J} W_{ij}(t+\Delta t), \quad \forall i \in A^*.$ [Eq. (32)]

g. * $\tilde{V}_i(t+\Delta t) = \tilde{V}_i(t) + \sum_{j\in J} \tilde{D}_{ij}(t), \quad \forall i \in A^*.$ [Eq. (61)]

h. * $V_{ij}(t+\Delta t) = V_{ij}(t) + D_{ij}(t), \quad \forall i \in A^*.$ [Eqs. (6)]

i. * $V_i(t+\Delta t) = \sum_{j\in J} V_{ij}(t), \quad \forall i \in A^*.$ [Eq. (32)]

j. * $\tilde{V}_{ip}(t+\Delta t) = U_{ip}\left(U_i^{-1}\left(\tilde{V}_i(t+\Delta t)\right)\right), \quad \forall i \in A^*, \forall p \in P.$

Step 8: Terminate or repeat.

a. Set $t := t + \Delta t$.

b. If $t = T$ then stop. Otherwise, return to Step 2.

Appendix F: FIFO condition in flow propagation

If $\tau(t)$ is Webster's uniform delay for a vehicle entering the link at time t, then it should hold that

$$W(t) = \vec{W}\left(t + \tau(t)\right). \tag{62}$$

where $W(t)$ is the realised cumulative inflow into the virtual link (for link *i* or movement *ij*), and $\vec{W}(t)$ is the potential cumulative outflow. Differentiating Eq. (62) with respect to time yields

$$\frac{dW(t)}{dt} = \frac{d\vec{W}\left(t+\tau(t)\right)}{dt} \quad \Rightarrow \quad w(t) = w\left(t+\tau(t)\right)\left(1+\frac{d\tau(t)}{dt}\right), \tag{63}$$

where $w(t)$ is defined in Eq. (17). Rearranging terms results in the following flow propagation equation,

$$w\left(t+\tau(t)\right) = \frac{w(t)}{1+\frac{d\tau}{dt}}. \tag{64}$$

The avoid negative flows, it should hold that

$$\frac{d\tau(t)}{dt} > -1. \tag{65}$$

This is also known as the FIFO condition in flow propagation, see for example Astarita (1996).

References

Akçelik, R. (1980). *Time-Dependent Expressions for Delay, Stop Rate and Queue Length at Traffic Signals*.

Asamer, J., & Reinthaler, M. (2010). Estimation of road capacity and free flow speed for urban roads under adverse weather conditions. *13th International IEEE Conference on Intelligent Transportation Systems*, 812–818. https://doi.org/10.1109/ITSC.2010.5625152

Astarita, V. (1996). A continuous time link model for dynamic network loading based on travel time function. *Proceedings Of The 13th International Symposium on Transportation and Traffic Theory*, 79–102.

Beckmann, M. J., McGuire, C. B., & Winsten, C. B. (1956). Studies in the Economics of Transportation. *Yale University Press*. https://www.rand.org/pubs/research_memoranda/RM1488.html

Bifulco, G., & Crisalli, U. (1998). Stochastic user equilibrium and link capacity constraints: Formulation and theoretical evidences. *Proceedings of the European Transport Conference*.

Bliemer, M. C. J. (2007). Dynamic queuing and spillback in analytical multiclass dynamic network loading model. *Transportation Research Record*, *2029*(1), 14–21. https://doi.org/10.3141/2029-02

Bliemer, M. C. J., & Bovy, P. H. L. (2003). Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem. *Transportation Research Part B: Methodological*, *37*(6), 501–519. https://doi.org/10.1016/S0191-2615(02)00025-5

Bliemer, M. C. J., & Raadsen, M. P. H. (2019). Continuous-time general link transmission model with simplified fanning, Part I: Theory and link model formulation. *Transportation Research Part B: Methodological*, *126*, 442–470. https://doi.org/10.1016/j.trb.2018.01.001

Bliemer, M. C. J., & Raadsen, M. P. H. (2020). Static traffic assignment with residual queues and spillback. *Transportation Research Part B: Methodological*, *132*, 303–319. https://doi.org/10.1016/j.trb.2019.02.010

