Doctoral dissertation at the faculty of Engineering of KU Leuven

INTERSECTION MODELLING AND MARGINAL SIMULATION IN MACROSCOPIC DYNAMIC NETWORK LOADING

Ruben CORTHOUT

Supervisors:
Prof. ir. L.H. Immers
Prof. dr. ir. C.M.J. Tampère

Members of the
examination committee:
Prof. dr. ir. P. Verbaeten, chair
Prof. dr. C. Buisson
Prof. dr. ir. D. Cattrysse
Prof. dr. ir. S. Hoogendoorn
Prof. dr. ir. J. Vandewalle

Dissertation presented in
partial fulfilment of the
requirements for the degree
of Doctor of Engineering

May 2012

ISBN 978-94-6018-506-9
D/2012/7515/42

"*Science never solves a problem without creating ten more"*
George Bernard Shaw

DANKWOORD

Hoewel het welslagen en het wetenschappelijk aanzien van een doctoraat er niet door bepaald wordt, heeft het dankwoord onmiskenbaar een belangrijke functie. Het is het stukje tekst waarin iedereen op zoek gaat naar literaire virtuositeit, een emotionele noot, een grappige kwinkslag, die ene dt-fout of een dankbetuiging aan hun adres. Vooral bij het ontbreken van dit laatste dreig ik een bron van verontwaardiging te vormen, en een doelwit van misnoegde blikken[1]. Bovenop deze hoge kwaliteitseisen komt dan nog de nood tot bondigheid (wat niet mijn sterkste punt is). Immers, de meesten onder u - al bent u al verder gegaan dan een snelle blik op de kaft en een goedkeurend "schoon boekske" of een opmerkzaam "precies dikker dan dat van Rodric" - haken wellicht na twee bladzijden af. Dus, als ik zonder verdere inleiding aan de eigenlijke bedankingen begin, kan u straks misschien nog een stukje van de samenvatting meepikken...

Ik prijs me gelukkig een innemende persoonlijkheid als Ben Immers als promotor gehad te hebben. Ben, je hartelijkheid ontlokte zelfs op de meest grijze dagen een glimlach en je aanstekelijk enthousiasme verspreidde zich voelbaar en hoorbaar doorheen de hele afdeling. Bedankt voor de aangename werkomgeving die je gecreëerd hebt, en het vertrouwen dat je steeds in mij hebt gesteld.
Aan mijn dagelijkse begeleider en promotor-op-de-valreep Chris Tampère, bedankt voor de bouwstenen die je mij voor dit doctoraat hebt aangereikt en voor de rol van meester-architect die je steeds voortreffelijk hebt vervuld. Je gaf me de vrijheid om op een zelfstandige manier aan onderzoek te doen. Tegelijkertijd heb je, ondanks je steeds voller lopende agenda, steeds tijd vrijgemaakt om samen het hoofd te breken over een nieuw opgedoken probleem, en om mee de lijnen uit te zetten op weg naar het volgende. Want hoewel een doctoraatsonderzoek soms een hobbelige weg is, heb ik hierdoor nooit – nu ja, zeer zelden – gevreesd dat ik het einde van de rit niet zou halen. Ik hoop dat ik ook in de toekomst het geluk zal hebben om te werken met motiverende mensen zoals jij.

[1] Hierbij wens ik dan ook mijn oprechte dank uit te drukken aan al wie ik hiernavolgend vergeten ben.

Special thanks go to the other members of my examination committee, Pierre Verbaeten, Dirk Cattrysse, Joos Vandewalle, Serge Hoogendoorn and Christine Buisson, who helped refining this dissertation to its current form. In that respect, I also express my gratitude to Gunnar Flötteröd for joining forces with me to tackle the node model problem. Gunnar, although your perseverance can indeed be annoying at times, I think it's one of the qualities that makes you one of the best traffic researchers of our generation (and certainly the one who can handle the most Leffes!).

Bedankt aan alle collega's en gerenommeerde receptie-crashers van CIB. In het bijzonder aan mijn directe (ex-)collega's, Jim, Thomas, Wouter, Wouter, Shengrun, Xin, Wei, Willem, Bart, Rodric, Marco, Francesco en Francesco: bedankt voor de vele interessante gesprekken (al dan niet verkeersgerelateerd) en voor de aangename werksfeer al die jaren. In het nog-net-iets-meer-bijzonder bedankt aan Rodric en Francesco Viti voor de talloze brainstormsessies die geholpen hebben mijn onderzoek te ontwikkelen, artikels bij te schaven en mijn kennis van verkeersmodellering te verbreden.

Familie en vrienden, sinds lang voordat ik mijn eerste verkeersstroomvergelijking op papier zette en tot lang nadat deze thesis uit het digitaal archief zal gedelete zijn (in de 20^{e} eeuw had hier gestaan "in de universiteitsbibliotheek onder het stof zal bedolven zijn"), ben ik jullie dankbaar; gewoon omdat jullie er waren. Jullie steun en gezelschap zijn privileges die ik enorm waardeer en koester. Een speciale vermelding hierbij gaat uit naar mijn ouders. Mams en paps, bedankt voor de troost in jullie stem bij twijfel en tegenslag, en voor de trots in jullie ogen.
Tenslotte, Sonja. Ik zou nog eens 250 bladzijden vol kunnen schrijven en er nog niet in slagen het gevoel te verwoorden jou aan mijn zijde te weten. Dankzij en met jou kijk ik gelukzalig achterom naar wat geweest is, en reikhalzend uit naar wat komen zal.

SUMMARY

The ever increasing demand for mobility causes traffic jams and delays on road networks all over the world, generating immense social and economic losses. Dynamic Network Loading (DNL) models represent the propagation of traffic and congestion. This PhD thesis focuses on macroscopic simulation-based DNL modelling based on first-order traffic flow theory (Lighthill & Whitman, 1955; Richards, 1956). These models support the decision-making of road managers and the information provision to road users.

The general aim of this PhD research is to further advance first-order macroscopic DNL models and their practical applicability. More specifically, this thesis pursues two quite distinct research directions:

1. Enhancing the theoretical knowledge and soundness as well as the realism and practical applicability of macroscopic DNL intersection models
2. Developing marginal DNL simulation models, which combine significant computation time savings with realistic congestion dynamics in repeated (iterative, finite difference or Monte-Carlo) DNL simulations

The function of the intersection model in the DNL model is twofold. Firstly, it must find a consistent solution for the traffic flows (veh/h) from each incoming to each outgoing link, considering the external constraints from these links (i.e. the local demands and supplies). The second function is to impose internal supply constraints, i.e. additional flow restrictions due to conflicts internal to the intersection (e.g. crossing streams hindering each other).
From a thorough literature review, it is concluded that the vast majority of existing models fails to properly fulfil the first function. In response, we compile a list of seven generic requirements for first-order macroscopic DNL intersection models and present a general intersection model (limited to external constraints) that meets all of these requirements.
Since most existing models neglect internal supply constraints, they are not well suited for busy urban and regional intersections. In this thesis, it is proposed to introduce internal supply constraints analogous to how external supply constraints are universally treated in the

state-of-the-art. This implies that the internal supplies are distributed according to the proportionality of predefined priority parameters of the incoming links. The possibility of non-unique solutions is identified (in general; not only in our model specification). It is found that solution uniqueness is only guaranteed if the priority parameters are single-valued. This implies that all movement flows from an incoming link have the same competitive strength for all the internal and external conflicts of the intersection. Since this uniqueness condition is intuitively contradictory to the observation that priorities often differ per conflict (for instance a straight movement usually has priority, while a left turn has not), it hinders the definition of the intersection model.
Specific intersection models for different types of intersections are developed that solve this ambiguity via a weighted pre-processing of different priority parameters per conflict into a single representative value. While there are still several ways to further improve these models, they are – to the best of our knowledge - the first to combine both functions of the intersection model into a unique, consistent solution.

Secondly, the novel concept of marginal DNL simulation is introduced. Marginal DNL algorithms are derived from a maternal base model, from which they adopt the modelling principles and (the majority of) the simulation algorithm. They perform partial (marginal) simulations of local variations to a base scenario, rather than running full simulations. This provides a considerable computational advantage if many repeated simulations with large overlap need to be performed.
In this thesis, two marginal DNL algorithms are proposed. For both, the maternal base model is the Link Transmission Model (LTM) of Yperman (2007). Hence, the congestion dynamics in both algorithms are consistent with first-order traffic flow theory as in Newell (1993).
The first, the Marginal Incident Computation (MIC) algorithm, is designed for fast Monte-Carlo simulation of incidents. Compared to LTM, MIC may reduce the computation time to less than 1 % (depending on the network size), at the cost of acceptable approximation errors (of aggregated outputs such as vehicle hours lost).
Secondly, the Marginal Computation (MaC) algorithm is presented. MaC has an extended functionality (both demand and supply variations) and higher accuracy compared to MIC, enabling analysis of fine-grained output such as (link) flows. Possible applications include variability studies, optimization problems such as dynamic origin-destination estimation and optimal control, robust network design and (real-time) dynamic traffic management support; all of which are currently infeasible or at least highly computationally demanding.
Finally, the applicability of marginal simulation can be enhanced by tailoring specific marginal algorithms for specific purposes. Moreover, we estimate that the advantage of marginal simulation could carry over to other research domains with similar needs (for instance pedestrian modelling and supply chain management). In fact, it has turned out that very similar techniques have been used for the design of digital hardware circuits (Hwang et al., 1988; Salz & Horowitz, 1989).

SAMENVATTING

De steeds toenemende vraag naar mobiliteit zorgt voor vertragingen en files op verkeersnetwerken wereldwijd, wat enorme sociale en economische verliezen met zich meebrengt. Met dynamische verkeerspropagatie (DNL) modellen kunnen verkeersstromen en congestie gemodelleerd worden. Dit doctoraatsonderzoek richt zich op macroscopische DNL simulatiemodellen gestoeld op eerste-orde verkeersstroomtheorie (Lighthill & Whitman, 1955; Richards, 1956). Deze modellen ondersteunen de beslissingen van wegbeheerders (bv. i.v.m. het inzetten van verkeersmanagement maatregelen) en de informatievoorziening naar de weggebruiker toe.

Het algemene objectief van dit doctoraatsonderzoek is het verder ontwikkelen van eerste-orde macroscopische DNL simulatiemodellen en het vergroten van hun praktische toepasbaarheid. Meer specifiek volgt deze thesis twee vrij gescheiden onderzoeksrichtingen:

1. Het verbeteren van de theoretische kennis over en de degelijkheid van het macroscopische kruispuntmodel in DNL simulatiemodellen, alsook hun realiteitsgehalte en praktische inzetbaarheid.
2. Het ontwikkelen van zogenaamde marginale DNL simulatiemodellen, die significantie rekentijdwinsten combineren met een realistische modellering van congestie in herhaalde (bv. iteratieve of Monte-Carlo) DNL simulaties.

Het kruispuntmodel heeft twee functies in het DNL model. Ten eerste dient het een consistente oplossing te vinden voor de verkeersstromen (vtg/u) van elke ingaande schakel naar elke uitgaande schakel van het kruispunt, rekening houdend met de externe beperkingen opgelegd vanuit deze schakels (m.n. de vraag- en aanbod-beperkingen). De tweede functie bestaat erin bijkomende, interne aanbod-beperkingen op te leggen, afkomstig van conflicten op het kruispunt zelf (bv. vanwege kruisende stromen die elkaar hinderen).
Na een gedetailleerd literatuuroverzicht, concluderen we dat het merendeel van de bestaande kruispuntmodellen de eerste functie niet goed vervult. In deze thesis wordt een lijst opgesteld van zeven vereisten die de correcte uitvoering van de eerste functie waarborgt. Tevens

ontwikkelen we een algemeen kruispuntmodel (met voorlopig enkel externe beperkingen) dat aan elk van deze vereisten voldoet.
De meeste bestaande modellen verwaarlozen de interne aanbod-beperkingen, waardoor ze niet geschikt zijn voor drukke stedelijke en regionale kruispunten. In deze thesis worden deze interne beperkingen geïntroduceerd analoog aan de externe aanbodbeperkingen, en wel zodanig dat het aanbod van een intern conflict verdeeld wordt aan de hand van voorgedefinieerde prioriteiten van de ingaande schakels. Het blijkt dat de oplossing van het kruispuntmodel niet altijd uniek is (niet enkel in onze modelspecificatie), tenzij slechts één prioriteit per ingaande schakel geldt. Deze voorwaarde strookt echter niet met de werkelijkheid, aangezien verschillende stromen uit een ingaande schakel vaak een verschillende prioriteit hebben in verschillende conflicten (bv. rechtdoor heeft voorrang; linksaf niet). Dit vormt een hinderpaal bij de verdere ontwikkeling van het kruispuntmodel.
Tenslotte worden kruispuntmodellen voor specifieke kruispunttypes voorgesteld die bovenstaand probleem oplossen door de verschillende prioriteiten per conflict te wegen tot één uiteindelijke prioriteit. Hoewel er nog verschillende mogelijkheden zijn om deze modellen verder te verbeteren, zijn dit de eerste modellen die beide functies van het kruispuntmodel vervullen en een unieke, consistente oplossing leveren.

Een tweede onderwerp in deze thesis betreft marginale DNL simulatie. Marginale DNL simulatiealgoritmes worden afgeleid van een basis moedermodel, waarvan ze de modelprincipes en (het merendeel van) het algoritme erven. Ze voeren gedeeltelijke (marginale) simulaties uit in de vorm van lokale variaties op een basisscenario, in plaats van volledige simulaties. Dit brengt een grote rekentijdwinst met zich mee wanneer vele grotendeels overlappende simulaties uitgevoerd dienen te worden.
Twee marginale DNL algoritmes zijn ontwikkeld. Voor beide is het basis moedermodel het Link Transmissie Model (LTM) van Yperman (2007). Dit betekent dat de modellering van congestie consistent is met eerste-orde verkeersstroomtheorie zoals beschreven door Newell (1993). Het eerste algoritme, het Marginal Incident Computation (MIC) algoritme, is ontworpen met het oog op snelle Monte-Carlo simulatie van incidenten. Vergeleken met LTM, kan MIC de rekentijd reduceren tot minder dan 1 % (afhankelijk van de grootte van het netwerk), en dit ten koste van beperkte benaderingsfouten (van geaggregeerde output zoals voertuigverliesuren). Het tweede is het Marginal Computation (MaC) algoritme. MaC heeft een uitgebreidere functionaliteit (zowel variaties in vraag als in aanbod) en een grotere nauwkeurigheid dan MIC, wat analyses van de verkeersstromen op individuele schakels toelaat (naast analyses van geaggregeerde output). Mogelijke toepassingen zijn o.a. betrouwbaarheidstudies, het numeriek bepalen van gradiënten voor optimalisatieproblemen zoals dynamische herkomst-bestemmingsschatting en evacuatieplanning, robuust netwerkontwerp en ondersteuning van dynamisch verkeersmanagement. Deze toepassingen zijn momenteel vaak onhaalbaar omdat ze zeer rekenintensief zijn.
Tenslotte kan de toepasbaarheid van marginale simulatie vergroot worden door meer marginale algoritmes te ontwikkelen voor specifieke doeleinden. Onze inschatting is dat dit niet beperkt hoeft te blijven tot wegverkeer, maar dat marginale simulatie ook voordelig kan zijn in gelijkaardige onderzoeksdomeinen zoals voetgangersmodellering en productieprocesbeheer. Het is bovendien gebleken dat zeer gelijkaardige technieken reeds zijn toegepast op het ontwerp van digitale hardware (Hwang et al., 1988; Salz & Horowitz, 1989).

NOTATION

Acronyms

AWSC	All-Way-Stop-Controlled
CDF	Cumulative Distribution Function
CFL	Courant-Friedrichs-Lewy
CTF	Conservation of Turning Fractions
CTM	Cell Transmission Model
CVN	Cumulative Vehicle Numbers
DNL	Dynamic Network Loading
DTA	Dynamic Traffic Assignment
DTD	Day-To-Day
DTM	Dynamic Traffic Management
DUE	Deterministic User Equilibrium
DUO	Dynamic User Optimal
FIFO	First-In-First-Out
LTM	Link Transmission Model
MaC	Marginal Computation (model)
MC	Multi-Commodity
MIC	Marginal Incident Computation (model)
MSA	Method of Successive Averages
OD	Origin-Destination
PTTR	Priority-To-The-Right
PUE	Probabilistic User Equilibrium
RAF	Rightfully Affected Flows
SC	Single-Commodity
SCIR	Supply Constraint Interaction Rules
SO	System Optimal
SUE	Stochastic User Equilibrium
TTS	Total Time Spent

TTV Travel Time Variability
VHL Vehicle Hours Lost
WAF Wrongfully Affected Flows
WUF Wrongfully Unaffected Flows

Symbols

α_i single-valued priority parameter determining competitive strength of link i for all conflicts
α_{ij} priority parameter determining competitive strength of link i for supply R_j
α_{ik} priority parameter determining competitive strength of link i for internal supply N_k
$\alpha_{ij,P}$ priority parameter determining competitive strength of link i for supply $R_{j,P}$ during combination of green phases P
$\alpha_{ik,P}$ priority parameter determining competitive strength of link i for internal supply $N_{k,P}$ during combination of green phases P
a_j reduction factor due to supply R_j
C_i capacity of link i (veh/h)
C_{ij} oriented capacity of link i to link j (veh/h)
c_r travel cost for a route r
ε_{up} accuracy threshold below which upstream moving flow changes are neglected (%)
ε_{down} accuracy threshold below which downstream moving flow changes are neglected (%)
f_{ij} turning fraction from link i to link j
f_{ik} (turning) fraction from link i passing internal conflict k
G_i fraction of total cycle time during which the light is green for link i
G_j fraction of total cycle time during which traffic flows to link j
G_p green fraction of total cycle time of green phase p
i incoming (upstream) link to an intersection
I number of incoming links of an intersection
j outgoing (downstream) link to an intersection; external conflict
J number of outgoing links of an intersection
$j^*(i)$ outgoing link of which the supply is (most) restrictive to incoming link i
k internal conflict (in part I – Intersection Modelling)
K number of internal conflicts of an intersection
k density (veh/km) (in part II – Marginal Simulation)
k_c critical density at capacity (veh/km)
k_j (maximum) jam density (veh/km)
L length of a link (km)
n intersection or node
N_k internal supply available in conflict k
$\widehat{N}_k$ internal supply constraint function
$N_{k,p}$ partial supply of conflict k during green phase p
$N(x,t)$ cumulative vehicle number at location x at time t (veh)
p green phase

P	combination of green phases
p_r	route fraction of OD demand choosing route r
q	traffic flow (veh/h)
q_i	traffic flow from link i (veh/h)
q_{ij}	partial flow from link i to link j (veh/h)
q_{ik}	partial flow from link i passing conflict k (veh/h)
q_r	flow or demand of route r (veh/h)
r	route
R_j	external supply available in link j (veh/h)
$\widehat{R}_j$	external supply constraint function
$\widetilde{R}_j$	reduced supply of link j that remains after links i not constrained by j have taken their share (veh/h)
R_j^i	rightful share of link i of supply R_j (veh/h)
$R_{j,p}$	partial supply of link j during green phase p (veh/h)
S_i	demand of link i (veh/h)
S_{ij}	partial demand of link i to link j (veh/h)
S_{ik}	partial demand of link i passing conflict k (veh/h)
t_{ik}	consumption time during which one vehicle of link i occupies conflict k (s)
T	turning fraction interval (min)
tt	(instantaneous or experienced) travel time (h or min)
U_j	set containing links constrained by supply constraint function $\widehat{R}_j$
U_k	set containing links constrained by internal supply constraint function $\widehat{N}_k$
v_f	free flow speed (km/h)
w	maximum negative shockwave speed (km/h)
w_s	negative shockwave speed (km/h)
w_{ij}	weight given to priority parameter α_{ij}
w_{ik}	weight given to priority parameter α_{ik}

CONTENTS

1
INTRODUCTION

This chapter provides a general introduction on Dynamic Traffic Assignment (DTA) and Dynamic Network Loading (DNL) models. Only uni-modal vehicular road traffic is considered; other transportation methods and modes such as pedestrians, bicycles and public transportation are not included in our research.
The intention of this introductory chapter is to create an understanding of the broad context in which the specific research topics of this thesis (which itself will be introduced in detail in subsequent chapters) are situated. We do not aim to provide an extensive literature overview. Rather, citations are limited to pioneering work and discussion and survey papers that are good starting references for additional reading.

Section 1.1 first briefly mentions the well-known problems in road traffic and potential measures from road authorities and services to road users. DTA and DNL models are useful tools in this context and are discussed thereafter. Section 1.2 familiarizes the reader with the objectives and scope of this PhD research, while Section 1.3 clarifies its contributions. Finally, Section 1.4 provides an overview of the subsequent chapters of this thesis.

1.1 Background and context

The ever growing demand for mobility increases the pressure on road networks all over the world. Even without considering externalities like the impact on safety, liveability and the environment and only looking at traffic jams and delays, the social and economic losses are immense. In response, both road users and road authorities and managers take actions to minimize these losses. For road authorities, this includes improving or expanding the infrastructure, providing public transportation and carpool facilities as alternatives to driving (individually), and efficiently managing the traffic flows through travel demand management or (Dynamic) Traffic Management (DTM) measures such as ramp metering, route guidance, (dynamic) road pricing, high occupancy vehicle lanes, dynamic hard shoulder running and signal coordination. Meanwhile, road users optimize their personal use of the available road network and transportation systems, partially based on information and (route) guidance provided to them via Variable Message Signs (VMS) or in-vehicle systems (GPS, smart phone applications). All of the above and many more factors directly or indirectly determine the capacity of the available road network (called supply) and the aggregated travellers' demand, i.e. the magnitude of the traffic streams and when (departure time), where (origin, destination, route) and how (transportation mode, vehicle type) they want to make use of the network. Under these complex and highly interrelated circumstances, traffic assignment and propagation models are indispensable tools that support the decision-making of road authorities and managers and information provision to road users. These models can be applied for mapping and analyzing the current and expected traffic conditions, as well as for testing and optimizing improvements to (the management of) the existing network.
The focus of this dissertation is on the dynamic traffic propagation or Dynamic Network Loading (DNL) model. The DNL model can be applied by itself - mainly in (online) estimation and prediction applications - or as a part of a Dynamic Traffic Assignment (DTA) model. This section first continues with a short discussion on DTA in general. Then, we elaborate on DNL; more precisely, macroscopic simulation-based DNL, which is where our research is situated.