Bliemer, M. C. J., Raadsen, M. P. H., Brederode, L. J. N., Bell, M. G. H., Wismans, L. J. J., & Smith, M. J. (2017). Genetics of traffic assignment models for strategic transport planning. *Transport Reviews*, *37*(1), 56–78. https://doi.org/10.1080/01441647.2016.1207211

Bliemer, M. C. J., Raadsen, M. P. H., Smits, E.-S., Zhou, B., & Bell, M. G. H. (2014). Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model. *Transportation Research Part B: Methodological*, *68*, 363–384. https://doi.org/10.1016/j.trb.2014.07.001

Bureau of Public Roads. (1964). *Traffic Assignment Manual*. U.S. Dept. of Commerce, Ubran Planning Division, Washington, DC, USA.

Cascetta, E. (2009). *Transportation Systems Analysis: Models and Applications*. Springer Science & Business Media.

Chabini, I. (1998). Discrete Dynamic Shortest Path Problems in Transportation Applications: Complexity and Algorithms with Optimal Run Time. *Transportation Research Record: Journal of the Transportation Research Board*, *1645*(1), 170–175. https://doi.org/10.3141/1645-21

Chen, H.-K. (1999). *Dynamic Travel Choice Models*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-59980-4

Coclite, G. M., Garavello, M., & Piccoli, B. (2005). Traffic Flow on a Road Network. *SIAM Journal on Mathematical Analysis*, *36*(6), 1862–1886. https://doi.org/10.1137/S0036141004402683

Corthout, R., Flötteröd, G., Viti, F., & Tampère, C. M. J. (2012). Non-unique flows in macroscopic first-order intersection models. *Transportation Research Part B: Methodological*, *46*(3), 343–359. https://doi.org/10.1016/j.trb.2011.10.011

Covidence. (2014). [Computer software]. https://www.covidence.org/

Daganzo, C. F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, *28*(4), 269–287. https://doi.org/10.1016/0191-2615(94)90002-7

Daganzo, C. F. (1995). The cell transmission model, part II: Network traffic. *Transportation Research Part B: Methodological*, *29*(2), 79–93. https://doi.org/10.1016/0191-2615(94)00022-R

Daganzo, C. F. (1997). A continuum theory of traffic dynamics for freeways with special lanes. *Transportation Research Part B: Methodological*, *31*(2), 83–102. https://doi.org/10.1016/S0191-2615(96)00017-3

Daganzo, C. F., Lin, W.-H., & Del Castillo, J. M. (1997). A simple physical principle for the simulation of freeways with special lanes and priority vehicles. *Transportation Research Part B: Methodological*, *31*(2), 103–125. https://doi.org/10.1016/S0191-2615(96)00032-X

Durlin, T., & Henn, V. (2005). A Delayed Flow Intersection Model for Dynamic Traffic Assignment. *10th EURO Working Group on Transportation Meeting and 16th Mini-EURO Conference*.

Durlin, T., & Henn, V. (2008). Dynamic Network Loading Model with Explicit Traffic Wave Tracking. *Transportation Research Record: Journal of the Transportation Research Board*, *2085*(1), 1–11. https://doi.org/10.3141/2085-01

Fitzpatrick, K., Schneider, W. H., & Park, E. S. (2006). Predicting Speeds in an Urban Right-Turn Lane. *Journal of Transportation Engineering*, *132*(3), 199–204. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:3(199)

Flötteröd, G., & Osorio, C. (2017). Stochastic network link transmission model. *Transportation Research Part B: Methodological*, *102*, 180–209. https://doi.org/10.1016/j.trb.2017.04.009

Flötteröd, G., & Rohde, J. (2011). Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B: Methodological*, *45*(6), 903–922. https://doi.org/10.1016/j.trb.2011.04.001

Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L., & Wie, B. W. (1993). A Variational Inequality Formulation of the Dynamic Network User Equilibrium Problem. *Operations Research*, *41*(1), 179–191.

Friesz, T. L., Han, K., Neto, P. A., Meimand, A., & Yao, T. (2013). Dynamic user equilibrium based on a hydrodynamic model. *Transportation Research Part B: Methodological*, *47*, 102–126. https://doi.org/10.1016/j.trb.2012.10.001

Gao, M. (2012). A Node Model Capturing Turning Lane Capacity and Physical Queuing for the Dynamic Network Loading Problem. *Mathematical Problems in Engineering*, *2012*, 1–14. https://doi.org/10.1155/2012/542459

Gentile, G. (2010). The General Link Transmission Model for Dynamic Network Loading and a Comparison with the DUE Algorithm. In *New Developments in Transport Planning* (pp. 153–178). Edward Elgar Publishing. https://doi.org/10.4337/9781781000809.00016

Gentile, G., Meschini, L., & Papola, N. (2007). Spillback congestion in dynamic traffic assignment: A macroscopic flow model with time-varying bottlenecks. *Transportation Research Part B: Methodological*, *41*(10), 1114–1138. https://doi.org/10.1016/j.trb.2007.04.011

Gibb, J. (2011). Model of Traffic Flow Capacity Constraint through Nodes for Dynamic Network Loading with Queue Spillback. *Transportation Research Record: Journal of the Transportation Research Board*, *2263*(1), 113–122. https://doi.org/10.3141/2263-13

Gong, X., Bliemer, M., & Raadsen, M. (2024). *Extended Macroscopic Node Model for Multilane Traffic*. https://doi.org/10.2139/ssrn.4772868

Guo, Q., Ban, X. (Jeff), & Aziz, H. M. A. (2021). Mixed traffic flow of human driven vehicles and automated vehicles on dynamic transportation networks. *Transportation Research Part C: Emerging Technologies*, *128*, 103159. https://doi.org/10.1016/j.trc.2021.103159

Hajiahmadi, M., Corthout, R., Tampère, C., De Schutter, B., & Hellendoorn, H. (2013). Variable Speed Limit Control Based on Extended Link Transmission Model. *Transportation Research Record: Journal of the Transportation Research Board*, *2390*(1), 11–19. https://doi.org/10.3141/2390-02

Han, K., & Gayah, V. V. (2015). Continuum signalized junction model for dynamic traffic networks: Offset, spillback, and multiple signal phases. *Transportation Research Part B: Methodological*, *77*, 213–239. https://doi.org/10.1016/j.trb.2015.03.005

Han, K., Gayah, V. V., Piccoli, B., Friesz, T. L., & Yao, T. (2014). On the continuum approximation of the on-and-off signal control on dynamic traffic networks. *Transportation Research Part B: Methodological*, *61*, 73–97. https://doi.org/10.1016/j.trb.2014.01.001

Han, K., Piccoli, B., & Szeto, W. Y. (2016). Continuous-time link-based kinematic wave model: Formulation, solution existence, and well-posedness. *Transportmetrica B: Transport Dynamics*, *4*(3), 187–222. https://doi.org/10.1080/21680566.2015.1064793

Herty, M., & Klar, A. (2003). Modeling, Simulation, and Optimization of Traffic Flow Networks. *SIAM Journal on Scientific Computing*, *25*(3), 1066–1087. https://doi.org/10.1137/S106482750241459X

Himpe, W., Corthout, R., & Tampère, M. J. C. (2016). An efficient iterative link transmission model. *Transportation Research Part B: Methodological*, *92*, 170–190. https://doi.org/10.1016/j.trb.2015.12.013

Holden, H., & Risebro, N. H. (1995). A Mathematical Model of Traffic Flow on a Network of Unidirectional Roads. *SIAM Journal on Mathematical Analysis*, *26*(4), 999–1017. https://doi.org/10.1137/S0036141093243289

Jabari, S. E. (2016). Node modeling for congested urban road networks. *Transportation Research Part B: Methodological*, *91*, 229–249. https://doi.org/10.1016/j.trb.2016.06.001

Janson, B. N. (1991). Convergent algorithm for dynamic traffic assignment. *Transportation Research Record*, *1328*.