1.1.1 Dynamic Traffic Assignment (DTA)

While the earlier developed static assignment models are still often used, also practitioners are becoming increasingly aware of the advantages of DTA models, following scientific advances since the pioneering work of Merchant & Nemhauser (1978a, 1978b). DTA models, contrary to static models, are time-dependent. Hence, they give a much more realistic representation of congested traffic conditions, which may be highly time-variable.
In the following, a brief introduction on DTA is provided. First, the purpose and typical composition of a DTA model is explained. Then, various characteristics that are often used to categorize DTA models are discussed.

1.1.1.1 Components of the DTA model

Simply put, the purpose of a DTA model is to determine the traffic conditions in a given network for a given travel demand. The most important steps in the DTA problem are illustrated in Figure 1-1, taken from chapter 7 in Cascetta (2001).

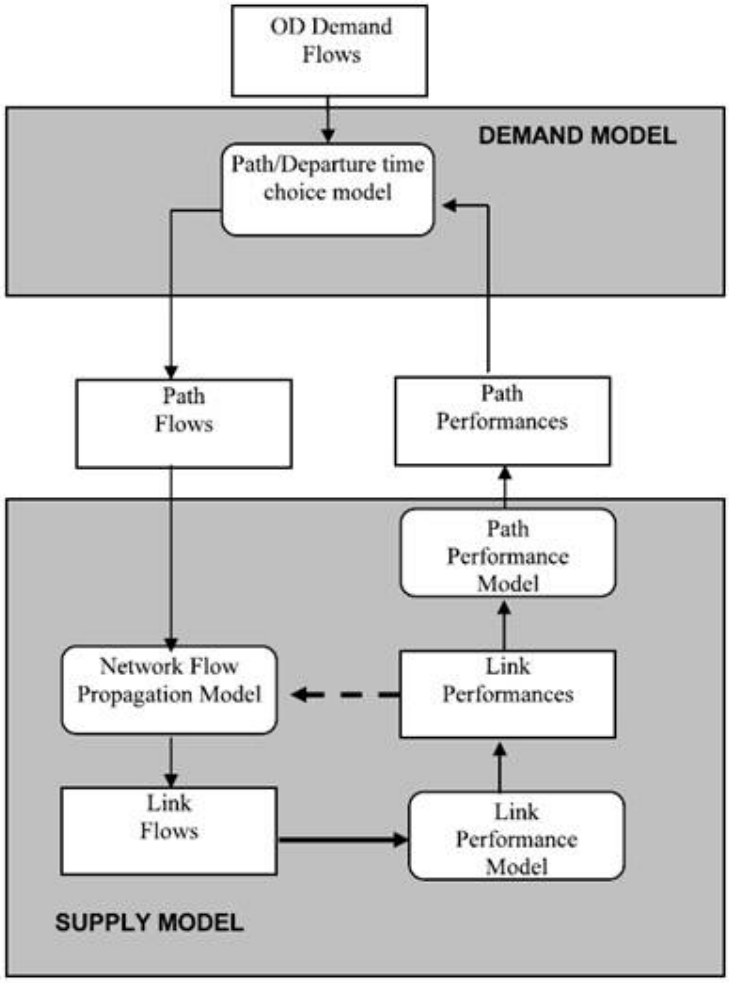

Figure 1-1: General framework of the DTA problem (Cascetta, 2001)

In this dissertation, the discussion of the DTA model is limited to a brief, general introduction. Generally speaking, two main components can be identified in DTA models, namely the choice model (i.e. the demand model in Figure 1-1) and the propagation or DNL model (i.e. the supply model in Figure 1-1).
The extent to which the two are separated or interwoven differs among various types of DTA models. Below, these two main model components as well as the typical input and output of a DTA model are briefly discussed[2].

- Input:
  Input for the DTA model is needed from the demand and supply side. The geometry and characteristics of the network constitute the supply. The network consists of stretches of road (links) and highway on- and off-ramps and intersections (nodes).

[2] This outline is aimed to be general and simple. Therefore, it does not correspond entirely with each specific DTA model. For instance, some DTA models may include a control component that governs the system response to the travel costs in the network in the form of (D)TM actions such as ramp metering and road pricing. However, in its present form, this outline serves its purpose of creating a general understanding of DTA.

The input that is required to fully specify the network depends on the specific type of DTA model. This may include for instance link lengths, capacities, desired speeds, link-node configurations, etc.
The travel demand is the amount of traffic that wishes to make use of the network in the considered time period. Typically, this is represented by an Origin-Destination (OD) matrix, indicating the desired flow from each origin to each destination. Dependent on whether or not departure time choice is included in the choice model, the required input may be a static or time-dependent OD matrix (usually the latter).
Determining and calibrating all the necessary input may be a difficult task for practical applications. In particular, dynamic OD estimation forms a separate research domain on its own.

- Choice model:
  The purpose of the choice model is to represent the decisions that travellers make in response to the travel costs of their desired trip. Since the travel costs are determined by the DNL model, the choice and DNL model are either solved concurrently or iteratively. Usually, the choice model is limited to a route choice model that assigns (groups of) vehicles to specific routes between each origin and each destination. In addition, a departure time choice model may be included[3]. Thus, the choice model divides the total travel demand into dynamic route demands.

- DNL model:
  This model propagates the route demands determined by the choice model through the network. The traffic flows are calculated in time and space, congestion may form, delays may be encountered and the travel times (or, more generally, the travel costs) on the network are determined.

- Output:
  The output of the DTA model ultimately is that of the DNL model, once correspondence between the choice and DNL model has been obtained. Different correspondences may be assumed, however (see Section 1.1.1.2).
  The output primarily consists of the dynamic traffic flows - sometimes presented in the form of Cumulative Vehicle Numbers (CVN) – and travel times (costs) on the network during the considered time period. Other results may be derived, for instance Total Time Spent (TTS), Vehicle Hours Lost (VHL) and – via post-processing models – fuel consumption, emissions, etc.

A large number of research directions have been pursued in the vast literature involving DTA. Yet, a lot of further research is still necessary to further analyze and develop all of the above components. As this thesis focuses on the DNL model, the other components are discussed if necessary in the context, but not studied or further developed.

[3] In multimodal models, also mode choice is to be included.

1.1.1.2 Categorization of DTA models

Before zooming in on the DNL model, this section briefly familiarizes the reader with different types of DTA models. While DTA models are still discussed as a whole here, the DNL model is obviously also characterized by the addressed features. Hence, afterwards, this categorization allows defining the scope of our research more precisely.
DTA models exist in many forms. In the literature, the following are the three most common categorizations of different DTA approaches. More details and literature reviews can be found in discussion papers such as Peeta & Ziliaskopoulos (2001) and Viti & Tampère (2010).

1. Type of assignment:
   This refers to the correspondence that is assumed between the driver decisions determined by the choice model and the travel costs determined by the DNL model. A common assumption is that each driver optimizes his own travel cost, so that driver decisions and travel costs are in a Deterministic User Equilibrium (DUE). This is an extension of the static UE corresponding to the first principle of Wardrop (1952), expressing a state in which no driver can individually lower his travel cost by changing his route. A popular variant is the Stochastic User Equilibrium (SUE), introduced by Daganzo & Sheffi (1977), in which drivers optimize their perceived travel cost. The SUE thus accounts for human judgment errors and heterogeneity among drivers.
   System Optimal (SO) assignment was also adopted from earlier static approaches. SO expresses a state in which the total travel cost on the network is minimized. SO assignment is thus useful to indicate the ideal situation in the network (from the system's perspective) and directions to improve towards it. This SO situation (or at least the directions for improvement) could then be pursued with control strategies such as road pricing and traffic guidance.

   UE (and SO) are the traditional types of assignment, in which the driver decisions are based on the actually experienced travel costs, i.e. the costs that will be experienced given the choice that is made. This means that the choice model and the DNL model are either solved concurrently, or as two separate models that are iterated until convergence. In the past two decades, advances to two non-equilibrium approaches are made.
   The first is reactive or en-route assignment - also referred to as Dynamic User Optimal (DUO) assignment - introduced by Ran et al. (1993). In reactive assignment, drivers' route choice is made at decision points en-route rather than only at departure. The choices minimize instantaneous travel costs instead of experienced costs. The instantaneous travel cost on a route is the cost that would arise if the current traffic conditions would remain unchanged for the remainder of the trip.
   The second non-equilibrium assignment is called dynamic process or doubly dynamic assignment (see e.g. Cantarella & Cascetta, 1995). In dynamic process models, day-to-day (DTD) dynamics are explicitly considered. Drivers make adjustments in a learning choice model, based on past experiences. Hence, decisions are not based on (to be) experienced travel costs, but (expected costs predicted from) past costs and

possibly information available to the drivers. Contrary to the traditional UE assignment, not only a fixed (equilibrium) state is sought. Also the DTD evolution itself can be studied. Depending on the model assumptions and initialization, this dynamic process may or may not convergence to an equilibrium state as found in the traditional assignments, or settle into a periodic motion (see Watling & Hazelton, 2003).

The properties, similarities, differences and appropriateness of the different types of assignment have been and continue to be studied and debated by the research community. The traditional equilibrium assignment is considered simpler and familiar and is therefore still by far the most popular. Yet, it is well-known that the existence and uniqueness of the solution cannot be guaranteed in general networks; the earliest study of Smith (1979) is followed by many more. Regardless of these theoretical issues, practical objections are increasingly voiced against the UE assumption (see e.g. Tampère & Viti, 2010). Indeed, UE intuitively corresponds to modelling the drivers' response to well-known, typical conditions, i.e. a 'typical day'. However, travel demand (e.g. due to random daily fluctuations, mass events) and network supply (e.g. due to weather conditions and incidents) and in result traffic conditions are often highly variable from one day to another. Traditionally, DTA models are deterministic and thus ignore this DTD variability. However, an evolution towards stochastic DTA is necessary and currently pursued by various researchers (see Appendix F.1). This will render network analysis and planning much richer and more meaningful, and DTM more effective. Rendering the DNL model employable in such a stochastic DTA to account for variability has been a starting motivation of our doctoral research. The implications for the type of assignment - mainly whether or not UE is a defendable assumption in stochastic DTA - are not studied in this thesis. Still, it seems safe to say that a dynamic process approach - likely to be combined with reactive assignment - has strong potential in this context.

2. Solution methodology:
   The distinction between analytical and simulation-based DTA models mainly connotes a difference in solution methodology, although it obviously also affects the problem formulation.
   An analytical DTA model is usually based on the transformation of the UE problem into an equivalent variational inequality (see e.g. Friesz et al., 1993) or optimization problem (based on Beckmann et al., 1956). The advantage is twofold. On the one hand, these equivalent mathematical formulations allow for formal proofs of existence and uniqueness of the equilibrium flow patterns. On the other hand, universal solution algorithms from the broader domains of optimization and variational inequalities can be adopted to solve the DTA problem. However, for real-world networks, the mathematical solvability (to guarantee solution existence and uniqueness) and computational feasibility of this analytical approach requires simplifications to the representation of traffic propagation and congestion dynamics (especially with regard to the intersection model). Because of this, analytical models are mainly useful for gaining theoretical insight.

Simulation-based models disregard rather than solve the theoretical issues of solution existence and uniqueness. However, they can circumvent the problems regarding realism that exist in analytical models by solving the problem step-by-step and piece-by-piece. Hence, simulation-based models discretize time (in simulation update steps)[4]. Because of their higher potential realism, simulation-based models have the advantage for real-world applications. However, this does not imply that computation time is not an issue. Large-scale problems are usually computationally cumbersome or even infeasible.

3. Aggregation level:
   Simulation-based DTA models can be further subdivided into microscopic, mesoscopic and macroscopic models – for a detailed overview, see Hoogendoorn & Bovy (2001). Microscopic models - for example VISSIM, see e.g. Fellendorf & Vortisch (2010) – assign and propagate traffic as individual vehicles. Since microscopic models are computationally demanding and usually require a significant calibration effort due to the large number of parameters, they are generally not well-suited for large-scale problems. Macroscopic models – see e.g. Lebacque (1996), the Cell Transmission Model (CTM) of Daganzo (1994, 1995) and the Link Transmission Model (LTM) of Yperman (2007) - represent traffic as continuous streams. They do not distinguish between individual vehicles. While this aggregated approach may be insufficient for detailed, small-scale applications, it is clearly advantageous in large-scale networks. Mesoscopic models – e.g. DynaMIT (Ben-Akiva et al., 1998) – describe the behaviour of individuals in an aggregate way, for instance using probability distribution functions (Hoogendoorn & Bovy, 2001). Hence, the underlying behavioural rules stem from the aggregate level, whereas the representation is in terms of individual vehicles (or packets of vehicles).

While the above constitute the most common categorizations of DTA models, there are other ways to make a further distinction. Deterministic and stochastic DTA models have been mentioned above. Furthermore, models can account for a single user class or multiple user classes (for instance separating vehicle types or groups of drivers with substantially different choice behaviour).
The DNL model in itself can be further specified. This is explained in the next section, with the focus limited to macroscopic, simulation-based DNL.

1.1.2 Macroscopic simulation-based Dynamic Network Loading (DNL)

Corresponding to the characteristics of the DTA model, different types of DNL models can be distinguished. In this thesis, the focus is on macroscopic simulation-based DNL. This category of models can be further subdivided, as different theoretical bases exist to describe the propagation of traffic and congestion on links. A brief summary is given below; for more details and references, we refer to the overview papers of Nie & Zhang (2005) and Mun (2007).

[4] Analytical models on the other hand can be either time-discrete or time-continuous.

- Exit flow functions:
  In DNL models applying exit flow functions, the outflow (rate) of a link at a certain time is defined as a function of the instantaneous amount of traffic on the entire link. An important drawback of such models is their violation of causality. This means that the propagation of vehicles is not only influenced by downstream traffic, but also by upstream traffic that entered the link at a later time. Hence, these models do not represent traffic propagation and congestion in a realistic way.

- Link travel time functions:
  Travel time functions propagate traffic by assigning a link travel time to a group of vehicles upon entry of the link. This travel time depends on the inflow or amount of traffic on the link. Congestion dynamics are thus not explicitly modelled. Hence, these models are insufficiently realistic in (heavily) congested conditions. Moreover, in case of non-linear travel time functions – which are the more realistic assumption – the First-In-First-Out (FIFO) principle may be violated. The FIFO principle states that vehicles exit a link in the same order as they entered it. While this should not necessarily hold in free flowing conditions (overtaking), this is universally considered as a necessary property when modelling congestion in a DNL model.

  While these first two approaches are mostly used in analytical DNL models, they are also employable in simulation-based models. For instance, travel time functions are one of the modelling options in INDY (Bliemer et al., 2004).

- Vertical queuing:
  Contrary to the previous two types of models, vertical queuing models consider an outflow and/or inflow capacity, i.e. a maximum flow (veh/h) that may exit or enter a link. If this capacity is exceeded by the desired flow (i.e. the local demand), congestion is modelled as a dimensionless, vertical queue. Hence, congestion spillback over a link and further upstream onto other links is not accounted for. Therefore, these models are still insufficiently realistic for applications involving congested networks.

- Horizontal queuing:
  Horizontal queuing models physically account for congestion. Their realism is still limited (although higher than that of vertical queuing models). While the queue tail can move in the up- or downstream direction to model how congestion grows or dissolves, the queue head is always located at the end of a link. Moreover, the density in the congested part of the link is typically limited to one fixed value. This may cause unrealistic congestion spillback behaviour. Indeed, in reality, the head of a queue may propagate upstream as congestion dissolves, and a wide range of densities is possible in congestion. Congestion dynamics and spillback are more adequately modelled by traffic flow theory.

- First-order traffic flow theory:
  Inspired by the similarities with compressible fluids, traffic flow or kinematic wave theory was introduced by Lighthill & Whitham (1955) and Richards (1956). Traffic

states at a location x and time t are defined by three macroscopic variables, namely the flow (rate) or intensity q (veh/h), the density k (veh/km) and speed v (km/h). These variables are related as in (1.1):

$$q(x,t) = k(x,t)v(x,t) \tag{1.1}$$

For notational convenience, *(x,t)* is omitted in the remainder. Secondly, conservation of vehicles holds, expressed by:

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{1.2}$$

Finally, a fundamental relationship is assumed - represented by a fundamental diagram - between q, k and v; only two of which are independent due to (1.1). Often, a triangular-shaped diagram is assumed for this fundamental relationship (see Figure A-2 in Appendix A for an example). In first-order traffic flow theory, the fundamental relationship is assumed to be stationary. Hence, each tempo-spatial traffic state can be expressed by this stationary relationship and is located on the fundamental diagram.
Combination of the above basic relations leads to a partial differential equation with only one independent variable, e.g. the density k. With given initial and boundary conditions, this can be solved in space and time. This solution describes the propagation of traffic (states) in the form of characteristic waves with constant q, k and v. Different traffic states are separated by shock (or rarefraction) waves (notably congestion fronts) that may propagate up- or downstream.

DNL models adopting first-order traffic flow theory are universally acknowledged to represent traffic propagation and congestion dynamics more realistically than the aforementioned model types. On the downside, their computation time and memory use is substantially larger. The best known first-order DNL model is CTM (Daganzo, 1994). More recently, Yperman (2007) developed the Link Transmission Model (LTM). LTM adopts the simplified solution procedure to first-order traffic flow theory of Newell (1993), which is based on constructing and evaluating CVN curves. Readers unfamiliar with LTM or Newell's theory are referred to Appendix A.

- Second-order traffic flow theory:
  Second-order traffic flow theory - first introduced by Payne (1971) - differs from first-order theory in that the fundamental relationship is considered to be not a stationary, but only an equilibrium relation. Transition states are incorporated by assuming (for instance) a relaxation equation between k and v. Additional traffic phenomena such as acceleration and deceleration and traffic instabilities (e.g. in the form of stop-and-go waves) can be modelled, which are not included in first-order models. However, there is debate regarding how realistic these higher-order phenomena are reproduced. Some researchers adhere to the superiority of second-order DNL models – such as METANET (Messmer & Papageorgiou, 1990) -, whereas others claim that the simpler first-order models are just as (or in fact more)

realistic for most applications (when second-order phenomena are not relevant, for instance in traffic flows interrupted by intersections). Moreover, second-order models may have difficulties satisfying causality[5]. The discussion on the (dis)advantages of first- versus second-order models is not continued in this thesis. A good starting point in that matter is the discussion paper of Lebacque & Lesort (1999).

In conclusion, models based on traffic flow theory are more realistic (particularly regarding the representation of congestion dynamics) than simpler approaches such as travel time functions or vertical queuing. Simpler models, on the other hand, have the advantage of lower computation time and memory use. The choice of a proper model should be made in the context of the research aim or application.

More subdivisions between DNL models can be made, for instance between single-lane and multi-lane and Single-Commodity (SC) and Multi-Commodity (MC) DNL. The latter approach keeps track of separate flows during propagation, disaggregated by route or by destination. The former does not retain information regarding route or destination, but only considers one aggregate flow and governs propagation by means of turning fractions at the link ends. This is computationally advantageous but requires additional considerations in a DTA environment. More details on SC DNL are provided in Appendix B.

Finally, while the subdivisions discussed above are generally considered to correspond to the DNL model as a whole, they mainly concern the traffic propagation on the links of the network. This is governed by a submodel of the DNL model called the link model. However, if capacity inflow restrictions are considered – without which a realistic modelling of congestion is generally not possible - also an intersection or node model is needed[6]. The first function of the intersection model is then to find a solution for the flows from each incoming to each outgoing link, considering the constraints imposed by the maximum possible inflows into the outgoing links (the local supplies) and the maximum possible outflows from the incoming links (the local demands). A second function of the intersection model, which is not less important in busy urban and regional networks, is to impose additional flow restrictions due to limited supply within the intersection itself. The vast majority of existing DNL intersection models does not encompass the latter. This is because macroscopic simulation-based DNL models have been originally designed for simulation of highway networks. These include mainly simple three-legged nodes (merges and diverges) to represent on- and off-ramps, to which internal supply does not apply. As a result, much more attention has been devoted to modelling traffic propagation and congestion on the links, leaving the DNL intersection model rather undeveloped. For a more profound problem statement, we refer to Chapter 2. Indeed, the first main goal of this thesis is to advance the macroscopic modelling of intersections.

[5] A violation of causality means that the propagation of vehicles may be influenced by upstream vehicles.
[6] In absence of capacity inflow restrictions, the outflows of the incoming links of an intersection into each outgoing direction merely have to be added up to determine the inflows into the outgoing links. This can hardly be considered an intersection model.

1.2 Objectives and scope

1.2.1 Objectives

The objective of this PhD research is twofold:

1. To further advance the modelling of intersections in macroscopic DNL. This constitutes enhancing the theoretical soundness and knowledge as well as the realism and practical applicability of these intersection models, particularly in (congested) urban and regional networks.

2. To unite two important challenges in DNL, namely to limit computation time and to realistically capture congestion dynamics. This is particularly important in applications that require a large number of DNL simulations to be performed. We note that originally, the objective was more narrowly defined, namely to improve the suitability of simulation-based DNL for the stochastic modelling of the variable traffic conditions.

While these two topics appear quite distinct (which, technically, they are), the developments in this thesis are united in their general aim to improve state-of-the-art macroscopic DNL modelling and the practical applicability of these models.

1.2.2 Scope

The focus in this PhD research is on the DNL model. The choice component of the DTA model and the type of assignment are not specifically studied and only discussed when necessary in the context. Route choice, as well as other necessary input such as a dynamic OD matrix and network characteristics are considered externally given to the DNL model, unless stated otherwise.