Jin, W. L. (2010). Continuous kinematic wave models of merging traffic flow. *Transportation Research Part B: Methodological*, *44*(8), 1084–1103. https://doi.org/10.1016/j.trb.2010.02.011

Jin, W. L. (2015). Continuous formulations and analytical properties of the link transmission model. *Transportation Research Part B: Methodological*, *74*, 88–103. https://doi.org/10.1016/j.trb.2014.12.006

Jin, W. L., & Zhang, H. M. (2003). On the distribution schemes for determining flows through a merge. *Transportation Research Part B: Methodological*, *37*(6), 521–540. https://doi.org/10.1016/S0191-2615(02)00026-7

Jin, W. L., & Zhang, H. M. (2004). Multicommodity Kinematic Wave Simulation Model for Network Traffic Flow. *Transportation Research Record: Journal of the Transportation Research Board*, *1883*(1), 59–67. https://doi.org/10.3141/1883-07

Jin, W. L. (2012). A kinematic wave theory of multi-commodity network traffic flow. *Transportation Research Part B: Methodological*, *46*(8), 1000–1022. https://doi.org/10.1016/j.trb.2012.02.009

Jin, W. L. (2015). Analysis of Kinematic Waves Arising in Diverging Traffic Flow Models. *Transportation Science*, *49*(1), 28–45. https://doi.org/10.1287/trsc.2013.0499

Jin, W. L., & Zhang, H. M. (2013). An instantaneous kinematic wave theory of diverging traffic. *Transportation Research Part B: Methodological*, *48*, 1–16. https://doi.org/10.1016/j.trb.2012.12.001

Jin, Y., Yao, Z., Han, J., Hu, L., & Jiang, Y. (2022). Variable Cell Transmission Model for Mixed Traffic Flow with Connected Automated Vehicles and Human-Driven Vehicles. *Journal of Advanced Transportation*, *2022*, 1–15. https://doi.org/10.1155/2022/6342857

Laval, J. A., & Daganzo, C. F. (2006). Lane-changing in traffic streams. *Transportation Research Part B: Methodological*, *40*(3), 251–264. https://doi.org/10.1016/j.trb.2005.04.003

Lebacque, J. P. (1996). The Godunov scheme and what it means for rst order tra c ow models 1. *Proceedings of the 13th ISTTT Symposium*, 647-678.

Lebacque, J. P., & Khoshyaran, M. M. (2005). First-Order Macroscopic Traffic Flow Models: Intersection Modeling, Network Modeling. *Proceedings of the 16th International Symposium on Transportation and Traffic Theory (ISTTT)*, 365–386. https://trid.trb.org/view/758861

Lebacque, J.-P. (1996). The Godunov scheme and what it means for first order traffic flow models. *Proceedings of the 13th International Symposium on Transportation and Traffic Theory (ISTTT)*.

Levin, M. W., & Boyles, S. D. (2016a). A cell transmission model for dynamic lane reversal with autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *68*, 126–143. https://doi.org/10.1016/j.trc.2016.03.007

Levin, M. W., & Boyles, S. D. (2016b). A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transportation Research Part C: Emerging Technologies*, *62*, 103–116. https://doi.org/10.1016/j.trc.2015.10.005

Levin, M. W., & Kang, D. (2023). A Multiclass Link Transmission Model for a Class-Varying Capacity and Congested Wave Speed. *Journal of Transportation Engineering, Part A: Systems*, *149*(10), 04023096. https://doi.org/10.1061/JTEPBS.TEENG-7940

Levin, M. W., Smith, H., & Boyles, S. D. (2019). Dynamic Four-Step Planning Model of Empty Repositioning Trips for Personal Autonomous Vehicles. *Journal of Transportation Engineering, Part A: Systems*, *145*(5), 04019015. https://doi.org/10.1061/JTEPBS.0000235

Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *229*(1178), 317–345. https://doi.org/10.1098/rspa.1955.0089

Liu, H., Wang, J., Wijayaratna, K., Dixit, V. V., & Waller, S. T. (2015). Integrating the Bus Vehicle Class Into the Cell Transmission Model. *IEEE Transactions on Intelligent Transportation Systems*, *16*(5), 2620–2630. https://doi.org/10.1109/TITS.2015.2413995

Liu, S., Jiang, H., & Zhang, Y. (2023). CTM-based traffic signal optimization with variable speed limit and dynamic lane allocation. *2023 6th International Symposium on Autonomous Systems (ISAS)*, 1–6. https://doi.org/10.1109/ISAS59543.2023.10164396

Lu, J., & Osorio, C. (2018). A probabilistic traffic-theoretic network loading model suitable for large-scale network analysis. *Transportation Science*, *52*(6), 1509–1530.