As explained earlier, many different types of DNL models exist. We do not aim to evaluate or compare these different types. Our focus is on macroscopic simulation-based DNL based on first-order traffic flow theory. While at some points the superiority of this type of DNL models may be claimed in a specific context, this does not imply that they are universally preferable. In general, we can state that, on the one hand, the improved congestion dynamics render them advantageous to simpler model types for applications in congested networks. On the other hand, for many applications, the additional capabilities of more complex models are infeasible (e.g. large-scale problems for microscopic models) or unnecessary (e.g. in an urban environment with traffic flows interrupted by intersections, second-order traffic flow phenomena (instability, capacity drop, stop-and-go traffic) are not very relevant). Consequently, first-order macroscopic simulation-based DNL models are deemed the simplest models that are sufficiently realistic for a wide range of applications in congested networks. Hence, it makes sense to let research that is fundamental (intersection modelling) or entirely new to transportation modelling (marginal simulation) depart here and explore extensions to more complex models in future research.

Departing from previous research at our department (Yperman, 2007), LTM in particular forms a theoretical and practical starting point for the presented developments. Below, four delineations are listed that further define the scope of our research. The first three of these are adopted from LTM.

- Link-based:
  Many DNL models in the literature apply a spatial discretization based on cells (smaller than links), notably the CTM of Daganzo (1994) and variants thereof. We adopt the approach of LTM throughout this thesis, i.e. the spatial discretization is equal to the links. Extensions to cell-based models are not discussed, but are usually straightforward.

- Single-class:
  Multiple user classes, which differentiate for example between different types of vehicles or driver preferences, are not considered.

- Single-lane:
  The number of lanes on a link is not specified, so that links are considered to have a homogeneous cross-section (with a specific capacity that reflects the number of lanes that exist in reality of course).

- Multi-Commodity (MC) and Single-Commodity (SC):
  While LTM is a MC DNL model that keeps track of route flows, the marginal DNL models developed in this thesis are SC, which means they consider traffic as one 'undirected' flow on the link level which is propagated according to turning fractions at link ends. However, for the intersection model, this makes no difference and implementation in both MC and SC DNL models is possible.

1.3 Contributions

The contributions of this thesis are situated along the two research paths that are pursued:

1. Intersection modelling:
   The contribution of this part is primarily theoretical. An advanced understanding is developed of the necessary modelling properties and potential problems. This theoretical knowledge is translated into intersection models. More specifically, the main contributions are:

   - From the analysis of returning shortcomings in state-of-the-art models, a set of seven requirements is composed with which first-order DNL intersection models should comply in order to produce a consistent solution.

   - A general intersection model is presented that complies with these seven requirements. This model, however, still only considers the demand and external supply constraints of the adjacent links.

- Internal supply constraints, accounting for limited supply within the intersection itself (due to internal conflicts, for instance between crossing flows), are introduced into the model in a general way. It is shown that the solution of the intersection model may be non-unique. A uniqueness condition is found that does not intuitively correspond to a realistic representation of driver behaviour. On the other hand, a unique solution is necessary to allow using the intersection model in state-of-the-art deterministic DNL models. Hence, the finding of solution non-uniqueness has fundamental as well as practical implications.

- Intersection models for specific types of intersections are presented that include internal constraints that arise from conflicts inherent to the intersection itself (e.g. due to crossing flows). This significantly enhances their realism in urban applications. These intersection models are the first that include internal constraints, satisfy the seven requirements and guarantee a unique solution.

2. Marginal simulation:
   This second part has a strong practical tendency. The novel concept of marginal DNL simulation is introduced. This is a computationally efficient, approximate methodology for repeated DNL simulations. Marginal DNL algorithms are developed that perform partial (marginal) DNL simulations of local variations to a base scenario instead of a full simulation for each variation. At the cost of acceptable approximation errors, this provides a considerable computational gain compared to explicit simulation with traditional DNL simulation models (in this case LTM). Marginal simulation is thus suited for a wide range of applications that require a large number of repeated simulations with large overlap, such as variability studies, gradient-based OD estimation, optimal control and robust network design. More specifically, the contributions of this part of the thesis are the following:

   - The Marginal Incident Computation (MIC) algorithm is presented as the first marginal DNL simulation algorithm. On the one hand, it is limited to simulating the impact of incidents. Also, the approximation it provides may be too rough for some applications. The large computation time savings that are obtained, however, render MIC suited for estimating incidental travel time losses in large-scale applications that do not require detailed analysis of each particular incident. An example is robust network design; see Snelder (2010) for an application of MIC in that context.

   - A second marginal DNL algorithm, Marginal Computation (MaC), is developed. Compared to MIC, it provides a smaller (but still significant) computational gain. However, its versatility – it can simulate changes in both demand and supply - and improved accuracy render it applicable to a wide range of possible applications as mentioned above. In Frederix et al. (2011), MaC is used for dynamic OD estimation.

- The general philosophy and critical issues of marginal simulation are thoroughly explained, which facilitates the transfer of this concept to other (types of) DNL simulation models and to other fields of study where high computation times of repeated simulations limit the research scope.

1.4 Overview of thesis

This chapter has provided a general introduction on DTA and DNL modelling. In this context, two developments are presented in the remainder of this thesis. The first, macroscopic intersection modelling, is covered in Chapters 2-6. Secondly, Chapters 7-9 treat marginal DNL simulation.

Chapter 2 provides a specific introduction on macroscopic first-order DNL intersection modelling. Foremost, the two main functions of the intersection model are explained. The first function is to find a consistent solution in terms of flows transferred over the intersection, considering all demand and supply constraints. Secondly, the intersection model should impose internal supply constraints to account for limited supply within the intersection itself (e.g. due to crossing flows hindering each other).
The subsequent chapters analyze and partially solve the modelling difficulties regarding the proper fulfilment of these two functions. Thereto, we start with a broad, general scope and add details and complexity along the way. At first, the first function of the intersection model is separately treated while only considering the external constraints of the adjacent links. Chapter 3 provides an overview of the state-of-the-art. In response to returning shortcomings in existing models, we compose a list of seven requirements for the proper fulfilment of the first function.
Chapter 4 then presents a general intersection model – limited to external constraints - complying with these requirements, accompanied by an efficient solution algorithm.
In Chapter 5, the second function of the intersection model is added by introducing internal supply constraints in a general way into the model of Chapter 4. It is found that this may lead to non-unique solutions for the flows. It is shown that this solution non-uniqueness is not an artefact of the presented model. Rather, it is an issue that affects some other existing models and should definitely be reckoned with when developing future models. Also, while this dissertation focuses on the point-like modelling approach, a digression on spatial models is given, showing that this does not solve the problem. Furthermore, Chapter 5 presents a sufficient and necessary condition for solution uniqueness.
Finally, Chapter 6 presents specific intersection models for different intersection types. These models are, to our knowledge, the first that incorporate both functions of macroscopic DNL intersection models while complying with the seven requirements of Chapter 3 and producing a unique solution. However, it is also explained that this theoretical consistency seems difficult to reconcile with further enhancing the realism of the intersection model in future research. It is envisaged that careful deliberation is needed on this matter.

Chapter 7 starts with an overview of a broad range of DNL applications in which the computational burden due to the need for a large number of scenarios or iterations limits the

research scope. The general concept of marginal simulation is then presented as a solution to the computational limitations of these types of problems.

The following two chapters describe two marginal first-order DNL algorithms, namely Marginal Incident Computation (MIC) and Marginal Computation (MaC). Since MIC and MaC adopt the modelling assumptions of LTM, they reproduce congestion formation and spillback far more realistically than existing fast models (e.g. static and analytical models).

Chapter 8 presents MIC, which is specifically designed to assess the effects of link capacity reductions due to incidents. At the time of development, MIC was conceived more as an independent post-processing method than as a marginal algorithm in the philosophy explained in Section 7.2. As a result, the accuracy and applicability are more limited than that of the later developed MaC. On the other hand, its computational speed renders it well suited for coarse large-scale incident scenario evaluations. A case study (on the Sioux Falls network) is presented, showing a good approximation of the aggregated output (VHL) and a large computational gain compared to explicit simulation with LTM.

MaC, introduced in Chapter 9, more faithfully adopts the principles of marginal simulation. Because of this, the approximation errors are reduced, enabling a wide range of possible applications. MaC is able to simulate variations in both traffic demand and (link) supply. A case study on the network around Ghent is performed, in which the sensitivity of the link flows to each route demand is analyzed. It is demonstrated that the results of explicit LTM simulations are well approximated with a considerable computational gain.

Finally, Chapter 10 formulates the main conclusions of this PhD research and highlights important future research directions.

PART I:
INTERSECTION MODELLING

2

INTERSECTION MODELS: INTRODUCTION

Macroscopic DNL models separately treat traffic flows on links and through intersections in a link model and an intersection (or node) model respectively. Although the term 'node model' is more common in the state-of-the-art for these typically dimensionless, point-like models, we prefer 'intersection model' in this dissertation. This term corresponds more intuitively to the extended definition of the intersection model that we aim to establish, developing it from merely a node that transfers flows and shock waves between the adjacent links into a complex model that also accounts for interactions within the intersection itself.
Flow propagation on links has been extensively studied and various adequate link models exist in the literature. DNL intersection models have attracted much less attention and are still quite underdeveloped. Yet, they are an equally important component of DNL models, especially in congested urban and regional networks. Indeed, the intersection model too has a decisive influence on how shock waves propagate and are initiated. It determines if, to what extent and in which directions congestion is formed or spills back over intersections. Consequently, it determines the travel times in the network and the variability thereof, the resulting driver decisions (e.g. regarding route), and with that the accuracy of most DNL- or DTA-related studies. Hence, adequate intersection models are vital to ensure the effectiveness of network improvements and (dynamic) traffic management.

Section 2.1 introduces the two functions of the intersection model in a first-order macroscopic DNL model. Section 2.2 distinguishes different types of intersection models, both from a modelling approach perspective as in terms of real-world design and priority rules. Section 2.3 mentions two simplifying assumptions that are made in this thesis.

2.1 Functions of the intersection model in DNL

The function of the intersection model in DNL models with regards to the traffic flows transferred over an intersection is twofold. The first and primary function of the intersection model is to find a consistent solution in terms of flows transferred over the intersection. This flow solution must unambiguously define each movement flow, i.e. from each incoming link that sends flow into the intersection towards each outgoing link through which flows leave the intersection. Constraints apply on these flows in the form of demands, which are the maximum, desired outflows of the incoming links, and supplies, which is the space (or time) available to the flows. The supplies are to be distributed among the competing incoming flows. Therein, not only the dependency of this supply distribution on the demand levels is to be accounted for, but also the interdependency between the distribution of different supplies.
Foremost, the constraints that apply in the intersection model stem from the adjacent links, i.e. the demands of incoming links and the supplies of outgoing links. These are inputs provided to the intersection model by the link model. These are the external constraints. In addition, constraints due to limited supply within the intersection itself may be imposed; these are called internal supply constraints. The addition of these internal supply constraints constitutes the second function of DNL intersection models. They typically do not apply to highway junctions but can be decisive at regional and urban intersections. Internal supply constraints result from the following conflicts:

- Traffic controls (traffic lights, ramp metering)
- Crossing conflicts
    - at (un)signalized intersections, between movements originating from different incoming links, heading towards different outgoing links
    - with non-motorized traffic (pedestrians, cyclists)
- Merging conflicts
    - between flows entering a roundabout, merging with flows already on the roundabout
    - between flows merging into an outgoing link

Conflicts between flows merging into the same outgoing link – included here for completeness - are typically considered as external constraints in the form of the outgoing link's supply as described above. In fact, they can be considered both internal and external: on the one hand, they can be dominated by congestion spilling back from the outgoing link; on the other hand, drivers (usually) evaluate crossing and merging conflicts simultaneously before traversing the intersection. In this dissertation, they are considered external, which is the universal assumption in the state-of-the-art. Furthermore, only motorized traffic is considered. Also conflicts due to traffic controls are only occasionally discussed since these are either – in case of non-adaptive control – quite straightforward to deal with, or – in case of adaptive control – to the best of our knowledge not included in state-of-the-art macroscopic DNL models.
Finally, we note that a third function of imposing additional intersection delay is needed in models that do not explicitly capture stochastic queue formations in under-saturated conditions (e.g. Durlin & Henn, 2005 and Yperman et al., 2007). This can be done using delay formulas such as those of Akcelik & Troutbeck (1991) and Webster (1958). This third

function is not discussed in this dissertation, which aims to analyze and improve the first two functions that determine the traffic flows. If the flows are modelled realistically, realistic delays can be found using existing delay formulas as mentioned above.

Properly fulfilling the first function - unambiguously determining all movement flows - is obviously obligatory for any intersection model. The second function of adding internal supply constraints is optional from a theoretical point of view. As it also further complicates the first function by adding more constraints to the problem, this second function is disregarded by most state-of-the-art intersection models. Such models, however, assume unlimited supply of the intersection itself. This implies for instance that crossing vehicles do not hinder each other in any way. In urban and regional applications, this simplification significantly reduces the realism of the flows predicted by the intersection model. In summary, both of these functions aim to determine the flows over the intersection and in result affect the congestion dynamics in the adjacent links[7].

2.2 Types of intersection models

Distinction between different types of macroscopic DNL intersection models can be made in two ways. Firstly, the modelling approach may be point-like or spatial (Section 2.2.1). Secondly, different models may be developed for different types of intersections (Section 2.2.2).

2.2.1 Modelling approach

Roughly speaking, two possible approaches to DNL intersection modelling can be distinguished, namely point-like and spatial modelling. Point-like intersection models do not have physical dimensions. A point-like model combines all (external and internal) constraints into a strongly coupled set of equations for which the solution is calculated. A spatial intersection model on the other hand, disconnects (some of) the interdependencies by spatially separating (some of) the conflicts. Hence, a spatial model can be considered as a mini-network in which the conflict zones of crossing and/or merging flows are represented by dummy nodes, connected by dummy links. As such, a spatial model subdivides one large problem into several smaller ones that are easier to solve. While it is possible to define specific functionalities for these dummy links and nodes (as in Buisson et al., 1995), we will consider them to function exactly like real links and intersections. The latter is probably more common in practice, since networks serving as input to the DNL model are often adopted from static models, in which intersections are often represented spatially. Figure 2-1 illustrates the difference between a point-like and a spatial model (the intersection model is indicated by the dashed circle). Note that different configurations of a spatial model are conceivable (see further below).

[7] And the intersection delay in under-saturated conditions, if implemented (see e.g. Yperman, 2007).

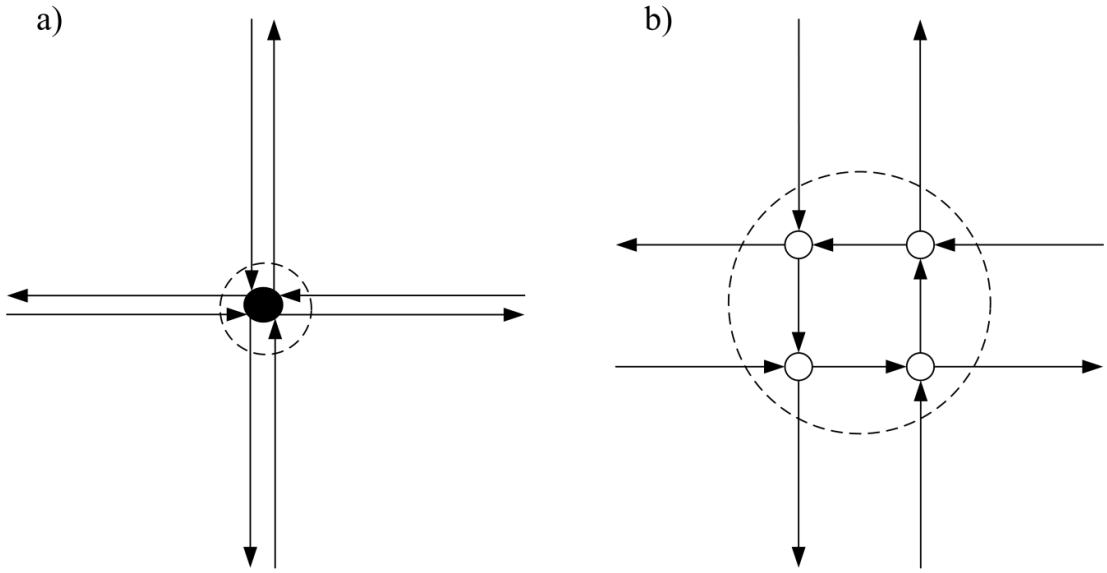

Figure 2-1: Example of a point-like (a) and a spatial (b) intersection model

The vast majority of existing macroscopic DNL intersection models is of the point-like type. For models limited to the first function that only serve as a connector of the external constraints of the adjacent links, this approach is undisputed. Indeed, if internal conflicts within the intersection itself (e.g. due to crossing conflicts, or the merging conflicts on the arcs of a roundabout) are neglected, there is little reason to consider the spatial dimensions of the intersection. The only possible motivation to choose for a spatial model in this case, would be to split up an intersection so that only a group of merges and diverges remains. The latter, having only one outgoing and incoming link respectively, are indeed much simpler to model. It will be shown in Chapters 3-4, however, that point-like models, when limited to external constraints, can also be very efficiently solved.
For intersection models that also implement the second function, the spatial characteristics of the intersection itself largely determine what the internal supply constraints should look like, i.e. how many distinct conflict points can be distinguished, which movements are in conflict, etc. In a point-like model, the set of equations may become very large and difficult to solve. Hence, a spatial modelling approach intuitively becomes more appealing, as this would alleviate this problem by spatially separating and disconnecting most of the interdependencies. However, the following disadvantages of spatial models can be identified:

- How to represent an intersection in a spatial model is a delicate matter. Caution is needed to make sure that the conflicts between the various movements are modelled realistically. For example, Chen et al. (2008) define their spatial intersection model as a grid of two by two cells, in which some movements that in reality are not in conflict do hinder each other, whereas some other movements that are in conflict in reality are not in the model. Moreover, the configuration in Chen et al. (2008) can easily lead to model-induced, unrealistic gridlock. Simply put, gridlock describes a complete standstill (zero flow over the intersection) caused by a continuously decreasing flow due to a circularly moving spillback wave. In Figure 2-2, an example of a gridlock-

sensitive representation (a) and a non-gridlock-sensitive (b) spatial model is given. In essence, grouping movements into circular patterns may cause gridlock.

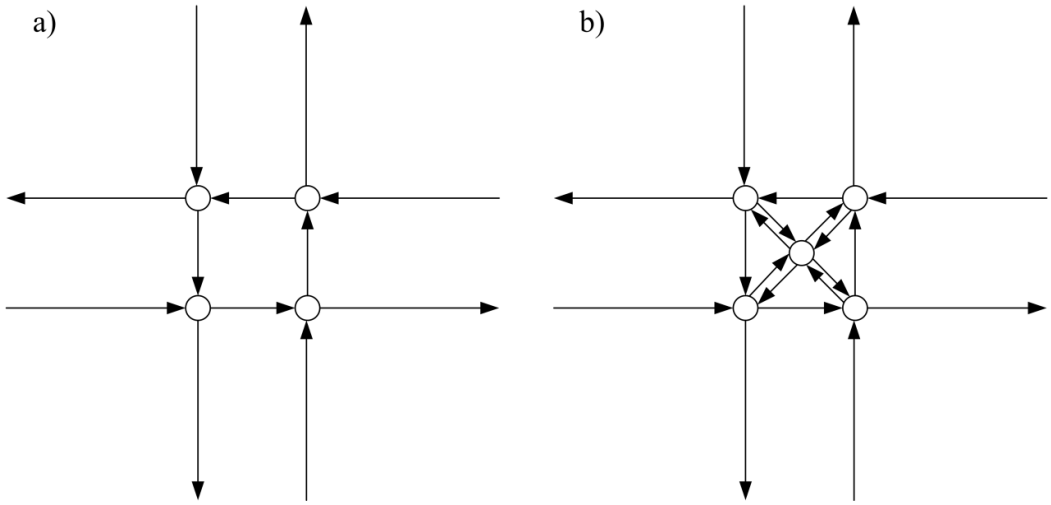

Figure 2-2: Gridlock-sensitive (a) and non-gridlock-sensitive (b) spatial model

- A second problem of spatial models is caused by the spatial separation of conflicts that are in reality subject to simultaneous driver decisions. Consider the example in Figure 2-3, where an eastbound minor flow has to yield to bi-directional traffic. Unless some storage space for crossing vehicles is present on the intersection, a vehicle from the minor street will only cross if no prioritized vehicle is approaching from either side.

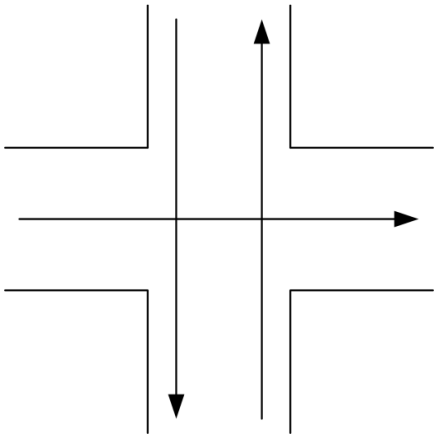

Figure 2-3: Two conflicts with bi-directional prioritized traffic subject to simultaneous driver decisions

While in reality the two internal conflicts are evaluated simultaneously by a driver, they would be separated into two dummy nodes in a spatial model, which seems to leave no option to capture the dependency between the two. While it will turn out that these microscopic dependencies also pose a severe difficulty for point-like

intersection models (see Section 6.5), still, they seem even more problematic in a spatial model.

- From a computational perspective, it should be noted that numerical solution procedures that need to comply with the Courant-Friedrichs-Lewy (CFL) condition (Courant et al., 1928) - stating in the given context that the simulation time step cannot be larger than the link travel time -, such as the Godunov scheme (see Lebacque, 1996), are forced to operate at very small time steps because of the short dummy links in the spatial intersection model. This increases the overall computation time tremendously. However, this criticism does not carry over to numerical schemes following the variational formulation of kinematic waves (Daganzo, 2005a; Daganzo 2005b), or the fixed-point formulation (Gentile et al., 2007), both of which avoid this constraint.

Due to the abovementioned disadvantages of spatial models, we hold on to the point-like approach, also for extended intersection models including internal supply constraints.

2.2.2 Intersection type

Most existing DNL intersection models represent general nodes rather than a specific type of intersection. Such general models are suitable for standard unsignalized intersections (usually excluding roundabouts[8]). Signal phases are typically not considered[9].
While the main objective of this dissertation is on developing a theoretically sound general intersection model (more consistent and elaborate than existing models), also specific models for particular types of intersections are presented. Below, an overview is given of the different types of intersections that we identify based on lay-out and priority rules. In this discussion and throughout the thesis, right-hand driving is assumed. Simple merges and diverges are not included as separate types of intersections.

- Priority-To-The-Right (PTTR):
  At PTTR intersections priority always goes to the vehicle approaching from the right. This is the default priority rule in most European countries in absence of marks, signs or signals. Typically, it is only used for intersections with low traffic volumes. Therefore, this type is usually not very important in DNL applications and it is rarely explicitly considered in the state-of-the-art. Still, in city centres, oversaturated intersections can be found that are governed by PTTR.

[8] To our knowledge, the only specific first-order macroscopic DNL roundabout models are those of Chevallier & leclercq (2008) and Mercier (2009). Both are spatial models, connecting multiple merges and diverges and T-junctions respectively.
[9] General intersection models can be applied to signalized intersections by alternately blocking the flows to simulate the red phases. This, however, requires shorter simulation update steps.

- All-Way-Stop-Controlled (AWSC):
  The arrival order determines the departure order at AWSC intersections. All incoming traffic first has to yield to a stop sign. Then, priority goes to the vehicle that arrived at the intersection first. AWSC intersections are common in the United States, Canada and South Africa.

- Roundabout:
  A roundabout avoids crossing conflicts by letting all incoming links merge onto the unidirectional circular roundabout. In general, vehicles already on the roundabout have priority over vehicles that wish to enter the roundabout. Exceptionally, roundabouts may be governed by PTTR or signals.

- Priority-controlled:
  Although obviously, priority rules exist at all types intersections, the term 'priority-controlled' is reserved to indicate intersections where priority is prescribed to movements coming from the prioritized (or major) streets over movements from the minor streets. In addition, left-turning traffic has to give way to right-turning and straight traffic.

  These first four intersection types may be commonly addressed as unsignalized intersections.

- Signalized:
  At signalized intersections, traffic lights alternatively allow different movements to pass. This way, crossing conflicts between straight flows are avoided. Remaining conflicts within the same green phase are again solved by granting priority to the straight and right-turning movements over the left-turning traffic.

2.3 Simplifying assumptions

The intersection model developments in the following chapters aim to unite theoretical soundness and practical accuracy regarding both congestion spillback and congestion formation due to limited supply of the intersection itself. Still, there are several ways to further improve upon the presented models. The two main simplifying assumptions that are adopted throughout this thesis are the following.

- The influence of microscopic dependencies that occur at the level of individual drivers (for instance due to simultaneous driver decisions as discussed in Section 2.2.1) on the macroscopic flows is not yet included in this dissertation. The reasons, consequences and future intentions regarding this simplification are further elaborated on in Section 6.5.

- The separation of movements into turning lanes is not considered. Instead, the simplifying assumption is made that all incoming links have a homogenous cross-section. If necessary, a diverge could be introduced on a link where the turning lanes

start, so that the lanes are modelled as different incoming links. Of course, this introduces short links into the DNL model which may have a negative impact on the computation time in DNL models that need to satisfy the CFL condition.

2.4 Conclusion

This chapter has provided an introduction of macroscopic DNL intersection modelling. In Section 2.1, a distinction is made between two functions that influence the flow solution of the intersection model. The first function is to find a consistent solution in terms of flows transferred over the intersection. The second is to impose additional, internal supply constraints on the flow solution due to limited supply of the intersection itself (e.g. due to crossing conflicts). In the next chapter, it will be explained that the vast majority of existing models does not properly fulfil the first function. Consequently, the solution of these models may be inconsistent with the prevailing constraints or with sensible traffic flow dynamics. In addition, the second function of taking into account conflicts within the intersection itself is typically entirely disregarded. This significantly compromises the realism and applicability in urban and regional applications.

Following a point-like approach as explained in Section 2.2.1, the subsequent chapters thoroughly analyze the modelling difficulties regarding these two functions. General requirements and conditions for their proper fulfilment are presented. This builds to specific models for the intersection types listed in Section 2.2.2. These models are still simplified in light of the assumptions mentioned in Section 2.3.

3

A CONSISTENT SOLUTION UNDER EXTERNAL CONSTRAINTS

This chapter discusses the first function of macroscopic DNL intersection models, namely to find a consistent solution in terms of flows transferred over the intersection from each incoming link to each outgoing link, accounting for all demand and supply constraints. In this chapter and the next, only the external constraints of the adjacent links are included.

First, Section 3.1 formally defines the external constraints. As an introductory step towards intersections with multiple incoming and outgoing links, Section 3.2 explains simple merge and diverge models. Section 3.3 provides an overview of the state-of-the-art on first-order macroscopic DNL intersection models, highlighting both their contributions and shortcomings. This literature overview reveals an existing lack of knowledge and understanding of the properties these models should possess. In response, a set of seven requirements is presented that guarantees consistency with the prevailing constraints and with sensible traffic flow dynamics (Section 3.4). These requirements are generic rules for first-order macroscopic DNL intersection models, regardless of its type (e.g. priority-controlled, signalized or roundabout). To build an intersection model from these foundations requires the specification of Supply Constraint Interaction Rules (SCIR). This is explained in Section 3.5. The SCIR govern how the supplies affect each flow, and the interaction between demand and supply constraints in that process. By doing so, it completes the first function of the intersection model.

This chapter is an edited version of the first part of 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 45 (1), pp. 289-309.*

3.1 Formal definition of the external constraints

The demand constraint S_i is the maximum flow that incoming link i (i = 1..I) could possibly send if the intersection and outgoing link(s) would impose no constraint whatsoever on the outflow of link i, as if link i was directly connected to a reservoir with infinite capacity. Obviously, S_i is determined purely by traffic conditions on link i and is thus delivered as an input to the intersection model by the link model.
The supply constraint R_j is the maximum inflow that outgoing link j (j = I+1..I+J) could receive if the intersection and incoming link(s) would impose no inflow constraint whatsoever, as if link j was connected to a reservoir capable of sending an infinite flow. Also R_j is determined by traffic conditions on link j and is provided by the link model.
For more details on how S_i and R_j are determined by the link model, see Appendix A or Yperman (2007).

Note that, since this thesis deals with dynamic models, of course the demands and supplies are variable over time. Obviously, this holds as well for the flows q, the turning fractions f and internal supplies N (which are defined further on). For notational convenience, however, the time dimension (t) is omitted in the equations. All equations hold for one time step in time-discrete models (assuming all variables to be constant within a time step). Alternatively, the equations could hold for one point in time for time-continuous models – although time-continuous models are not the focus of this dissertation.

3.2 Merge and diverge models

Many intersection models are generalizations of simple merge and diverge models, which are described by e.g. Daganzo (1995), Lebacque (1996), Jin & Zhang (2003) and Ni & Leonard (2005). This section introduces the basics of merge and diverge modelling to facilitate the understanding of the following sections.

3.2.1 Merge model

The merge model connects two incoming links i (i = 1, 2) to one outgoing link, maximizing the total flow q into the outgoing link. More general merge models with more than two incoming links i can be solved entirely analogously, so we limit the explanation to the base case with I = 2. Considering conservation of vehicles, this yields the following optimization equation, with q_1 and q_2 being the outflow from incoming link 1 and 2 respectively:

$$\max\ q = q_1 + q_2 \tag{3.1}$$

The demand and supply constraints of a merge model express that the total inflow q into the outgoing link is constrained by supply R, and the outflows q_1 and q_2 by the demands S_1 and S_2 respectively. The demand constraints can simply be written as:

$$\begin{aligned} q_1 &\le S_1 \\ q_2 &\le S_2 \end{aligned} \tag{3.2}$$

The supply constraint is represented as a function $\hat{R}(q_1, q_2) \le 0$ of the incoming flows (3.3). The circumflex is added to make a distinction between the supply constraint function $\hat{R}$ and the actual supply R (veh/h) that the outgoing link can receive. The supply constraint function $\hat{R}$ indicates whether the supply constraint is active ($\hat{R} = 0$), i.e. supply R is fully used ($\sum_i q_i = R$), inactive ($\hat{R} < 0 \Rightarrow \sum_i q_i < R$) or violated ($\hat{R} > 0 \Rightarrow \sum_i q_i > R$). Of course, only flow patterns for which $\hat{R} \le 0$ are valid solutions.

$$\hat{R}(q_1, q_2) = q_1 + q_2 - R \le 0 \tag{3.3}$$

The formulation in (3.3) corresponds entirely with how external supply constraints are universally treated in the state-of-the-art. While for the moment explicitly defining a function $\hat{R}$ may seem unnecessarily complex for this linear external supply constraint, it will prove convenient when generally introducing internal supply constraint functions (see Chapter 5).
It is straightforward to see that (3.2)-(3.3) imply $q = \min\{S_1 + S_2, R\}$. The solution of the merge model, however, needs to specify the flows q_1 and q_2 of both incoming links. If the sum of the demands is smaller than the supply, the demand constrained solution (3.4) results immediately. In case the demand exceeds the supply, a queue will form on at least one of the incoming links. Daganzo (1995) uses parameters d_i ($d_1 + d_2 = 1$) to reflect priorities in the distribution of the supply R. These parameters represent the strength of the incoming links in the competition for R, so that a share d_iR is assigned to each link i. If $S_i < d_iR$ for i, $R - S_i$ is appointed to the other link i'. For coherence with the remaining sections and chapters, we redefine the merge model of Daganzo in terms of general priority parameters α_i that do not necessarily add up to one. The rightful share of link i in the supply distribution is then given by $\frac{\alpha_i}{\alpha_i + \alpha_{i'}} R$. Summarizing, three types of solutions can occur:

1. Both incoming links are demand constrained:

$$q_1 = S_1;\ q_2 = S_2;\ q = S_1 + S_2 < R \tag{3.4}$$

2. Link i is demand constrained; link i' is supply constrained:

$$q_i = S_i;\ q_{i'} = R - S_i;\ q = R \tag{3.5}$$

3. Both incoming links are supply constrained:

$$q_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2} R;\ q_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2} R;\ q = R \tag{3.6}$$

From (3.6), it follows immediately that $\frac{q_2}{q_1} = \frac{\alpha_2}{\alpha_1}$ in case both incoming links are supply constrained. This implies that the ratio of the priority parameters determines directly the ratio of the resulting flows if both links are supply constrained. For the second type of solution holds that $\frac{q_{i'}}{q_i} \geq \frac{\alpha_{i'}}{\alpha_i}$. Indeed, because link *i* is demand constrained, it does not fully use its rightful share of *R*. The other, supply constrained link *i'* thus receives more than (or at least) its rightful share of the determinative supply constraint.

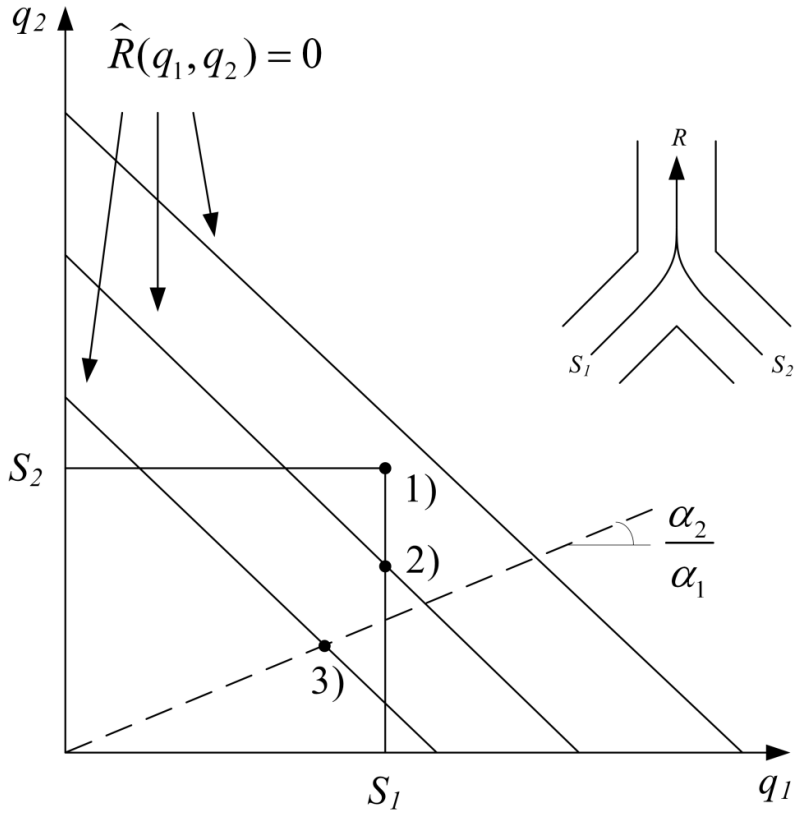

Figure 3-1: Three types of solutions in the merge model of Daganzo (1995)

The solution of Daganzo's merge model can be graphically constructed starting from the constraints and the priority ratio. Figure 3-1 shows how this results in one of the three solution types. Firstly, the demand constraints S_1 and S_2 set upper bounds on the flow from their respective incoming link. Secondly, the external supply constraint function $\widehat{R}(q_1, q_2) \leq 0$ further reduces the solution space. In this illustration, the three different solutions are obtained by comparing three supply constraint functions (as supply *R* is increased, the line

representing $\hat{R} = 0$ shifts upwards). Other boundary conditions could have been varied, leading to an analogous example. Intuitively, the solution can be found by following the priority ratio line until it meets some constraint. If it first meets the supply constraint, the intersection point is the type 3 solution (3.6). If a demand constraint is reached, the remaining supply is allocated to the other, not yet demand constrained link. If this results in the supply being fully consumed, a solution of type 2 is obtained (3.5). If also the demand constraint of the other link is reached before the supply is used up, one ends up in a type 1 solution (3.4).

Jin & Zhang (2003) incorporate "fairness" into Daganzo's merge model, with priority parameters equal to the demands: $\alpha_i = S_i$. This is called a demand proportional distribution. Many intersection models are based on this assumption (see Section 3.3). Demand proportional models can only produce solutions of the first and third type, i.e. either no incoming link is supply constrained, or both. This is because the priority ratio line coincides with the demand ratio (see Figure 3-2).

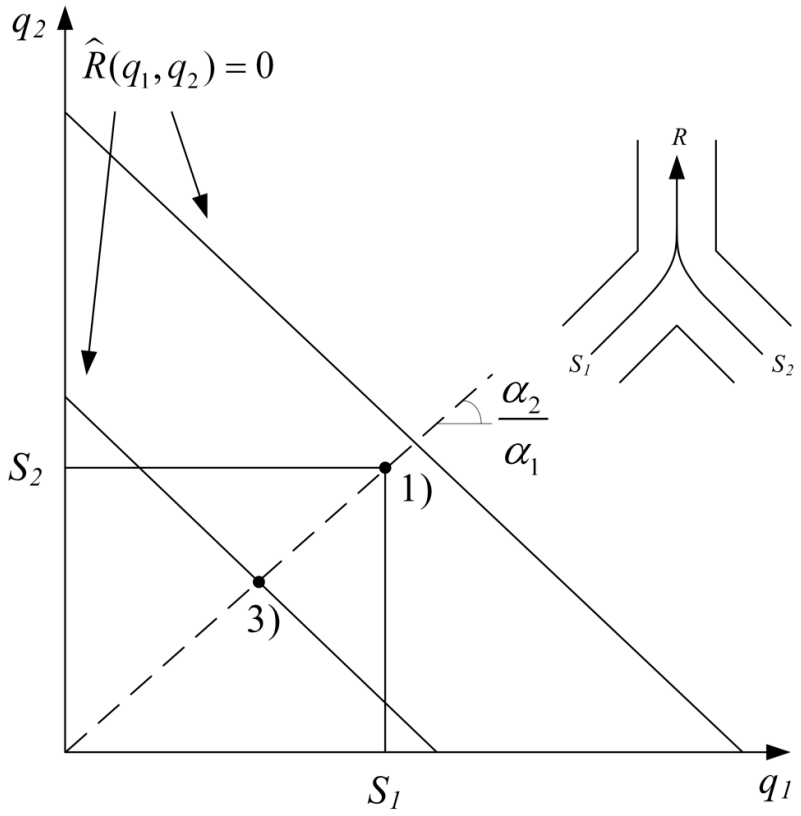

Figure 3-2: Only two types of solutions in the merge model of Jin & Zhang (2003)

Finally, Ni & Leonard (2005) suggest a model that applies a capacity proportional distribution. Therein, $\alpha_i = C_i$ with C_i being the physical capacity of incoming link *i*. They also make the extension to a general merge with any number of *i*. For a traditional merge model with two incoming links as discussed in this section, the possible outcomes can again

be described in the form of three solution types as in (3.4)-(3.6). Ni & Leonard (2005) support the validity of their model with empirical tests at a highway merge.

3.2.2 Diverge model

A diverge connects one incoming link to two outgoing links j (j = 1, 2). Cases with $J > 2$ can be solved analogously. Again, the total flow q is maximized (3.1); q now being the outflow of the incoming link, which is divided over the outgoing links according to turning fractions f_j ($f_1 + f_2 = 1$). Daganzo (1995) assumes for his diverge model that vehicles exit in a FIFO sequence, delaying successive vehicles regardless of their destination. As will be explained in more detail in Section 3.4.1, this implies that the turning fractions are given, external boundary conditions. Consequently, if the inflow into one of the links j is constrained, the inflow into the other is constrained accordingly. The demand constraint is simply given by $q \le S$. The supply constraint functions can now be written as:

$$\begin{aligned} \hat{R}_1(q) &= f_1 q - R_1 \le 0 \\ \hat{R}_2(q) &= f_2 q - R_2 \le 0 \end{aligned} \tag{3.7}$$

Given the maximization requirement (3.1) and the demand and supply constraints, the solution is simply determined by the most stringent of these constraints:

$$q = \min\left\{S, R_1 / f_1, R_2 / f_2\right\} \tag{3.8}$$

3.3 Overview of first-order macroscopic DNL intersection models

Macroscopic DNL intersection models are typically general extensions of the merge and diverge models discussed in the previous section, not aimed at a particular type of intersection as distinguished in Section 2.2.2. This literature overview is limited to articles published prior to our own work. The mentioned studies and models formed the starting ground for our findings and developments in this chapter and the next. More recent models will be discussed further in the text, when appropriate.

In intersection models, the solution needs to specify the flows from every i to every j, i.e. all partial flows q_{ij} ($q_i = \sum_j q_{ij}$ and $q_j = \sum_i q_{ij}$). Some models distinguish between the outflows of incoming links i headed towards j (q_{ij}^-) and the inflows of outgoing links j coming from i (q_{ij}^+). However, conservation of vehicles requires that the two are the same, so we refrain from using a double notation and simply refer to single-defined movement flows q_{ij} ($= q_{ij}^- = q_{ij}^+$).

One of the simplest intersection models is that of Holden & Risebro (1995). They define an entropy condition that maximizes the total flow, subject to the following constraints:

$$\begin{gathered} 0 \le q_i \le S_i,\ 0 \le q_j \le R_j \quad \forall i, j \\ \sum_i q_i = \sum_j q_j \end{gathered} \tag{3.9}$$

Turning fractions are not an exogenous input to the model of Holden & Risebro (1995), who assume that drivers choose their destination link solely based on the principle of least resistance. We find this to be an oversimplification that does not properly represent vehicular traffic. Drivers have the intention of travelling towards a specific destination, following a certain route. It is necessary to consider this route choice instead of merely maximizing flows irrespective of the destination link after the intersection. The latter would be acceptable for fluids or any other commodity having no preference for a route, but not for vehicular traffic with autonomous route choice. Herty & Klar (2003) also adopt this assumption and formulate a multi-lane model. Coclite et al. (2005) applied a similar approach as Holden & Risebro (1995), maximizing the total flow subject to the constraints in (3.9) and considering fixed turning fractions so that $q_{ij} = f_{ij} q_i$. Hence, they consider linear external supply constraint functions in the form of (3.10) that extend the merge supply constraint of Daganzo (1995). Virtually all existing DNL intersection models implement supply constraints of outgoing links in this form.

$$\hat{R}_j(\mathbf{q}) = \sum_i f_{ij} q_i - R_j \le 0 \quad \forall j \tag{3.10}$$

In (3.10), the vector **q** is composed of all q_i for which $f_{ij} > 0$.
However, Coclite et al. (2005) need to impose the artificial constraint $f_{ij} \neq f_{i'j}$ (for each j, with $i' \neq i$) to guarantee a unique solution.

The main issue with the simple models described above is that they do not include a distribution rule that governs how the supplies R_j are to be divided among the competing incoming links i. Simply put, a distribution rule defines a rightful share of R_j that is reserved for each i, which we denote as R_j^i. Typically, this rightful share is defined as[10]:

$$R_j^i = \frac{\alpha_{i(j)}}{\sum_{i'=1..I} \alpha_{i'(j)}} R_j \tag{3.11}$$

In (3.11), the rightful share R_j^i is determined by a proportionality rule that again extends Daganzo (1995). The priority parameters are, for now, denoted as $\alpha_{i(j)}$ as they may or may not be specific to j.

[10] To our knowledge, the model of Gibb (2011) is currently the only exception that does not use such a proportionality rule – see Section 3.4.2.

Several models exist that apply a distribution proportional to number of lanes or capacities (the two are of course largely equivalent). Such models are proposed by Gentile et al. (2007), Taale (2008) and Ni et al. (2006)[11]. It is also suggested as a possibility in Adamo et al. (1999), who propose several different intersection models.
Now, it may happen that some incoming links i are unable to fully consume their rightful share of an overloaded supply, for instance if $S_i < R_j^i$. In this case, the leftover supply has to be redistributed among the other competing incoming links that can still send more flow, analogous to (3.11). This interaction with the demand constraints is usually accounted for in state-of-the-art intersection models. However, the same interaction exists between different supply constraints, at least in models that enforce FIFO. This interaction is to be captured in Supply Constraint Interaction Rules (SCIR), as will be explained in detail in Section 3.5. However, the models of Ni et al. (2006), Taale (2008) and some of the models in Adamo et al. (1999) are not FIFO. We will argue in Section 3.4.1 that FIFO is a necessary model assumption. The model of Gentile et al. (2007) and again some models in Adamo et al. (1999) are FIFO, but do not implement SCIR that properly account for the redistribution of supply resulting from the interaction between different supply constraints[12]. This holds equally for many of the demand proportional models that are discussed below.