Lu, J., & Osorio, C. (2022). On the Analytical Probabilistic Modeling of Flow Transmission Across Nodes in Transportation Networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2676*(12), 209–225. https://doi.org/10.1177/03611981221094829

McNeil, D. R. (1968). A Solution to the Fixed-Cycle Traffic Light Problem for Compound Poisson Arrivals. *Journal of Applied Probability*, *5*(3), 624–635. https://doi.org/10.2307/3211926

Merchant, D. K., & Nemhauser, G. L. (1978a). A Model and an Algorithm for the Dynamic Traffic Assignment Problems. *Transportation Science*, *12*(3), 183–199.

Merchant, D. K., & Nemhauser, G. L. (1978b). Optimality Conditions for a Dynamic Traffic Assignment Model. *Transportation Science*, *12*(3), 200–207. https://doi.org/10.1287/trsc.12.3.200

Miller, A. (1968). The capacity of signalized intersections in Australia. *Australian Road Research Board*, *Bulletin No. 3*. https://www.semanticscholar.org/paper/The-capacity-of-signalized-intersections-in-Miller/70fcc42d3eb7df1068a1a443d05a681c6aadfe1c

Miller, A. J. (1969). Some operating characteristics of fixed-time signals with random arrivals. https://trid.trb.org/View/114910

Nagurney, A. (1998). *Network Economics: A Variational Inequality Approach*. Springer Science & Business Media.

Newell, G. F. (1960). Queues for a Fixed-Cycle Traffic Light. *The Annals of Mathematical Statistics*, *31*(3), 589–597. https://doi.org/10.1214/aoms/1177705787

Newell, G. F. (1965). Approximation Methods for Queues with Application to the Fixed-Cycle traffic Light. *SIAM Review*, *7*(2), 223–240.

Newell, G. F. (1993a). A simplified theory of kinematic waves in highway traffic, part I: General theory. *Transportation Research Part B: Methodological*, *27*(4), 281–287. https://doi.org/10.1016/0191-2615(93)90038-C

Newell, G. F. (1993b). A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, *27*(4), 289–303. https://doi.org/10.1016/0191-2615(93)90039-D

Newell, G. F. (1999). Delays caused by a queue at a freeway exit ramp. *Transportation Research Part B: Methodological*, *33*(5), 337–350. https://doi.org/10.1016/S0191-2615(98)00039-3

Ngoduy, D., Hoang, N. H., Vu, H. L., & Watling, D. (2021). Multiclass dynamic system optimum solution for mixed traffic of human-driven and automated vehicles considering physical queues. *Transportation Research Part B: Methodological*, *145*, 56–79. https://doi.org/10.1016/j.trb.2020.12.008

Ni, D., & Leonard, J. D. (2004). Simplified kinematic waves at a diverge. *The 8th World Multiconference on Systemics*, *VII*, 425–430.

Ni, D., & Leonard, J. D. (2005). A simplified kinematic wave model at a merge bottleneck. *Applied Mathematical Modelling*, *29*(11), 1054–1072. https://doi.org/10.1016/j.apm.2005.02.008

Ni, D., Leonard, J. D., & Williams, B. M. (2006). The Network Kinematic Waves Model: A Simplified Approach to Network Traffic. *Journal of Intelligent Transportation Systems*, *10*(1), 1–14. https://doi.org/10.1080/15472450500455070

Ohno, K. (1978). Computational Algorithm for a Fixed Cycle Traffic Signal and New Approximate Expressions for Average Delay. *Transportation Science*, *12*(1), 29–47. https://doi.org/10.1287/trsc.12.1.29

Osorio, C., Flötteröd, G., & Bierlaire, M. (2011). Dynamic network loading: A stochastic differentiable model that derives link state distributions. *Transportation Research Part B: Methodological*, *45*(9), 1410–1423. https://doi.org/10.1016/j.trb.2011.05.014

Petigny, B. (1967). Remarks on the equivalent of a vehicle in passenger car units (p.c.u.). *Transportation Research*, *1*(4), 339–348.

PRISMA. (2015). Transparent reporting of systematic reviews and meta-analyses. http://www.prisma-statement.org/

Qin, Y., & Wang, H. (2019). Cell Transmission Model for Mixed Traffic Flow with Connected and Autonomous Vehicles. *Journal of Transportation Engineering, Part A: Systems*, *145*(5), 04019014. https://doi.org/10.1061/JTEPBS.0000238

Raadsen, M. P. H., & Bliemer, M. C. J. (2019). Continuous-time general link transmission model with simplified fanning, Part II: Event-based algorithm for networks. *Transportation Research Part B: Methodological*, *126*, 471–501. https://doi.org/10.1016/j.trb.2018.01.003

Raadsen, M. P. H., & Bliemer, M. C. J. (2021). Variable speed limits in the link transmission model using an information propagation method. *Transportation Research Part C: Emerging Technologies*, *129*, 103184. https://doi.org/10.1016/j.trc.2021.103184

Raadsen, M. P. H., & Bliemer, M. C. J. (2023). General solution scheme for the static link transmission model. *Transportation Research Part B: Methodological*, *169*, 108–135. https://doi.org/10.1016/j.trb.2022.11.012

Richards, P. I. (1956). Shock Waves on the Highway. *Operations Research*, *4*(1), 42–51. https://doi.org/10.1287/opre.4.1.42

Shiomi, Y., Taniguchi, T., Uno, N., Shimamoto, H., & Nakamura, T. (2015). Multilane first-order traffic flow model with endogenous representation of lane-flow equilibrium. *Transportation Research Part C: Emerging Technologies*, *59*, 198–215. https://doi.org/10.1016/j.trc.2015.07.002

Shirke, C., Bhaskar, A., & Chung, E. (2019). Macroscopic modelling of arterial traffic: An extension to the cell transmission model. *Transportation Research Part C: Emerging Technologies*, *105*, 54–80. https://doi.org/10.1016/j.trc.2019.05.033

Smith, M., Huang, W., & Viti, F. (2013). Equilibrium in Capacitated Network Models with Queueing Delays, Queue-storage, Blocking Back and Control. *Procedia - Social and Behavioral Sciences*, *80*, 860–879. https://doi.org/10.1016/j.sbspro.2013.05.047

Smits, E.-S., Bliemer, M. C. J., Pel, A. J., & van Arem, B. (2015). A family of macroscopic node models. *Transportation Research Part B: Methodological*, *74*, 20–39. https://doi.org/10.1016/j.trb.2015.01.002

Smits, E.-S., Bliemer, M., & van Arem, B. (2011). Dynamic Network Loading of Multiple User-Classes with the Link Transmission Model. *In 2nd International Conference on Models and Technologies for Intelligent Transportation Systems*, *57*, 73.

Tampère, C. M. J., Corthout, R., Cattrysse, D., & Immers, L. H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B: Methodological*, *45*(1), 289–309. https://doi.org/10.1016/j.trb.2010.06.004

Thonhofer, E., Palau, T., Kuhn, A., Jakubek, S., & Kozek, M. (2018). Macroscopic traffic model for large scale urban traffic network design. *Simulation Modelling Practice and Theory*, *80*, 32–49. https://doi.org/10.1016/j.simpat.2017.09.007

Van der Gun, J. P. T., Pel, A. J., & van Arem, B. (2017). Extending the Link Transmission Model with non-triangular fundamental diagrams and capacity drops. *Transportation Research Part B: Methodological*, *98*, 154–178. https://doi.org/10.1016/j.trb.2016.12.011