In numerous models, the priority parameters α_i and the turning fractions f_{ij} determine the competitive strength $f_{ij}\alpha_i$ of i for the proportional sharing of supply R_j in (3.11). As such, α_i is the maximal strength (if $f_{ij} = 1$). It is indeed logical that the turning fractions f_{ij} play an explicit role, because the competitive strength of i for R_j obviously decreases with f_{ij}. In the model of Herty et al. (2006), the f_{ij} exclusively determine the competitive strength (i.e. $\alpha_i = 1 \quad \forall i$). This assumption implies that $q_i = q_{i'}$ for any two incoming flows that are supply constrained by the same outgoing link. This is obviously an oversimplified assumption.
Notably, several demand proportional models exist that extend Jin & Zhang (2003) by imposing $\alpha_i = S_i$. Then, the competitive strength is $f_{ij}\alpha_{ij} = f_{ij}S_i = S_{ij}$. Bliemer (2007) formulates a demand proportional intersection model of which the solution in terms of q_{ij} can be formulated as in (3.12). Bliemer (2007) claims this model to be flow maximizing. However, the example provided in Section 3.4.3 proves otherwise. This model is also used in Raadsen et al. (2010).

$$\begin{aligned} q_i &= \min_{j' \backslash S_{ij'}>0} \left(S_{ij}, \frac{R_{j'}^i}{f_{ij'}}\right) \\ with: \quad R_{j'}^i &= \frac{S_{ij}}{\sum_{i'=1..I} S_{i'j'}} R_{j'} \end{aligned} \tag{3.12}$$

In (3.12), each flow q_i is limited by either demand or its most restrictive rightful share of supply. For the latter, only the supplies of j' to which i actually wants to send flow ($S_{ij'} > 0$) take effect on q_i. This means that it is assumed that drivers heading from i to j can only be

[11] Ni et al. (2006) only describe a model for a 2x2 intersection.
[12] In Gentile (2010), this problem is solved.

obstructed by vehicles on *i* itself (because of FIFO) or by vehicles from *i'* headed for the same *j*, but not by vehicles driving from other *i'* to other *j'*. Jin & Zhang (2004) do assume hindrance between all movements over the intersection, by omitting the '$|S_{ij} > 0$' condition in (3.12) and minimizing over all *j*. We do not find this a solid modelling assumption for two reasons. Firstly, in reality, only some (crossing) movements will hinder each other while others will not. Secondly, if crossing movements do hinder each other; this hindrance is not correctly captured by the shares of external supply. Rather, such conflicts are to be solved using internal supply constraints (see Chapters 5-6). Nie et al. (2008) propose a variant of the model of Jin & Zhang (2004) in which the priority parameter α_{ij} is bounded by the supply R_j; should this be higher than the partial demand S_{ij}. Other demand proportional intersection models are developed by Rubio-Ardanaz et al. (2001), Liu et al. (2008) and Kurzhanskiy & Varaiya (2010).

Lebacque & Khoshyaran (2005) criticize demand proportional models, showing that they may exhibit discontinuous changes in the flows. They formulate the invariance principle as a rule to avoid this problem (see Section 3.4.1). Following this principle, Lebacque & Khoshyaran (2005) assume that demand and supply cannot be linked directly in order to find the flows. Their model incorporates intersection demands and supplies, to be derived from a global zone fundamental diagram (Buisson et al., 1995, 1996a, 1996b; Lebacque, 2003). The intersection supply is distributed over the incoming links proportional to the number of lanes. However, it is not specified clearly how these intersection demands and supplies are to be obtained. Deriving a global zone fundamental diagram for each intersection does not seem practicable in large-scale macroscopic applications. This cumbersome procedure can be omitted, since defining intersection demands and supplies is not necessary to ensure compliance to the invariance principle, as is shown in the remainder. Finally, in response to Lebacque & Khoshyaran (2005), Jin (2010) incorporates interior states into merge models so that the invariance principle is asymptotically satisfied. In our opinion, this unnecessarily renders the problem more complex, since satisfaction of the invariance principle can also be assured (and not just asymptotically) without these interior states.

In summary, the above discussed intersection models all exhibit some or several deficiencies. In response to this, we drafted a list of seven requirements to which first-order macroscopic DNL intersection models should comply, see the next section.

3.4 Requirements for first-order macroscopic DNL intersection models

The literature overview in the previous section indicates the lack of a clear delineation of the necessary properties of macroscopic DNL intersection models. This delineation is given by the set of seven requirements listed below. Only through compliance with these requirements, general consistency is achieved with the prevailing constraints and with sensible traffic flow dynamics. However, the vast majority of existing models fails to satisfy some or several of these requirements.

3.4.1 Set of seven requirements

1. General applicability to any number of incoming and outgoing links and any combination of boundary conditions

Firstly, intersection models should be applicable to any (finite) number of incoming links (*I*) and outgoing links (*J*). Note that this requirement is by definition not met by merge or diverge models. However, adequate merge and diverge models (that comply with all other requirements listed below) exist in literature; for instance Daganzo (1995) and Ni & Leonard (2005).
Secondly, intersection models should be able to operate under any valid combination of boundary conditions (demands, supplies and turning fractions), so that they are generally applicable to any traffic situation. By 'valid', we mean that of course demands and supplies must be non-negative:

$$\begin{aligned} S_i &\geq 0 \quad \forall i \\ R_j &\geq 0 \quad \forall j \end{aligned} \tag{3.13}$$

Furthermore, turning fractions can only take values between (and including) 0 and 1 and must sum up to 1 for each incoming link:

$$\begin{aligned} 0 \leq f_{ij} &\leq 1 \quad \forall i, j \\ \sum_j f_{ij} &= 1 \quad \forall i \end{aligned} \tag{3.14}$$

2. Non-negativity of flows

Traffic never travels backwards and therefore all flows need to be non-negative:

$$q_i \geq 0 \quad \forall i \tag{3.15}$$

3. Conservation of vehicles

Vehicles neither disappear nor are created at an intersection. Therefore, the outflow of the incoming links must equal the inflow of the outgoing links, both in terms of total flow (i.e. $\sum_i q_i = \sum_j q_j$) as in terms of partial flows. The latter implies that the flow q_{ij}^- coming from *i* and heading towards *j* must equal the flow q_{ij}^+ going into *j* and originating from *i*. We advise to define only one flow q_{ij} ($= q_{ij}^- = q_{ij}^+$) between *i* and *j*, dependent on turning fractions f_{ij}, such that:

$$\begin{aligned} & q_{ij} = f_{ij} q_i \quad \forall i, j \\ & with: \quad q_i = \sum_j q_{ij} \quad \forall i \\ & \qquad\quad q_j = \sum_i q_{ij} \quad \forall j \end{aligned} \tag{3.16}$$

This way, conservation of vehicles is guaranteed; and it simplifies the formulation.

4. Compliance with the demand and supply constraints

The demand constraints state that the outflow q_i from an incoming link i can never exceed the demand S_i at the downstream boundary of i. Demand constraints are imposed only on the flows from i:

$$q_i \le S_i \quad \forall i \tag{3.17}$$

The supply constraints state that the inflow q_j into an outgoing link j can never exceed the supply R_j that is available at the upstream boundary of j. Such constraints are typically imposed on all flows competing for R_j (i.e. $i|f_{ij} > 0$). This is expressed by the earlier defined formulation (3.10).

$$\hat{R}_j(\mathbf{q}) = \sum_i f_{ij} q_i - R_j \le 0 \quad \forall j \tag{3.10}$$

where the vector $\mathbf{q}$ is composed of all $q_i|f_{ij} > 0$.

5. Ensuring FIFO: Conservation of Turning Fractions (CTF)

FIFO over an intersection implies that traffic flows out of an incoming link and into different outgoing links in the same order as they reached the end of the incoming link. Vehicles that are unable to exit into their preferred outgoing link prevent all those behind, regardless of destination, to continue (Daganzo, 1995). In other words, if either one of the partial flows q_{ij} is supply constrained, all other flows $q_{ij'}$ from the same link i are restricted accordingly. Hence, at the intersection level, ensuring FIFO is equivalent to ensuring Conservation of Turning Fractions (CTF).

The turning fractions f_{ij} can be obtained in various ways. Fixed[13], predefined turning fractions could be provided (in a SC DNL model); or the link model could provide separate, partial demands S_{ij} from each i to each j (in a MC model). This can for instance be achieved by considering separate route flows or fixed destinations with some local route choice logic like DUO routing (Ran et al., 1993). Also, turning fractions could be imposed onto the drivers by some controller (e.g. by police officers in the case of evacuations). If fixed turning fractions are chosen, CTF is obviously guaranteed. If partial demands S_{ij} are considered, CTF implies that the outflow composition of a link i in terms of partial flows q_{ij} must be identical to its

[13] By this, we mean fixed per update time step, not fixed for the entire simulation.

demand composition in terms of partial demands S_{ij}. Consequently, all outflows q_{ij} are mutually coupled through the turning fractions that are obtained from the demand composition:

$$f_{ij} = \frac{S_{ij}}{S_i} = \frac{q_{ij}}{q_i} \tag{3.18}$$

This implies that the turning fractions are given, external boundary conditions. Due to CTF, the intersection model's solution is unambiguously defined by the incoming flows q_i. Moreover, it is advisable to have the intersection model produce a solution in terms of the total flows q_i and derive the partial flows as $q_{ij} = f_{ij}q_i$. Formulations of the solution in terms of q_{ij} are more susceptible to violation of the CTF requirement.

Now, one might argue that the FIFO assumption neglects the separation of traffic in different turning lanes. Indeed, if turning lanes are present, this may – depending on the specific geometry - initially allow a queue towards an outgoing link to be bypassed by vehicles headed for other directions. However, if congestion spills back farther than the end of the turning lanes, it will affect the other turning movements. Then, FIFO becomes the more realistic assumption. In situations where non-FIFO behaviour is more realistic, one can still use a FIFO intersection model and add a diverge at the beginning of the turning lanes, which are then modelled as separate incoming links. However, a non-FIFO intersection model (i.e. that violates the CTF requirement) can never reproduce FIFO behaviour.

6. Flow maximization from the users' perspective

The maximization of flows can be seen as an extension at the intersection level of the maximization of entropy presented by Ansorge (1990). It follows from the fact that drivers will always try to advance whenever possible. This corresponds to individual maximization of each flow q_i. This maximization from the users' perspective is to be understood as "each flow q_i should be actively constrained by either demand or a share of supply (which is fully used)". Otherwise, the flow would increase until it hits some constraint.

$$q_i = S_i \ \textit{and / or} \ \exists j \mid f_{ij} > 0 : q_{ij} = R_j^i \ \& \ \widehat{R}_j(\mathbf{q}) = 0 \quad \forall i \tag{3.19}$$

In order to fulfil this requirement, SCIR need to be defined that consistently distribute each supply among the competing i (determining the shares of supply R_j^i for each i). Thereby, the interaction with the demand and other supply constraints must be accounted for (see Section 3.5). Intersection models that fail to implement SCIR and instead seek for a global flow maximization (i.e. maximizing the total flow over the intersection without considering distribution rules for the supplies) are unacceptable. Indeed, global flow maximization would imply a behaviourally unrealistic cooperation of drivers. Therefore, the 'max' operator used in merge and diverge models (see Section 3.2) cannot simply be adopted in intersection models.

7. Compatibility with link traffic flow dynamics: compliance with the invariance principle

If the intersection model determines that $q_i < S_i$, then i enters a congested regime and a queue starts to build up on i. As a consequence of link traffic flow dynamics, the demand S_i increases after some infinitesimally small time increment to the link capacity (i.e. the queue discharge rate) C_i. Any intersection model that predicts a different outcome for q_i because of this change from S_i to C_i contradicts its own initial solution, which leads to discontinuous changes in flow. In response to this observation, Lebacque & Khoshyaran (2005) formulate the invariance principle. This principle states that the solution of the intersection model must be invariant to replacing S_i by C_i if q_i is supply constrained ($q_i < S_i$). Analogously, if $q_j < R_j$ the solution should be invariant to an increase of R_j to C_j. For more details, see Lebacque & Khoshyaran (2005).

$$\begin{aligned} &q_i \text{ is invariant to } S_i \to C_i \ \forall i \mid q_i < S_i \\ &q_j \text{ is invariant to } R_j \to C_j \ \forall j \mid q_j < R_j \end{aligned} \tag{3.20}$$

Violations of the invariance principle on the demand side are (far) more likely than on the supply side. It must be stressed that an intersection model applying a demand proportional supply distribution does not satisfy the invariance principle.

3.4.2 Discussion

Any first-order macroscopic DNL intersection model should comply with the above requirements, irrespective of the type of intersection and of the driver behaviour that might vary for instance due to visibility, signposting, legislation, interpersonal or cultural differences (e.g. urge to enter the intersection if the outflow is blocked).

The first four requirements are rather straightforward and well-known in the literature. Hence, most existing intersection models satisfy these four requirements. The last three, however, are often violated. Table 3-1 provides an overview of the compliance of state-of-the-art first-order DNL intersection models (limited to external constraints) with the seven requirements.

In Chapter 4, an intersection model is presented that generalizes our own model published in Tampère et al. (2011) and that meets all seven requirements. Independently of this, intersection models compliant with the seven requirements[14] have been developed by Gentile (2010) and Flötteröd & Rohde (2011) – Flötteröd & Nagel (2005) and Flötteröd (2008) contain a more rudimentary description of this model. Flötteröd & Rohde (2011) present their model as a stepwise procedure in which all demands and supplies are incrementally consumed. Gentile (2010) only presents a solution algorithm without properly defining the model equations. These two models are encompassed by our general intersection model. Instead of elaborating on these particular models, we simply refer to our formulation in Chapter 4.

[14] Also the model of Lebacque & Khoshyaran (2005) complies, but this – as mentioned above - requires a level of detail and data gathering and handling that is unjustifiably high for an intersection model that only accommodates external constraints.

Table 3-1: Compliance of state-of-the-art models with seven requirements

Model	Gen app	Non-neg	Cons veh	Satisf constr	CTF	Ind max	Inv prin
Holden & Risebro	x	x	x	x			x
Herty & Klar	x	x	x	x			x
Herty & al.		x	x	x	x		x
Coclite et al.		x	x	x	x		x
Gentile et al.	x	x	x	x	x		x
Taale	x	x	x	x		x	x
Ni et al.		x	x	x		x	x
Adamo et al. G1	x	x	x	x			x
Adamo et al. G2	x	x	x	x	x		x
Adamo et al. G3	x	x	x	x		x	
Adamo et al. G4	x	x	x	x	x		
Adamo et al. G5		x	x	x			x
Adamo et al. G6		x	x	x	x		x
Adamo et al. G7	x	x	x	x		x	x
Jin & Zhang	x	x	x	x	x		
Bliemer	x	x	x	x	x		
Nie et al.	x	x	x	x	x		
Rubio-Ardanaz et al.	x	x	x	x	x		
Liu et al.	x	x	x	x	x		
Kurzhanskiy & Varaiya	x	x	x	x	x		
Tampère et al.	x	x	x	x	x	x	x
Gentile	x	x	x	x	x	x	x
Flötteröd & Rohde	x	x	x	x	x	x	x
Gibb	x	x	x	x	x	x	x

To our knowledge, the model of Gibb (2011) is currently the only realistic model compliant with the seven requirements that cannot be described by our model formulation. Instead of applying a proportionality rule to govern the supply distribution, Gibb (2011) defines capacity consumption factors. These factors determine the portion of an incoming link's capacity that each vehicle consumes, depending on the overload that exists for its destination link (which again depends on the capacity consumption factors). In other words, the heavier the competition for a certain supply, the longer a vehicle headed in that direction must wait its turn, and hence the more of the incoming link's capacity that is lost to this vehicle. Consequently, the capacity consumption factors for the various outgoing directions together determine the flow restriction for an incoming link. Gibb's model is directly and explicitly designed for turn-taking behaviour. It is difficult (in our experience) to extend to other types of priority behaviour (see Section 5.1.1). Hence, while this model is a very strong candidate for some applications (e.g. for AWSC intersections or if the flow patterns are dominated by merging into congested outgoing links), it seems less flexible than our general model formulation presented in Chapter 4.

Now, the last two requirements, which are the most difficult to understand and (probably due to this) also most often violated, are demonstrated by a numerical example in the following section.

3.4.3 Numerical example of a violation of requirements 6 and 7

The invariance principle and the difficulty of flow maximization are illustrated by means of a numerical example. Hereto, the demand proportional intersection model of Bliemer (2007) is used. The solution of this model can be expressed in terms of partial flows q_{ij} as in (3.12). Since CTF is satisfied, this translates to total flows q_i as in (3.21), or, equivalently, (3.22).

$$\begin{aligned} q_i &= \min_{j|f_{ij}>0}\left(S_i, \frac{R_j^i}{f_{ij}}\right) \\ with: \quad R_j^i &= \frac{S_{ij}}{\sum_{i'=1..I} S_{i'j}} R_j \end{aligned} \tag{3.21}$$

$$q_i = \min_{j|f_{ij}>0}\left(S_i, \frac{S_i}{\sum_{i'=1..I} S_{i'j}} R_j\right) \tag{3.22}$$

A supply constraint takes effect on all links i that compete for this supply, i.e. $i|f_{ij} > 0$. If several supplies impose a constraint on i, the most restrictive one determines all q_{ij} from i (due to CTF). For each supply constrained i, there is thus one j that imposes the strongest supply constraint on i. This can be understood from (3.22).

For this example, a standard intersection is considered, with four incoming and four outgoing links (Figure 3-3).

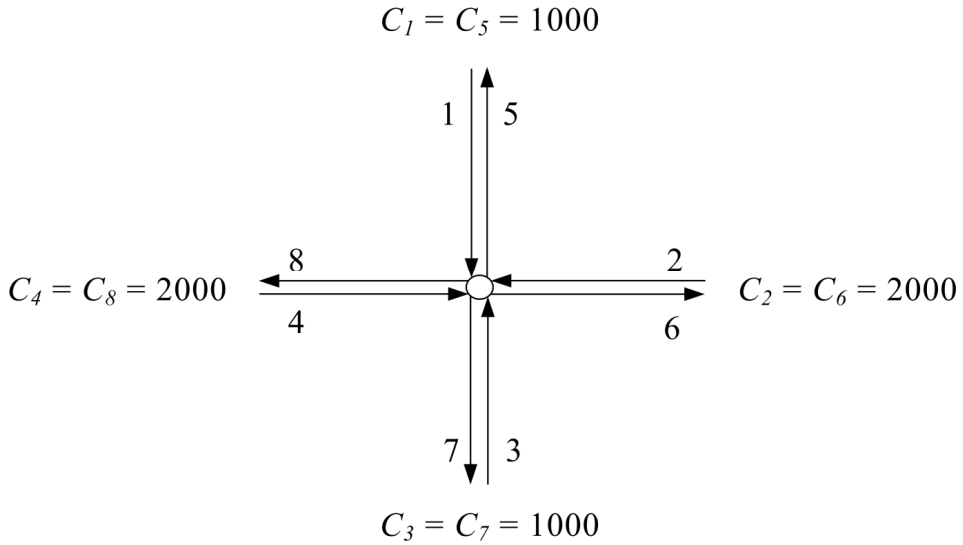

Figure 3-3: Standard general intersection

Only specific circumstances trigger a violation of the invariance principle and individual flow maximization. The demand in this example is specifically chosen to demonstrate these violations. The partial demands S_{ij} ($= f_{ij}S_i$) are given in Table 3-2, in veh/h.

Table 3-2: Partial demands S_{ij}

S_{ij}	5	6	7	8	S_i
1	0	50	150	300	500
2	100	0	300	1600	2000
3	100	100	0	600	800
4	100	800	800	0	1700
$\sum_i S_{ij}$	300	950	1250	2500	
R_j	1000	2000	1000	2000	

Assume that no congestion is spilling back from the outgoing links, so that $R_j = C_j$. From Table 3-2 and (3.22), it can easily be derived that the supplies R_7 and R_8 impose an equal reduction on the demands from all incoming links. This leads to the solution in Table 3-3.

Table 3-3: Flows q_{ij} (initial solution)

q_{ij}	5	6	7	8	q_i
1	0	40	120	240	400
2	80	0	240	1280	1600
3	80	80	0	480	640
4	80	640	640	0	1360
$\sum_i q_{ij}$	240	760	1000	2000	
R_j	1000	2000	1000	2000	

With this solution, congestion starts building up on all incoming links so that $S_i = C_i$. Conserving the turning fractions, the partial demands of Table 3-4 apply after an infinitesimal time increment.

Table 3-4: Partial demands S_{ij} after infinitesimal time increment

S_{ij}	5	6	7	8	S_i
1	0	100	300	600	1000
2	100	0	300	1600	2000
3	125	125	0	750	1000
4	118	941	941	0	2000
$\sum_i S_{ij}$	343	1166	1541	2950	
R_j	1000	2000	1000	2000	

From Table 3-4 and (3.22), it is clear that R_7 now constrains all its competing links (1, 2 and 4). Meanwhile, q_3 is constrained according to R_8^3. The new solution is shown in Table 3-5.

Table 3-5: Flows q_{ij} after infinitesimal time increment

q_{ij}	5	6	7	8	q_i
1	0	64.9	194.7	389.4	649
2	64.9	0	194.7	1038	1298
3	84.7	84.7	0	508.5	678
4	76.3	610.8	610.8	0	1298
$\sum_i q_{ij}$	226	760	1000	1936	
R_j	1000	2000	1000	2000	

This solution for link 1 now violates the invariance principle. Table 3-5 shows that the outflow from the queue on link 1 – which formed as a result of the first solution – is 649 veh/h. Since this is higher than the demand at the tail of the queue, which is still 500 veh/h (assuming unchanged conditions upstream), the queue on link 1 will dissolve. This intersection model thus produces a discontinuous solution, with the queue alternately growing and dissolving.

Furthermore, it can be observed that the inflow into outgoing link 8 is not maximized (1936 veh/h < R_8). Yet, q_3 is restricted assuming an insufficient supply R_8. As the other competing flows for R_8 (q_{18} and q_{28}) are more restricted by R_7, a bigger share of R_8 should have been attributed to q_{38}. In general: if q_{ij} of i is constrained in direction j, $q_{ij'}$ in other directions j' are constrained accordingly (due to CTF). Consequently, the other flows from other i' towards these j' experience less competition for $R_{j'}$. Hence, the share of $R_{j'}$ that is not used up by i should be distributed among the other flows towards j'. To account for this, it is necessary to consider the interaction of the various supply constraints when determining the flows. A mere distribution scheme is insufficient, but rather SCIR need to be formulated.

3.5 Supply constraint interaction rules (SCIR)

The seven requirements of the previous section are the basic rules that need to be followed when building an intersection model. They still do not fully define the model, however. In complete absence of a distribution logic, a simple flow maximizing model as those of Holden & Risebro (1995) and Coclite et al. (2005) is obtained. Such models are insufficiently realistic. As was demonstrated by the example in the previous section, a mere distribution scheme that does not take into account the interdependencies between different supply constraints is also insufficient. Rules describing both the distribution of supply over the competing flows and the interaction of constraints in that process are an obligatory addition to any DNL intersection model. These rules are denoted as Supply Constraint Interaction Rules (SCIR). Several definitions of the SCIR may be conceivable and plausible. In this section, the process of constructing SCIR and their function in the intersection model is generally explained. In Chapter 4, specific SCIR are proposed.

The SCIR should respect the requirements of Section 3.4.1. Also, they should realistically represent the aggregate driver behaviour at a (congested) intersection. The SCIR combine the demands and the (distribution of) supplies into a strongly coupled set of equations with a consistent solution. By doing so, it completes the first function of the intersection model. Generally speaking, in case of one or more active supply constraints, the SCIR need to answer two questions:

- For each flow: by which of the constraints is it limited? Usually, it is assumed that one most restrictive (demand or supply) constraint can be identified for each flow. This is also the case in all intersection models presented in this thesis. However, SCIR definitions in which multiple constraints together determine a flow are possible (e.g. as in Gibb, 2011). Particularly when introducing internal supply constraints this seems plausible. Indeed, flows often first have to pass some internal conflict (e.g. traffic light, crossing conflict) before being able to compete for the external supply of an outgoing link.

- For each supply constraint: how does the supply constraint restrict each of the competing flows? The supply constraints take effect on flows q_{ij} from different incoming links. Hence, the SCIR need to translate the supply constraints to individual constraints on each of these q_{ij} and therefore (due to CTF) on each q_i. This translation should be in accordance with the expected aggregate driver behaviour at a congested intersection. Typically[15], this is done by distributing a proportional, rightful share R_j^i of each R_j to each i. Formulation (3.11) provides a starting point in this process that, however, does not yet account for the interaction between supply constraints. Indeed, if some link i is unable to fully consume its rightful share of an R_j because it is more constrained by some other demand or supply constraint, the excess supply will be taken by other links i' (for which $f_{i'j} > 0$) that can still send more. Hence, each active supply constraint, i.e. that actively limits some flow(s), will be fully consumed (individual flow maximization).

In general, these two questions cannot be detached: whether or not some q_i is limited by some R_j depends on its share R_j^i. This share R_j^i again depends on which flows are limited by which constraint. Because of this interdependency, the SCIR may be difficult to define. Ideally, they should take into account all characteristics of driver behaviour, turning fractions, intersection geometry, priority rules and traffic control (if applicable).

Finally, it can be noticed that the difficulty of the mutual interaction between external supply constraints disappears if the CTF requirement would be released. Indeed, without CTF (i.e., without FIFO) the partial flows q_{ij} of each i are detached. Since each of these independent q_{ij} obviously only competes for one R_j, there is no mutual interaction between supply constraints in this case. Still, this should not be a reason for sacrificing realism by releasing CTF. Moreover, once internal supply constraints are introduced, even detached, non-FIFO q_{ij} are tied in several supply constraints (see Chapter 5).

[15] Gibb (2011) is to our knowledge the only valuable exception in the state-of-the-art.

3.6 Conclusion

First, an overview has been given of state-of-the-art of intersection models. Virtually all of these models either make unrealistic or artificial assumptions, are macroscopically impractical, or violate some or several fundamental modelling principles. In response, we formulate a set of seven generic requirements for first-order macroscopic DNL intersection models:

- General applicability to any number of incoming and outgoing links and any combination of boundary conditions
- Non-negativity of flows
- Conservation of vehicles
- Satisfying the demand and supply constraints
- Ensuring FIFO: Conservation of turning fractions (CTF)
- Individual flow maximization (each flow is actively constrained)
- Satisfying the invariance principle

This is an important contribution to the state-of-the-art, as it helps future model developers to build realistic, consistent intersection models.
Furthermore, the necessity and function of the SCIR are explained. The SCIR control the translation of the supply constraints into constraints on each competing flow, likely based on a proportional distribution. Also, it accounts for the interaction of the various constraints in this process, so that excess supply is reassigned. As opposed to the generic requirements, the SCIR should (ideally) be specific to a particular type of intersection and to the expected driver behaviour. As such, multiple plausible intersection models can be developed that build upon the generic requirements.

In this chapter, these fundamental modelling principles and components have been generally discussed. In the following chapter, an intersection model is drafted from this knowledge.

4

A GENERAL INTERSECTION MODEL

Following the theoretical discussion of the previous chapter, a new intersection model is presented. The SCIR define the intersection model's solution given the demand and external supply constraints, while ensuring consistency with the modelling requirements presented in Section 3.4.1. It is a general model, which is not designed specifically for one particular type of intersection. It can be applied to unsignalized intersections (with any number of incoming and outgoing links)[16]. Only for roundabouts, the presented supply distribution logic is not well suited.
Moreover, the presented model is simplified in the sense that internal supply constraints are not yet considered. The following chapters will build further upon this general model.

Section 4.1 generally defines the SCIR as a set of implicit equations that produces the intersection model's solution. An alternative, entirely equivalent SCIR formulation is given that simplifies graphical representation and that will be used in the subsequent chapters. In Section 4.2, the model is further specified by basing the distribution of supply on turning fractions and capacities. Yet another alternative SCIR formulation is presented that is convenient for developing an efficient iterative solution algorithm (in Section 4.3). The working of the intersection model and algorithm is demonstrated by means of the same example used earlier in Section 3.4.3. This numerical example shows that all requirements are satisfied and that the solution algorithm produces the exact solution of the presented model.

This chapter is an edited version of the second part of 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 45 (1), pp. 289-309.*

---

[16] For signalized intersections, this model could be used if the red and green phases are explicitly simulated.

4.1 General outline of the SCIR

In this section, the SCIR are defined. They implicitly define the flows q_i based on definitions of sets U_j that contain all incoming links i whose flows are constrained by the supply constraint in j - if any; some sets U_j may be empty. Furthermore, the SCIR prescribe how R_j is distributed over the constrained links $i \in U_j$. By doing so, the SCIR translate the supply constraints to individual constraints on each q_i. It turns out that no explicit definition of the sets U_j and the flows q_i can be written. Rather, they are interdependent and have to be implicitly defined. First, a formulation of the SCIR is given that clearly shows its construction and function. Then, an alternative SCIR formulation is presented which easily allows illustrative graphical representation. Also, it will act as the building format towards more complete and complex intersection models by adding internal supply constraints in Chapter 5.

4.1.1 Intuitive SCIR formulation

For a clear understanding, the demand and supply constraints are repeated first:

- The demand constraints imply:

$$q_i \le S_i \quad \forall i \tag{4.1}$$

- The supply constraint functions $\hat{R}_j$ express how the given supply R_j limits the partial flows q_{ij} that compete for it:

$$\hat{R}_j(\mathbf{q}) = \sum_i f_{ij} q_i - R_j \le 0 \quad \forall j \tag{4.2}$$

  where the vector **q** is composed of all $q_i | f_{ij} > 0$.

Then, we define SCIR that are based on set memberships. Sets U_j are defined that collect all incoming links i being constrained by supply constraint function $\hat{R}_j$. Writing "i is constrained by $\hat{R}_j$" as $i \mapsto \hat{R}_j$ allows to generally define U_j as:

$$U_j = \{i \mid i \mapsto \hat{R}_j\} \quad \forall j \tag{4.3}$$

More specifically, $i \mapsto \hat{R}_j$ implies that q_i is determined by the rightful share R_j^i (which is defined further below):

$$U_j = \{i \mid q_i = \frac{R_j^i}{f_{ij}}\} \quad \forall j \tag{4.4}$$

Analogously, sets could be introduced for demand constrained links. However, this is notationally cumbersome rather than convenient. Instead, it is simply stated that for a demand constrained link $q_i = S_i$. Otherwise, the flow is smaller than the demand ($q_i < S_i$). In this case, *i* must be supply constrained, i.e. *i* must belong to a set U_j (4.5).

$$q_i < S_i \Leftrightarrow \exists j \mid i \in U_j \quad \forall i \tag{4.5}$$

This implies that each q_i is determined by its most restrictive (demand or supply) constraint. Indeed, individual maximization of the flows q_i (see (3.19)) implies that each q_i has to be limited by some constraint. Otherwise, this flow could be trivially (and individually) increased by raising q_i until some constraint is met. This also implies that a constraint can only bind some flow(s) if it is completely used up:

$$\begin{aligned} U_j = \varnothing &\Leftrightarrow \hat{R}_j < 0 \quad \forall j \\ U_j \neq \varnothing &\Leftrightarrow \hat{R}_j = 0 \quad \forall j \end{aligned} \tag{4.6}$$

Also, note that from (3.12) follows logically that only *i* that wish to send flow to *j* (i.e. $f_{ij} > 0$) can belong to U_j. These are the links that compete for the supply in *j*.

The composition of the sets U_j depends, apart from the boundary conditions - the demand and supply constraints and turning fractions -, on how a supply constraint is translated into individual constraints on each of these competing q_i. For this, we adopt a proportional distribution of supply, which is universally applied in the state-of-the-art (see Section 3.3). This supply distribution should correspond to the aggregate driver behaviour (e.g. based on priorities, capacities, etc.). Hereby, it needs to be considered that if *i* is constrained by demand or some supply, it uses less than its rightful share of other supplies R_j it was competing for. Consequently, in order to avoid the underutilization of active supply as in the example in Section 3.4.3, the supply distribution must in turn depend on the composition of the sets U_j. Therefore, before distributing the available supply among $i \in U_j$, first the less-than-rightful shares of $i \notin U_j$ are subtracted from R_j. This reduced supply $\widetilde{R}_j$, which is available for distribution among the members of U_j, is defined as in (4.7). This way, the (mutual) interaction between demand and supply constraints is accounted for.

$$\widetilde{R}_j = R_j - \sum_{i \notin U_j} f_{ij} q_i \quad \forall j \tag{4.7}$$

This reduced supply $\widetilde{R}_j$ is then distributed among $i \in U_j$. Finite, strictly positive[17] priority parameters $\alpha_{i(j)}$ are introduced to govern the supply distribution. It would be plausible to assume that $\alpha_{i(j)}$ of *i* is different for each *j*. This will be elaborated on in Chapter 5. For now, we assume single-valued priority parameters α_i, which is a common assumption in the state-

[17] We define the priority parameters strictly positive for mathematical convenience. Algorithmically, zero priorities are allowable.

of-the-art (for external supply constraints). Then, the competitive strength of movement *ij* for R_j is given by $f_{ij}\alpha_i$, which determines the rightful share R_j^i of each R_j for each *i*:

$$R_j^i = \frac{f_{ij}\alpha_i}{\sum_{i' \in U_j} f_{i'j}\alpha_{i'}} \tilde{R}_j \quad if\ U_j \neq \varnothing \tag{4.8}$$

For $j \mid U_j = \varnothing$, no R_j^i have to be defined since in this case R_j is not determinant for the solution ($\hat{R}_j < 0$).
Then, each q_i is determined by its most restrictive demand or supply constraint and the solution of the intersection model can be written as:

$$q_i = \min_{j \mid U_j \neq \varnothing \ \&\ f_{ij} > 0} (S_i, \frac{R_j^i}{f_{ij}}) \quad \forall i \tag{4.9}$$

Now, if q_i in (4.9) is determined by a share of supply R_j^i, this implies (according to (4.4)) that $i \in U_j$. Hence, (4.9) actually defines the sets U_j based on the shares of supply R_j^i that in turn depend on the sets. In other words, the composition of sets U_j and the supply distribution, which leads to the solution of the flows, are interdependent; and (4.4)-(4.9) are implicit definitions. In order to solve this set of implicit definitions, an iterative solution algorithm is developed (see Section 4.3). It is straightforward to see that in (4.9), each q_i is actively constrained by either demand or supply and cannot be increased. In Appendix C it is shown that all the requirements in Section 3.4.1 are satisfied.

4.1.2 Alternative SCIR formulation

Now, an alternative, entirely equivalent formulation of the SCIR is presented. This formulation expresses the ratios of flows in terms of the ratios of the priory parameters (i.e. the priority ratios). This is convenient for graphical illustration of the solution space. Also, it is most comprehensible when elaborating on the solution non-uniqueness that is observed when introducing internal supply constraints in Chapter 5.

This formulation adopts the definitions (4.3)-(4.6) of the sets U_j. Of course, for each *i* that is not part of any U_j still holds that it is demand constrained ($q_i = S_i$). As explained before, finite, strictly positive priority parameters α_i determine the rightful shares R_j^i of each competing *i*. Now, it can be understood from (4.7)-(4.8) that the rightful share R_j^i of $i \in U_j$ increases if some other link(s) *i'* are more restricted by another constraint. This is described by (4.10). Due to CTF, this concerns the total flows q_i and not just the partial flows q_{ij}.

$$\frac{q_i}{q_{i'}} \geq \frac{\alpha_i}{\alpha_{i'}} \quad \forall i \in U_j, \forall i' \mid f_{i'j} > 0 \tag{4.10}$$

Although surprisingly concise, (4.10) and (4.4)-(4.6) completely define the SCIR. Indeed, comparing (4.8) and (4.9) for any link $i \in U_j$ and $i' | f_{i'j} > 0$ leads to (4.10). This implies that for two i that are constrained by the same supply constraint function $\hat{R}_j$, (4.10) holds for both. In this case, the priority ratio directly determines the flows:

$$\frac{q_i}{q_{i'}} = \frac{\alpha_i}{\alpha_{i'}} \quad \forall i, i' \in U_j \tag{4.11}$$

The graphical representation of the solution space as in Daganzo (1995) can now be extended to intersections with several incoming and outgoing links. Graphical clarity suggests limiting the example to a 2x2 case, as in Figure 4-1. Recall that due to CTF both inflows q_i are mutually dependent in both supply constraints. Hence, the solution is fully determined by the total inflows q_i. The partial flows $q_{ij} = f_{ij}q_i$ are directly derivable. Therefore, a two-dimensional representation of the solution space suffices.

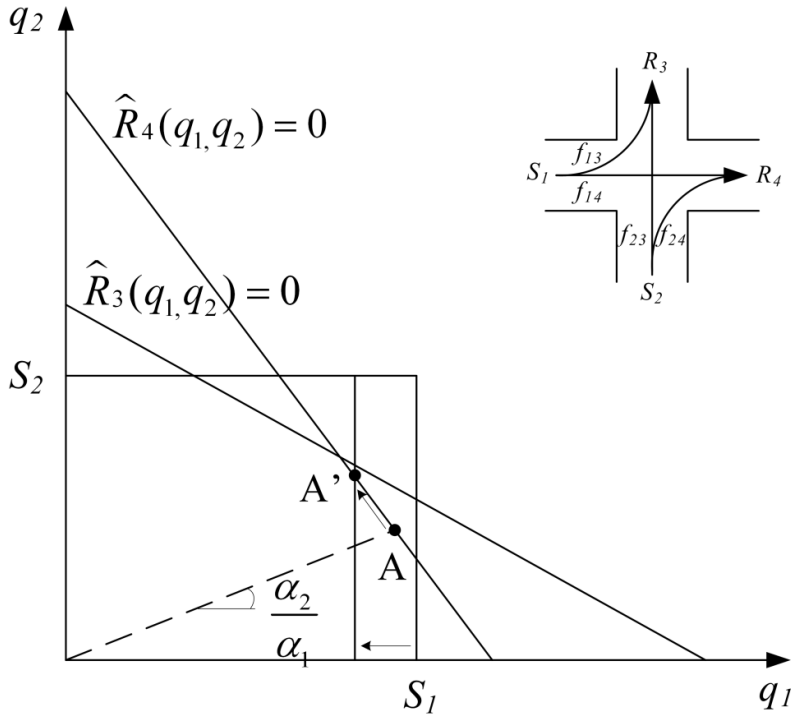

Figure 4-1: Example of the solution to the intersection model

Two possible solutions are demonstrated by shifting demand S_1. Other examples could of course be composed by varying other boundary conditions. If S_1 is high, both q_i are constrained by $\hat{R}_4$ and the solution A is determined by (4.11). Hence, this solution can be phrased in terms of the SCIR definitions as follows:

$$A: \quad \begin{array}{c} q_1 < S_1 \\ q_2 < S_2 \\ U_3 = \varnothing \Leftrightarrow \hat{R}_3 < 0 \\ U_4 = \{1,2\} \Leftrightarrow \hat{R}_4 = 0 \;\&\; \frac{q_2}{q_1} = \frac{\alpha_2}{\alpha_1} \end{array} \tag{4.12}$$

If S_1 is lowered, link 1 eventually becomes unable to fully consume its rightful share. The excess supply is reassigned to link 2. The solution point A' follows $\hat{R}_4$ along with the decreasing S_1. In A', (4.10) applies:

$$A': \quad \begin{array}{c} q_1 = S_1 \\ q_2 < S_2 \\ U_3 = \varnothing \Leftrightarrow \hat{R}_3 < 0 \\ U_4 = \{2\} \Leftrightarrow \hat{R}_4 = 0 \;\&\; \frac{q_2}{q_1} \geq \frac{\alpha_2}{\alpha_1} \end{array} \tag{4.13}$$

Note that lowering also S_2 would eventually lead to a fully demand constrained solution in which (4.10) would hold for neither link 1 nor 2 (as both U_3 and U_4 would be empty).

In conclusion this SCIR formulation allows finding the solution of the intersection model intuitively by following the priority ratio line. Whichever of the supply constraints is met or 'hit' by the priority ratio first, is identified as the most stringent one (at least, essentially, without considering the demand constraints). If a demand constraint is exceeded, the solution shifts by following the supply constraint function until this demand constraint is met.

This alternative formulation is added to improve the readers' understanding of the presented SCIR and intersection model. Moreover, in Chapter 5 this formulation will prove useful when introducing multiple-valued priority parameters α_{ij}.

4.1.3 Discussion

The general intersection model presented in this section also encapsulates the intersection model of Flötteröd & Rohde (2011). Furthermore, with only two incoming links *i* and one outgoing link *j*, or one *i* and two *j*, it reduces to respectively the merge and diverge model of Daganzo (1995). Finally, the core of the above SCIR - the distribution of supply through the proportionality of α_i (representing e.g. priorities, number of (turning) lanes or link capacities) - is shared by the majority of existing models. Compared to the above model, however, they exhibit some small incompleteness or flaw. This makes them violate some requirement - as discussed in Chapter 3. Alternative definitions of the SCIR than those presented in this section are conceivable. To the best of our knowledge, the only model compliant with all

requirements that has realistic SCIR substantially different from the above is that of Gibb (2011)[18].

This general model can be further detailed through the priority parameters α_i. In the next section, we set $\alpha_i = C_i$, so that the competitive strength of i for a supply in j is given by $f_{ij}\alpha_i = f_{ij}C_i = C_{ij}$. We call C_{ij} 'oriented capacities'; hence the term 'oriented capacity proportional distribution'.

4.2 Oriented capacity proportional distribution

First, the choice for an oriented capacity proportional supply distribution is motivated. Afterwards, yet another alternative formulation for the SCIR is presented, which is algorithmically convenient.

4.2.1 Motivation

Rather than being merely a mathematical construction to solve the intersection model, the SCIR are actually an aggregate representation of driver behaviour on intersections with at least one active supply constraint. An important behavioural difference between models therefore lies in how the priority parameters are defined. In Chapter 5, it is explained in more detail how the priority parameters can be used to model different driver behaviour in external and internal conflicts. In this chapter, we aim to define priority parameters such that a realistic distribution of the external supplies of outgoing links is obtained. Many previous studies have intuitively based the distribution on the demands S_i, which does not satisfy the invariance principle. Intersection models on the other hand that generalize the distribution scheme of Daganzo (1995) - based on priority constants – or that of Ni & Leonard (2005) – based on capacities – are valid in that sense. Hence, a distribution based on the priorities imposed by law at an intersection could be applied. However, priority rules for merging conflicts are obeyed less strictly under nearly-saturated and especially congested conditions (e.g. Troutbeck & Kako (1999) show this empirically at roundabouts). As empirical studies that investigate precisely how downstream supply is distributed at general intersections are lacking, we extrapolate the findings of studies at highway merges. Cassidy & Ahn (2005) find that congested highway merging occurs in a fixed, site-dependent ratio. This is confirmed by Bar-Gera & Ahn (2010), who show that this fixed ratio is well approximated by the ratio of the number of lanes of the incoming links of the merge. From a smaller data set, Ni & Leonard (2005) conclude that merging follows the ratio of the capacities, which is largely equivalent to number of lanes. This type of behaviour is often referred to as turn-taking, which can be regarded as the alternating use of the available supply by all competing flows with the turns or opportunities of each flow being determined by its capacity. Based on this, we specify the priority parameters as:

[18] However, in our experience extending Gibb's model to include internal conflicts is less straightforward, at least for driver behaviour other than turn-taking.

$$\begin{aligned} &\alpha_i = C_i \quad \forall i \\ &\Rightarrow f_{ij}\alpha_i = f_{ij}C_i = C_{ij} \quad \forall i, j \end{aligned} \tag{4.14}$$

Hence, the claim that i makes in the distribution of each R_j is determined by the capacity C_i and the turning fractions f_{ij}. The supply distribution is thus based on oriented capacities C_{ij}. For a clear understanding, we summarize the behavioural interpretation of having oriented capacities determine the supply distribution as follows. In case of an active supply constraint in j, several i are competing for a limited number of entering opportunities into j. The number of opportunities taken by a movement ij is proportional to:

- Turning fractions:
  a link i that sends all of its flow to j exploits twice as many opportunities than one with the same capacity that sends only half of its traffic to j and the other half to other j'.