Van der Gun, J. P. T., Pel, A. J., & Van Arem, B. (2019). The link transmission model with variable fundamental diagrams and initial conditions. *Transportmetrica B: Transport Dynamics*, *7*(1), 834–864. https://doi.org/10.1080/21680566.2018.1517060

Viti, F., & Van Zuylen, H. J. (2010). Probabilistic models for queues at fixed control signals. *Transportation Research Part B: Methodological*, *44*(1), 120–135. https://doi.org/10.1016/j.trb.2009.05.001

Wada, K., Ohata, T., Kobayashi, K., & Kuwahara, M. (2015). Traffic Measurements on Signalized Arterials from Vehicle Trajectories. *Interdisciplinary Information Sciences*, *21*(1), 77–85. https://doi.org/10.4036/iis.2015.77

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *Proceedings of Institution of Civil Engineers, Part II*, *1*, 325–378.

Webster, F. V. (1958). Traffic Signal Settings. *Road Research Technical Paper No.39*, *Road Research Laboratory*.

Wei, L., Waller, S. T., Mei, Y., Wang, Y., & Wang, M. (2023). *A Link Transmission Model with Variable Speed Limits and Turn-Level Queue Transmission at Signalized Intersections* (No. arXiv:2310.12249). arXiv. http://arxiv.org/abs/2310.12249

Wright, M. A., Gomes, G., Horowitz, R., & Kurzhanskiy, A. A. (2017). On node models for high-dimensional road networks. *Transportation Research Part B: Methodological*, *105*, 212–234.

Wright, M. A., Horowitz, R., & Kurzhanskiy, A. A. (2016). A dynamic system characterization of road network node models. *IFAC-PapersOnLine*, *49*(18), 1054–1059. https://doi.org/10.1016/j.ifacol.2016.10.307

Wu, J. H., Chen, Y., & Florian, M. (1998). The continuous dynamic network loading problem: A mathematical formulation and solution method. *Transportation Research Part B: Methodological*, *32*(3), 173–187. https://doi.org/10.1016/S0191-2615(97)00023-4

Xu, F., He, Z., Sha, Z., Zhuang, L., & Sun, W. (2013). Assessing the Impact of Rainfall on Traffic Operation of Urban Road Network. *Procedia - Social and Behavioral Sciences*, *96*, 82–89. https://doi.org/10.1016/j.sbspro.2013.08.012

Xu, Y. W., Wu, J. H., Florian, M., Marcotte, P., & Zhu, D. L. (1999). Advances in the Continuous Dynamic Network Loading Problem. *Transportation Science*, *33*(4), 341–353.

Yahyamozdarani, R., & Tampère, C. M. J. (2023). The continuous signalized (COS) node model for dynamic traffic assignment. *Transportation Research Part B: Methodological*, *168*, 56–80. https://doi.org/10.1016/j.trb.2022.12.003

Yahyamozdarani, R. (2024). *Advanced node modeling in macroscopic network loading*. Katholieke Universiteit Leuven.

Yperman, I. (2007). *The Link Transmission Model for Dynamic Network Loading*. Katholieke Universiteit Leuven.

Yperman, I., Logghe, S., & Immers, B. (2005). The link transmission model: An efficient implementation of the kinematic wave theory in traffic networks. *Proceedings of the 10th EWGT Meeting*, *24*, 122–127.

Yperman, I., & Tampère, C. (2006). Multi-commodity dynamic network loading with kinematic waves and intersection delays. *Proceedings of the DTA*.

Nie, Y., Ma, J., & Zhang, H. M. (2008). A Polymorphic Dynamic Network Loading Model. *Computer-Aided Civil and Infrastructure Engineering*, *23*(2), 86–103. https://doi.org/10.1111/j.1467-8667.2007.00525.x

Zhang, F., Lu, J., Hu, X., & Meng, Q. (2023). A stochastic dynamic network loading model for mixed traffic with autonomous and human-driven vehicles. *Transportation Research Part B: Methodological*, *178*, 102850. https://doi.org/10.1016/j.trb.2023.102850