- Capacity:
  a link i having more lanes, better visibility, or a smaller turning angle into j exploits more opportunities than one with less lanes, worse visibility, or a sharper turn into j.

Finally, note that this oriented capacity proportional intersection model – although independently developed – can be seen as an extension of the merge model of Ni and Leonard (2005).

4.2.2 Alternative oriented capacity proportional SCIR formulation

Now, let us consider an alternative formulation of the oriented capacity proportional SCIR. As will become clear in the next section, this formulation is algorithmically convenient. Definitions (4.3)-(4.7) apply here as well. Reduction factors a_j are introduced, defined as follows:

$$\begin{aligned} a_j &= \frac{\widetilde{R_j}}{\sum_{i \in U_j} C_{ij}} \quad \text{if } U_j \neq \varnothing \\ a_j &= 1 \qquad\qquad \text{if } U_j = \varnothing \end{aligned} \tag{4.15}$$

If $U_j = \varnothing$, the supply constraint of j is inactive ($\hat{R}_j < 0$). In this case, reduction factor a_j is by definition set to the default value 1. Furthermore, let us denote the outgoing link that imposes the most stringent supply constraint on i as $j^*(i)$:

$$j^*(i) = \arg\min_{j | f_{ij} > 0} a_j \tag{4.16}$$

Indeed, the most restrictive supply is that for which a_j is smallest. Equivalent to (4.9), the solution to the intersection model can be written as:

$$q_i = \min(S_i, a_{j^*(i)} C_i) \quad \forall i \tag{4.17}$$

In an oriented capacity proportional distribution, the level of reduction a_j has the following physical interpretation. Since the flows from $i \in U_j$ are supply constrained, a queue will build up on these links. As a result, links $i \in U_j$ wish to send flow with a rate of C_{ij} towards j. In an oriented capacity proportional distribution, the level of reduction a_j thus represents the overload of the (reduced) supply that is generated by the sum of these maximal claims C_{ij}. If a supply constraint in j is active, R_j is insufficient to grant the maximal, desired claims. Hence, only a fraction of this maximal claim can actually flow out of i towards j. This reduced outflow q_{ij} equals $a_{j^*(i)} C_{ij}$. With CTF, q_i then equals $a_{j^*(i)} C_i$.

4.3 Solution algorithm

The iterative algorithm described in this section determines the solution by finding the smallest level of reduction in each iteration. From (4.4), (4.8) and (4.15)-(4.17), it follows that:

$$a_j < a_{j'} \Rightarrow i \notin U_{j'} \quad \forall i \mid f_{ij} > 0 \tag{4.18}$$

In words, (4.18) states that q_i can never be constrained by a supply in j' if it also competes for a more restrictive supply in j. Hence, when calculating a reduced supply $\widetilde{R}_j$ according to (4.7), one only needs to account for the interaction with the demand constraints and the supply constraints that are more stringent than that of j. This implies that for the most stringent of all j – denoted as $\hat{j}$ (i.e. with the smallest level of reduction $a_{\hat{j}}$ over all j) -, only more restrictive demands may reduce $\widetilde{R}_{\hat{j}}$:

$$\widetilde{R}_{\hat{j}} = R_{\hat{j}} - \sum_{i \notin U_{\hat{j}}} S_{i\hat{j}} \tag{4.19}$$

All terms necessary to calculate $\widetilde{R}_{\hat{j}}$ are available as inputs to the intersection model. Therefore, if the solution algorithm can correctly pick out $\hat{j}$, the smallest level of reduction $a_{\hat{j}}$ can be calculated from (4.19) and (4.15). Given $a_{\hat{j}}$, the second smallest a_j can be calculated and so on, leading to the exact solution (4.17). It is explained below that the proposed solution algorithm indeed finds this exact solution. This is proven in Appendix D.

The iterative solution algorithm (k iterations, k starting at 0) for the presented intersection model can be subdivided into five steps. Note that the first two steps are initialization steps

that are only executed in the first run. The algorithm is first explained briefly in words and subsequently in commented code lines.

1. S_i, f_{ij}, C_i, R_j (to which $\widetilde{R}_j^{(0)}$ is initialized) - determined by the link model – are input to the intersection model. The sets $U_j^{(0)}$ are initialized as 'all $i|f_{ij} > 0$' and a set $J^{(0)}$ is initialized as 'all $j \mid \sum_i f_{ij} > 0$'. This set J thus contains all j that need to be considered by the algorithm. It is merely an aid to the algorithm and has no physical meaning in the solution.

2. The oriented capacities C_{ij} are determined from C_i and f_{ij} as in (4.14).

3. In iteration k, all $a_j^{(k)}$ for all j still under consideration ($j \in J^{(k)}$) are calculated from (4.15). The smallest is saved as $a_{\hat{j}}^{(k)}$; the corresponding j as $\hat{j}^{(k)}$. Recall that in the calculation of $a_j^{(k)}$, only the interaction with j' with smaller $a_{j'}$ (which have been definitively fixed in previous iterations) is to be considered. For $k = 0$, $\widetilde{R}_j^{(0)} = R_j \quad \forall j$.

4. $a_{\hat{j}}^{(k)}$ (determined in step 3) is imposed on all $i \in U_{\hat{j}}^{(k)}$.
   a) If link(s) i are found to be demand constrained ($q_i = S_i$), they are removed from all $U_j^{(k)}$. All j for which U_j is empty are removed from consideration (j removed from $J^{(k)}$). According to (4.7), each $\widetilde{R}_j^{(k)}$ is reduced by $q_{ij} = f_{ij}S_i$.

   b) If there are no demand constrained links in this iteration k, each $i \in U_{\hat{j}}^{(k)}$ is constrained according to $a_{\hat{j}}^{(k)}$. This results in $q_i = a_{\hat{j}}^{(k)} C_i$. Each $i \in U_{\hat{j}}^{(k)}$ is removed from all other $U_j^{(k)}$. All j for which U_j is empty, and $\hat{j}^{(k)}$ are removed from $J^{(k)}$. According to (4.7), each $\widetilde{R}_j^{(k)}$ is reduced by $q_{ij} = a_{\hat{j}}^{(k)} C_{ij}$.

5. Stop criterion: if there is no j left to consider in the next iteration, the algorithm stops. Otherwise, $k \to k+1$ and the algorithm returns to step 3.

In both of the substeps 4(a) and 4(b), the supplies R_j of j that are still under consideration (to be distributed in the next iteration) are reduced. With that, the corresponding levels of reduction a_j need to be recalculated in step 3 of the next iteration. The share that the treated link(s) i (that are no longer $\in U_j^{(k)}$) consume of these supplies R_j is less than their rightful share – either due to a more restrictive demand constraint (a) or due to a more restrictive supply constraint in $\hat{j}^{(k)}$ (b). Consequently, a bigger share of supply will be available for remaining i' (still $\in U_j^{(k)}$). As a result, the levels of reduction a_j can only increase (or be set to

1) as they are recomputed. This makes it possible to pick out $\hat{j}^{(k)}$ in each iteration and definitively fix the corresponding $a_{\hat{j}}^{(k)}$ (i.e. the smallest over all j still under consideration) and iteratively find the exact solution for all a_j and U_j, and thereby all q_i.
In code, the algorithm can be written as follows.

1. Retrieve link constraints & initialize supplies and sets

   For all *i, j*:
   S_i, f_{ij}, R_j and C_i are input from the link model.
   \\initialisation:
   $\widetilde{R}_j^{(0)} = R_j$ \\supply constraint determined by the link model
   $U_j^{(0)} = \{i \mid f_{ij} > 0\}$ \\add all *i* competing for R_j to initial set $U_j^{(0)}$
   $J^{(0)} = \{j \mid S_j > 0\}$ \\add all *j* towards which non-zero demand is directed to initial set *J*

2. Determine oriented capacities

   For all $i|S_i > 0$:
   $C_{ij} = f_{ij} C_i \quad \forall j$ \\calculate oriented capacities

3. Determine most restrictive constraint

   For all $j \in J^{(k)}$:
   \\level of reduction from *j*, according to (4.15):

$$a_j^{(k)} = \frac{\widetilde{R}_j^{(k)}}{\sum_{i \in U_j^{(k)}} C_{ij}} \tag{4.20}$$

   $a_{\hat{j}}^{(k)} = \min_{j \in J^{(k)}} a_j^{(k)}$ \\determine the most restrictive constraint
   Set $\hat{j}^{(k)} = j \mid a_j^{(k)} = a_{\hat{j}}^{(k)}$ \\save the corresponding *j* as $\hat{j}^{(k)}$

4. Determine flows of corresponding set $U_{\hat{j}}^{(k)}$ and recalculate $\widetilde{R}_j^{(k)}$

   \\step 4(a):
   If $\exists i \in U_{\hat{j}}^{(k)} \mid S_i \le a_{\hat{j}}^{(k)} C_i$: \\if at least one $i \in U_{\hat{j}}^{(k)}$ is demand constrained
      For all $i \in U_{\hat{j}}^{(k)} \mid S_i \le a_{\hat{j}}^{(k)} C_i$:
         $q_i = S_i$ \\solution for $q_{i(j)}$ found;
         $q_{ij} = f_{ij} S_i \quad \forall j$ \\equal to demand
      For all $j \in J^{(k)}$:

\\since i is demand constrained, it takes a share of each supply equal to $q_{ij} = S_{ij}$
\\and it is removed from all sets:

$$\widetilde{R}_j^{(k+1)} = \widetilde{R}_j^{(k)} - S_{ij} \tag{4.21}$$

Set $U_j^{(k+1)} = U_j^{(k)} \setminus \{i\}$

If $U_j^{(k+1)} = \varnothing$: \\if there are no i left that could be constrained by j

$a_j = 1$ \\definitive value for a_j
$U_j = \varnothing$ \\Set U_j is definitively empty
Set $J^{(k+1)} = J^{(k)} \setminus \{j\}$ \\$j$ is no longer considered

\\step 4(b):

ElseIf $S_i > a_{\hat{j}}^{(k)} C_i \quad \forall i \in U_{\hat{j}}^{(k)}$: \\if all $i \in U_{\hat{j}}^{(k)}$ are constrained by $\hat{j}^{(k)}$

For all $i \in U_{\hat{j}}^{(k)}$:

$q_i = a_{\hat{j}}^{(k)} C_i$ \\solution for $q_{i(j)}$ found;
$q_{ij} = a_{\hat{j}}^{(k)} C_{ij} \quad \forall j$ \\ equal to oriented capacity proportional share

For all $j \in J^{(k)}$:

\\since i is constrained in $\hat{j}^{(k)}$, it takes a share of each supply equal to
\\ $q_{ij} = a_{\hat{j}}^{(k)} C_{ij}$:

$$\widetilde{R}_j^{(k+1)} = \widetilde{R}_j^{(k)} - a_{\hat{j}}^{(k)} C_{ij} \tag{4.22}$$

If $j \neq \hat{j}^{(k)}$:

\\since i is constrained by $\hat{j}^{(k)}$, it is not constrained by other j;

\\hence, all $i \in U_{\hat{j}}^{(k)}$ are removed from all $U_j^{(k)} \mid j \neq \hat{j}^{(k)}$:

Set $U_j^{(k+1)} = U_j^{(k)} \setminus U_{\hat{j}}^{(k)}$

If $U_j^{(k+1)} = \varnothing$: \\if there are no i left that could be constrained by j

$a_j = 1$ \\definitive value for a_j
$U_j = \varnothing$ \\Set U_j is definitively empty
Set $J^{(k+1)} = J^{(k)} \setminus \{j\}$ \\$j$ is no longer considered

ElseIf $j = \hat{j}^{(k)}$:

\\the definitive level of reduction and set of $\hat{j}^{(k)}$ are saved:

$a_j = a_{\hat{j}}^{(k)}$
$U_j = U_{\hat{j}}^{(k)}$
Set $J^{(k+1)} = J^{(k)} \setminus \{\hat{j}^{(k)}\}$ \\ $\hat{j}$ is no longer considered

5. Stop criterion

   If $J^{(k+1)} = \varnothing$:
   Stop
   Else:
   $k = k + 1$
   Return to step 3

In Appendix D.2 it is shown that this algorithm needs at most I iterations - I being the number of incoming links - to find the solution.

4.4 Numerical example

The example of Section 3.4.3 is now examined with the new intersection model. The iterations of the algorithm in Section 4.3 are followed until the exact solution is reached.

$k = 0$:

1. The inputs from the link model (S_i, f_{ij} and R_j (=C_j)) are retrieved (see Table 4-1) and the sets U_j and J are initialized:

$$U_5^{(0)} = \{2,3,4\}$$
$$U_6^{(0)} = \{1,3,4\}$$
$$U_7^{(0)} = \{1,2,4\}$$
$$U_8^{(0)} = \{1,2,3\}$$
$$J^{(0)} = \{5,6,7,8\}$$

2. The oriented capacities C_{ij} are calculated from (4.14), see Table 4-1.

Table 4-1: First iteration (k = 0)

C_{ij}	5	6	7	8	C_i	S_i
1	0	100	300	600	1000	500
2	100	0	300	1600	2000	2000
3	125	125	0	750	1000	800
4	118	941	941	0	2000	1700
$\sum_{i \in U_j} C_{ij}$	343	1166	1541	2950		
$\widetilde{R}_j^{(0)}$	1000	2000	1000	2000		
$a_j^{(0)}$	2.92	1.72	0.649	0.678		

3. From (4.20), all $a_j^{(0)}$ are calculated (see Table 4-1). $a_7^{(0)}$ = 0.649 is selected as the smallest (= $a_{\hat{j}}^{(0)}$).

4. Since $S_1 = 500 < 649 = a_7^{(0)} C_1$, step 4(a) is entered. For now, only link 1 is demand constrained and the flows q_{1j} can be fixed (see 1st row in Table 4-4). The reduced supplies are recomputed according to (4.21) for the next iteration (see $\widetilde{R}_j^{(1)}$ in Table 4-2). Link 1 is removed from all sets:

   $U_5^{(1)} = \{2,3,4\}$
   $U_6^{(1)} = \{3,4\}$
   $U_7^{(1)} = \{2,4\}$
   $U_8^{(1)} = \{2,3\}$
   $J^{(1)} = \{5,6,7,8\}$

5. k = 1; return to step 3.

Table 4-2: Second iteration (k = 1)

C_{ij}	5	6	7	8	C_i	S_i
1	-	-	-	-	-	500
2	100	0	300	1600	2000	2000
3	125	125	0	750	1000	800
4	118	941	941	0	2000	1700
$\sum_{i \in U_j} C_{ij}$	343	1066	1241	2350		
$\widetilde{R}_j^{(1)}$	1000	1950	850	1700		
$a_j^{(1)}$	2.92	1.83	0.685	0.723		

<u>k = 1:</u>

3. From (4.20), all $a_j^{(1)}$ are calculated (see Table 4-2). Since extra supply has become available due to the restrictive demand constraint S_1, the levels of reduction have increased compared to the first iteration. The algorithm sets $a_{\hat{j}}^{(1)} = a_7^{(1)} = 0.685$.

4. Now, for all links that compete for R_7, i.e. $U_7^{(1)} = \{2,4\}$, $S_i > a_{\hat{j}}^{(1)} C_i$ and thus the algorithm enters step 4(b). The flows of links 2 and 4 are fixed to $a_{\hat{j}}^{(1)} C_{ij}$ (see 2nd and 4th row in Table 4-4).

The reduced supplies are recomputed according to (4.22) for the next iteration (see $\widetilde{R}_j^{(2)}$ in Table 4-3). Links 2 and 4 are removed from all sets except U_7, which is fixed to $U_7^{(1)}$. Also, a_7 is fixed to $a_7^{(1)}$ and link 7 is not further considered:

$U_5^{(2)} = \{3\}$
$U_6^{(2)} = \{3\}$
$U_7 = U_7^{(2)} = \{2,4\}$
$U_8^{(2)} = \{3\}$
$J^{(2)} = \{5,6,8\}$

5. $k = 2$; return to step 3.

Table 4-3: Third iteration ($k = 2$)

C_{ij}	5	6	7	8	C_i	S_i
1	-	-	-	-	-	500
2	-	-	-	-	-	2000
3	125	125	0	750	1000	800
4	-	-	-	-	-	1700
$\sum_{i \in U_j} C_{ij}$	125	125	0	750		
$\widetilde{R}_j^{(2)}$	851	1306	0	604		
$a_j^{(2)}$	6.81	10.45	-	0.805		

<u>$k = 2$:</u>

3. See Table 4-3 for $a_j^{(2)}$. $a_j^{(2)} = a_8^{(2)} = 0.805$.

4. Now, $a_8^{(2)} C_3 = 805 > 800 = S_3$, rendering link 3 demand constrained (q_{3j} fixed to S_{3j}; see 3rd row in Table 4-4). All reduced supplies become zero, a_5, a_6 and a_8 are fixed to 1 and all sets are fixed to:

$U_5 = \varnothing$
$U_6 = \varnothing$
$U_7 = \{2,4\}$
$U_8 = \varnothing$
$J^{(3)} = \varnothing$

5. $J^{(3)} = \varnothing$, so the algorithm stops.

The model solution (Table 4-4) is found after only three iterations. The model does not predict congestion on link 3, contrary to the demand proportional model of Bliemer (2007); see Section 3.4.3.

Table 4-4: Resulting flows q_{ij}

q_{ij}	5	6	7	8	q_i	S_i
1	0	50	150	300	500	***500***
2	68.5	0	205.5	1096	1370	2000
3	100	100	0	600	800	***800***
4	80.6	644.5	644.5	0	1369.6	1700
$\sum_i q_{ij}$	249.1	794.5	1000	1996	4039.6	
R_j	1000	2000	***1000***	2000		
a_j	1	1	0.685	1		

This solution satisfies the invariance principle. Hence, it is not subject to discontinuous changes. Table 4-4 illustrates that flows are maximized: no individual flow can be increased, since it is either limited by demand (dashed horizontal arrows), or by a supply that is distributed proportional to oriented capacities (vertical arrows) and that – due to CTF – limits all flows from the constrained incoming links (full horizontal arrows).

4.5 Conclusion

In this chapter, a basic, general intersection model is built from the foundations laid in the previous chapter. SCIR are defined that distribute the supplies of outgoing links, accounting for the interactions with the demand and other supply constraints. An intersection model is obtained that correctly fulfils the first function of intersection models in DNL (see Section 2.1) under external constraints.

First, the SCIR are presented in general terms in Section 4.1. Sets are introduced that collect incoming links of which the flow is constrained by each particular supply constraint. Based on set memberships, restrictive supplies are distributed by proportion of priority parameters. Hereby, it is taken into account that flows that are more constrained by some other supply or by demand take less than their rightful share and that this excess supply needs to be attributed among the remaining competing flows. Importantly, the composition of the sets (i.e. which flow is limited by which constraint) and the supply distribution are interdependent. This interdependency translates into a set of implicit definitions ((4.4)-(4.9)) that unambiguously defines a consistent solution for the intersection model.
The general formulation in Section 4.1 can be further specified through the priority parameters. In this generic form, it also encapsulates the independently developed models of Flötteröd & Rohde (2011) and Gentile (2010)[19]. It is applicable to various types of

[19] The model of Gentile (2010) applies multiple-valued priority parameters α_{ij} instead of single-valued α_i, however (see Chapter 5).

unsignalized intersections (but not roundabouts) and can also represent merges and diverges; in this case it reduces to the model of Daganzo (1995). Based on the scarcely available empirical research on supply distribution (at highway merges), a more specific model is suggested in Section 4.2. This model applies an oriented capacity proportional distribution of supply. This means that the competitive strength of an incoming link, determining its rightful share of supply, is influenced by the physical link capacity and the turning fraction towards this supply. For this specification, an alternative model formulation is suggested. This convenient formulation is deployed by the solution algorithm of Section 4.3. The numerical example of Section 4.4 shows that the algorithm provides the solution of the presented model in at maximum I iterations (I being the number of incoming links).

In conclusion, the intersection model presented in this chapter complies with the seven requirements of Section 3.4.1, adopts behaviourally realistic SCIR and produces a consistent solution. Hence, unlike the vast majority of existing intersection models, it properly fulfils the first function of the intersection model in first-order macroscopic DNL models. Still, it is simplified in the sense that it is limited to external constraints. Accounting for internal constraints imposed by limited supply of the intersection itself constitutes the second function of the intersection model. For highway merges and diverges or under certain traffic conditions (e.g. heavy congestion spilling back onto the intersection), it is justified to neglect the intersection supply constraints in favour of a simpler intersection model such as the one presented here. Especially in urban applications, however, internal supply constraints are vital in order to obtain a realistic representation of traffic flows. The next chapter discusses the difficulties, notably regarding solution uniqueness, that arise upon adding such constraints to the basic model that has been presented here.

5

INTERNAL SUPPLY CONSTRAINTS AND SOLUTION NON-UNIQUENESS

In the previous two chapters, the focus has been limited to external constraints. In this chapter, another level of detail and realism - but also complexity - is reached by adding internal supply constraints. Following and extending the discussion in Flötteröd & Rohde (2011), the possible occurrence of non-unique flow solutions is thoroughly discussed. In fact, this chapter is the result of joint research with Gunnar Flötteröd, in which ours and his previously parallel and independent research paths are united.

Section 5.1 translates the observed and expected driver behaviour in conflicts into a general formulation of the internal supply constraints and their distribution. This extends the general model presented in Section 4.1. The additional modelling formulations in Section 5.1 are general by intention. While allowing various ways of detailing, it enables a general elaboration on the observation of non-unique solutions of the intersection model in Section 5.2. Finally, Section 5.3 discusses the possibility of diverting from point-like to spatial intersection modelling. It is shown that also spatial models can produce different solutions under instantaneously identical boundary conditions, and that the result is inherently determined based on the history of flows. Moreover, some unrealistic and undesirable model behaviour is identified.

This chapter is an edited version of Corthout, R., Flötteröd, G., Viti, F. & Tampère, C.M.J. (2012). Non-unique flows in macroscopic first-order intersection models. *Transportation Research Part B 46 (3), pp. 343-359.*

5.1 From driver behaviour to internal supply constraints

The second function of intersection models, the infliction of additional supply constraints due to conflicts within the intersection itself, is often neglected in the state-of-the-art. Most macroscopic DNL intersection models are therefore currently not well suited for urban and regional applications. This is because these internal conflicts are often largely responsible for the traffic problems in such networks.
Here, the general intersection model of Section 4.1 is extended. As such, we propose the inclusion of internal supply constraints analogous to how external supply constraints of outgoing links are treated.

Section 5.1.1 first inquires into the various behavioural assumptions that can be made regarding intersection conflicts. This elaboration comprises both external and internal conflicts. This serves the translation of these behavioural considerations into (the distribution of) internal supply constraints in Section 5.1.2.

5.1.1 Driver behaviour in crossing and merging conflicts

Currently, there is a lack of empirical knowledge on driver behaviour at intersections under varying conditions. Therefore, it is difficult to outline the behavioural assumptions to be made in the model. In the following, we attempt to construct a classification of (expected) driver behaviour at intersections. This is drafted partly from existing empirical work, mostly at highway junctions, and partly from our expert judgment. Hence, until validated empirically, our findings are susceptible to discussion. Three types of driver behaviour are identified in solving intersection conflicts: absolute compliance to priority rules, limited compliance to priority rules and turn-taking. Different driver behaviour can be observed depending on traffic load, intersection type and geometry, and personal and cultural differences.

Absolute compliance to priority rules covers cases where an imposed ordering of the movements, rendering priority to some prioritized (or major) movements over other minor movements, is strictly obeyed by drivers. According to the traffic rules, this behaviour should apply to merging and crossing conflicts at priority-controlled (and the remaining conflicts at signal-controlled) intersections under all circumstances. In reality, it can be observed only at low traffic volumes (under-saturated).
As traffic volumes increase, priority rules are less strictly obeyed due to politeness from prioritized and forcing from non-prioritized drivers. This limited compliance commonly arises as the minor movements are in nearly- to over-saturated conditions. Empirical documentation is given for roundabout merging in Troutbeck & Kako (1999) and for crossing and merging flows in Brilon & Miltner (2005).
In over-saturated conditions, the behaviour typically tends towards turn-taking, in particular for merging conflicts. Turn-taking can be regarded as the alternating use of the available supply by all competing movements. Hereby, the ‘turns’ or opportunities that are available for each movement are determined by the movements’ outflow capacities, possibly reduced by other conflicts (crossing, control). Empirical studies of merging behaviour in over-saturated conditions at general intersections are lacking. Cassidy & Ahn (2005) find that

congested highway merging occurs in a fixed, site-dependent ratio, independently of the available downstream supply. This is further confirmed in Bar-Gera & Ahn (2010), who show that this fixed ratio is well approximated by the ratio of the number of lanes of the incoming links of the merge. From a smaller data set, Ni & Leonard (2005) conclude that merging follows the ratio of the capacities. Finally, AWSC intersections actually prescribe turn-taking for both merging and crossing conflicts. Although for AWSC intersections this type of behaviour is in compliance with the priority rules, the behaviour at AWSC conflicts is categorized as turn-taking and the concepts of limited and absolute compliance are preserved for situations as described above.

Naturally, the aggregate driver behaviour to be captured by macroscopic intersection models may be a mixture of different types of behaviour. In Figure 5-1, a suggestion is provided on how the aggregate driver behaviour at various intersection conflicts could be classified. This classification is partially derived from available empirical studies, as discussed above. Since not all types of conflicts have been (sufficiently) documented, additional assumptions are necessary. As stated before, until validated empirically, this classification is susceptible to discussion.

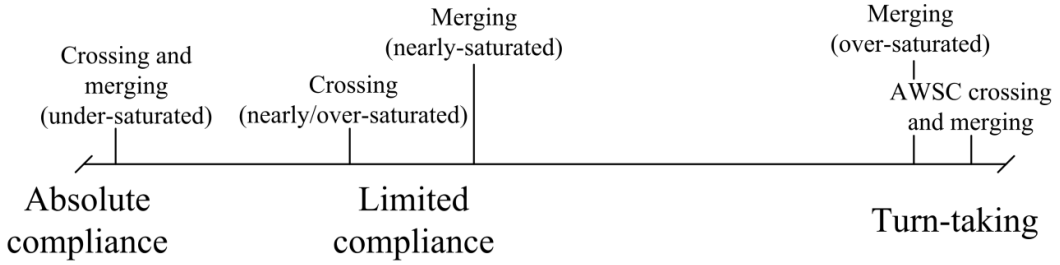

Figure 5-1: Driver behaviour in intersection conflicts

5.1.2 Modelling internal supply constraints in DNL intersection models

Although the internal supply constraints are largely responsible for the traffic problems in many regional and urban networks, they are rarely considered in macroscopic DNL intersection models. First, an overview of existing intersection models that do include internal supply constraints is presented. Afterwards, it is explained how we propose to include the internal supply constraints into the intersection model.

5.1.2.1 State-of-the-art

First, we mention that spatial intersection models that account for internal conflicts within the intersection have been presented by Buisson et al. (1995) – see also Buisson et al. (1996) - and Chen et al. (2008), both of which do not satisfy the invariance principle. Apart from the digression on spatial models in Section 5.3, we continue to focus on point-like intersection models.

The internal supply constraint functions are more difficult to define than the external supply constraint functions $\hat{R}_j$. For example, unlike $\hat{R}_j$, it is conceivable that the internal supply constraint functions are non-linear. For $\hat{R}_j$, definition (3.10) is universally adopted, which is typically distributed among the competing flows via SCIR that prescribe some proportionality (see Chapters 3-4). Such a universally adopted approach does not exist in the DNL state-of-the-art for modelling internal conflicts.
In the context of capacity estimation of single intersections (e.g. for design purposes), many studies have been performed since the 1960's - Chapters 8 and 9 in Gartner et al. (2000) summarize earlier work – to determine the restriction on minor movements from yielding to higher prioritized movements in crossing and merging conflicts. The majority of the (non-empirical) research efforts are communally categorized as gap acceptance theory. Gap acceptance theory aims to calculate the probability that drivers will accept gaps of various sizes when deciding to cross or merge. From this, the capacity of a minor movement – or, in our terminology, the restriction on this movement flow due to the internal supply constraint – is determined. Thereto, a distribution function of the gaps is considered. This distribution function typically depends on the prioritized flows, which are considered known. For most of the models described below, the internal supply constraint functions are derived from gap acceptance theory. An alternative, newer theory is conflict theory (Brilon & Wu, 2001; Brilon & Miltner, 2005). In Chapter 6, an explanation of conflict theory will be provided. An important difference with gap acceptance theory is that the restrictions on the minor flows are determined directly from the prioritized flows rather than via gap distributions. In that respect, a similar formulation has been proposed by Bovy (1991). To our knowledge, the only macroscopic intersection model that adopts the definition of Bovy (1991) to define the internal supply constraints is that of Bezembinder & Brandt (2007) - see further below.
When embedding these theories into DNL intersection models, two issues in particular need to be considered. Firstly, both gap acceptance and conflict theory are designed for free flowing conditions on the prioritized links. Gap acceptance theory in particular loses some validity in nearly-saturated and definitely in over-saturated conditions. Conflict theory is, at least theoretically, more easily extendable to over-saturated conditions; but this has not been validated empirically. Secondly, the prioritized flows are considered externally given. In a DNL intersection model, however, the prioritized flows may also be subject to (internal) supply constraints and therefore not known beforehand. As will be explained below, this is often erroneously solved by making the internal supply constraint functions dependent on the prioritized demands instead of the actual flows.

Besides which of the above theories to choose, it needs to be decided how to incorporate the effect of the internal conflicts into the intersection model. Ngoduy (2006)[20] – see Ngoduy et al. (2005) for an earlier version – incorporates the restriction due to internal conflicts as an additional reduction of the shares of external supply R_j^i. The solution to the intersection model of Ngoduy (2006) can, in essence, be formulated as:

[20] The author claims to divide the inner space of the intersection into spatial zones, suggesting that this is a spatial intersection model. However, this spatial discretization is not explained or demonstrated. Moreover, the internal conflicts are simultaneously considered, which would be impossible in a spatial model as in Section 5.3. Therefore, we consider it here as a point-like intersection model.

$$q_{ij} = \min(S_{ij}, P_{ij} R_j^i) \tag{5.1}$$

In (5.1), P_{ij} denotes the probability that the minor flow q_{ij} can cross, which is determined from gap acceptance theory with the cumulative gap distributions derived from the microscopic safety-distance model of Jepsen (1998). Apart from a violation of the invariance principle from the demand side[21], the CTF requirement and flow maximization, this model does not satisfy the invariance principle from the supply side. This is caused by the fact that the rightful shares R_j^i are further reduced by the internal conflicts. This way, if a flow is limited by R_j, it is not ensured that this supply is fully used up (q_j may be $< R_j$). From this, it can be concluded that combining the internal supply constraints and the (distribution of) external supplies R_j as in (5.1) is not a correct way to include the internal conflicts into the model. In several other intersection models, such as those of Yperman et al. (2007), van Hinsbergen et al. (2009) and Bezembinder & Brandt (2007) - see also Raadsen (2010) and Schilpzand (2008) -, the internal conflicts are considered as a constraint separate from the external supplies R_j. However, these models consider the internal supply constraint functions dependent on the demands of the prioritized movements. In their most general and basic form, the internal supply constraint functions of the above mentioned models can be formulated as:

$$q_i \leq \widehat{N}(S_{i'}) \tag{5.2}$$

Again, such formulations are prone to violations of the invariance principle. Moreover, it may cause an overestimation of the constraint imposed on the minor flow (as highlighted in Flötteröd & Rohde, 2011) if the prioritized flows are supply constrained themselves (i.e. $q_{i'} < S_{i'}$). The model of Bezembinder & Brandt (2007) starts from constraints dependent on the demands as in (5.2). If $q_{i'} < S_{i'}$, the calculations of the internal supply constraints are iterated with demand values lowered by 5%. Although this heuristic approach may be practically valuable in some or many cases, it does not qualify as a theoretically sound model as we are looking for here.

In Yperman (2007), the model of Yperman et al. (2007) has been improved upon so that the internal supply constraints on minor flows are dependent on the prioritized flows instead of the demands:

$$q_i \leq \widehat{N}(q_{i'}) \tag{5.3}$$

Yperman (2007) adopts these internal supply constraint functions in a simplified way from the conflict theory of Brilon & Miltner (2005). The merging conflicts for outgoing links are included in the internal supply constraint functions (5.3) and, separately, as external supply constraints. Hence, they are accounted for twice in the model. Yperman (2007) does actually not provide a sound formulation for his model. Rather, it is stated that the solution for all flows can be found by first calculating the prioritized flows and using these to solve lower

[21] This stems from the fact that the rightful shares R_j^i are determined demand proportionally.

ranked flows. It is claimed that by iteratively returning to the higher ranked flows, a solution (consistent with CTF) can be obtained. However, this procedure could cause the algorithm to switch between multiple solutions. This problem also occurs and is recognized in Flötteröd & Rohde (2011), where a uniquely converging solution algorithm in the presence of internal supply constraints is presented only for the special case in which the incoming links can be ranked such that the flows of higher ranked links are independent from those of lower ranked links. Indeed, the impossibility to design a solution algorithm that converges to a unique solution leads to the (seminal) investigation of solution uniqueness presented in that article. Motivated by the identification of a simple (three-legged) configuration that already yields non-unique flows; the authors also provide a heuristic algorithm with guaranteed convergence towards a compromise solution.

5.1.2.2 A general formulation of the internal supply constraints

The most viable models described in the previous section introduce internal supply constraints as separate functions that depend on the flow solution of the prioritized movements, in the form of (5.3). Usually the gap acceptance or conflict theory that forms the basis of these internal constraints assumes absolute compliance with the priority rules. Some modifications exist that define a less severe restriction on the minor flow due to limited (rather than absolute) compliance (see Troutbeck & Kako, 1999; Brilon & Miltner, 2005 and Chevallier & Leclercq, 2007 – only the latter work is in the context of DNL). Still, this approach more naturally relates to absolute compliance. Indeed, in (5.3) the minor flow is limited by the prioritized flow via their shared internal conflict, but not the other way around. With regard to modelling different driver behaviour, the following general formulation of the internal supply constraint function for an internal conflict k is more flexible.

$$\widehat{N}_k(\mathbf{q}) \leq 0 \tag{5.4}$$

where $\mathbf{q}$ is the vector of all $q_i | f_{ik} > 0$. The (turning) fraction f_{ik} denotes – entirely analogous to f_{ij} – the fraction of the total flow q_i that takes part in the internal conflict k. Obviously, each f_{ik} must be equal to a fraction f_{ij}, or to a part or the sum of some f_{ij}. Hence, CTF must hold for the fractions f_{ik} as for f_{ij}. In the remainder of the text, whenever outgoing links j and internal conflicts k are simultaneously considered, we uphold a consistent numbering of $i = 1..I$, $j = I+1..I+J$ and $k = I+J+1..I+J+K$ so that ambiguity between f_{ij} and f_{ik} (and other variables defined for j and k hereafter) is avoided.

The internal supply constraint function $\widehat{N}_k$ constrains all flows q_i that participate in the internal conflict k ($f_{ik} > 0$), contrary to formulations of restricted minor flows as functions of unrestricted prioritized flows as discussed before. Again, the internal supply constraint functions $\widehat{N}_k$ are to be derived from gap acceptance or conflict theory. Exact definitions of these functions $\widehat{N}_k$ are reserved for Chapter 6, when specific intersection models are presented.

The benefit of the general formulation (5.4) is its analogy with the external supply constraint functions $\widehat{R}_j$. It is proposed to generally extend the model of Section 4.1 with such internal supply constraint functions. The distribution of internal supply can then governed by priority

parameters, analogously to external supply. Analogous to the sets U_j in Chapter 4, sets U_k are introduced that contain all incoming links i whose flows are constrained by the internal supply constraint in k. We use the SCIR formulation of Section 4.1.2 here, which is the most convenient one for continuing the discussion in this chapter.

$$U_k = \{i \mid i \mapsto \widehat{N}_k\} \quad \forall k \tag{5.5}$$

$$\begin{aligned} U_k = \varnothing &\Leftrightarrow \widehat{N}_k < 0 \quad \forall k \\ U_k \neq \varnothing &\Leftrightarrow \widehat{N}_k = 0 \quad \forall k \end{aligned} \tag{5.6}$$

$$\frac{q_i}{q_{i'}} \geq \frac{\alpha_{ik}}{\alpha_{i'k}} \quad \forall i \in U_k, \forall i' \mid f_{i'k} > 0 \tag{5.7}$$

$$\frac{q_i}{q_{i'}} = \frac{\alpha_{ik}}{\alpha_{i'k}} \quad \forall i, i' \in U_k \tag{5.8}$$

Note that, in Section 4.1, we used single-valued priority parameters α_i to distribute external supplies (as is common in the state-of-the-art). Now, for internal supply constraints, it is much more realistic to assume multiple-valued priority parameters α_{ik} for a link i. These α_{ik} may differ depending on which conflict k is considered. Indeed, priority rules often subscribe that some movement(s) are prioritized, while another movement coming from the same link i (typically the left-turn) has to give way. For generality, we also consider multiple-valued α_{ij} instead of single-valued α_i from here on.

Below, it is more clearly explained and exemplified how the priority parameters can be used to model different driver behaviour. This discussion again draws from the graphical representation of an intersection model's solution space as in Section 4.1.2 and as used by Daganzo (1995). A simple intersection with one crossing conflict ($\widehat{N}_3$) is considered in Figure 5-2. Firstly, the demand constraints S_1 and S_2 limit the flow from their respective incoming link. Secondly, $\widehat{N}_3(q_1, q_2) \leq 0$ limits q_1 and q_2, which compete for the shared internal supply. Say that for this intersection, the priority rules prescribe that q_2 has priority over q_1. Absolute compliance (case A) can then be modelled by setting α_{13}^A arbitrarily small and $\alpha_{23}^A = 1$. Consequently, q_2 consumes as much as possible ($q_2 = S_2$) and leaves the remaining internal supply for q_1. Hence, point A is the solution in this case. Note that this case corresponds to formulation (5.3). Limited compliance (case L) can be modelled by less extreme priorities, i.e. $\alpha_{13}^L > 0$ and $\alpha_{23}^L < 1$. If turn-taking behaviour applies to the crossing conflict (case T), the priority parameters are based on the number of lanes or capacities of the incoming links, e.g. $\alpha_{13}^T = C_1$ and $\alpha_{23}^T = C_2$.

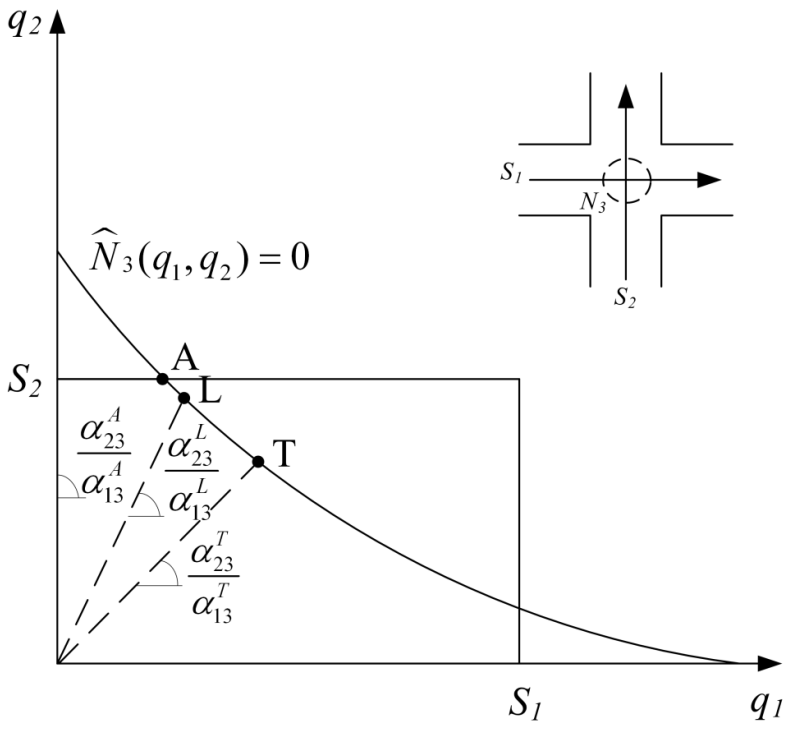

Figure 5-2: Priority parameters controlling driver behaviour in supply distribution

In the above, the priority parameters are presented as constant values. However, α_{ij} and α_{ik} could also be defined as functions of the flows. This way, transitions in driver behaviour, e.g. depending on the saturation level (see Figure 5-1), may be accounted for by these priority functions. This can be understood from Figure 5-2 by imagining a monotonically increasing function $\alpha_{23}(\mathbf{q})/\alpha_{13}(\mathbf{q})$ with varying slope, for instance starting as $\alpha_{23}^A/\alpha_{13}^A$ in the origin and approaching $\alpha_{23}^T/\alpha_{13}^T$ as the flows increase. If priority functions are introduced, the model formulation should be slightly adapted, so that instead of (5.8), (5.9) holds. That is, the value of the priority functions at some point **q** dictates the allowed relative increase from this point of two flows competing for an (internal) supply.

$$\frac{\partial q_i}{\partial q_{i'}} = \frac{\alpha_{ik}(\mathbf{q})}{\alpha_{i'k}(\mathbf{q})} \quad \forall i,i' \in U_k \tag{5.9}$$

The possibility of defining functions $\alpha_i(\mathbf{q})$ that govern the distribution of external supplies has been suggested (but not further pursued) by Flötteröd & Rohde (2011). Additional research on how to properly define the priorities as functions of the flow would be highly valuable. It should be noted that the approach in Flötteröd & Rohde (2011) of incrementally transferring the flows has algorithmic advantages in this case. The remainder of this dissertation focuses on constant priority parameters, which already create a great deal of complexity.

In summary, this section adds general internal supply constraint functions to the intersection model. The use of priority parameters to model different types of driver behaviour in the distribution of (internal) supply is demonstrated. This demonstrates that our general formulation of the internal supply constraint functions (5.4) encapsulates the more traditional form (5.3), while providing more flexibility regarding the modelling of driver behaviour other than absolute compliance. For now, the depth and complexity of this discussion is limited to the level necessary to introduce the modelling concepts and to allow a general elaboration on solution uniqueness in the next sections.

5.2 Solution non-uniqueness in intersection models

Based on the formal intersection modelling framework of the previous section, it is now shown that solution uniqueness in such models is anything but trivially guaranteed. As this model framework generally extends the proportional distribution of external supply as implemented in most state-of-the-art intersection models to internal supply constraints, the analysis in this section is general as well.

In Section 5.2.1, it is shown by means of a simple example that realistic behavioural assumptions can lead to multiple flow solutions under identical boundary conditions. In Section 5.2.2, technical conditions for solution (non-)uniqueness are formulated. In Section 5.2.3, the implications of these findings are discussed; the main conclusion being that the uniqueness condition for the priority parameters in the model does not (always) intuitively correspond to realistic driver behaviour. Therefore, careful consideration is needed on how to remedy the non-uniqueness and obtain one unambiguous result from the model. Section 5.2.4 suggests some approaches.

5.2.1 Explanatory example

Figure 5-3 shows a 2x4 intersection, which can be interpreted as a standard 4-leg intersection where only two inflows are considered[22]. Two crossing conflicts can be identified. The priority rules for such conflicts typically state that the left-turning movement has to give way to the straight movement coming from the opposite link, i.e. $\alpha_{23} = \alpha_{14} = 1$ and α_{13} and α_{24} arbitrarily small. In addition to the demand constraints, this results in the solution space as shown in Figure 5-3.

[22] Assume that the other flows can be neglected (at least for determining the solution of the considered flows).

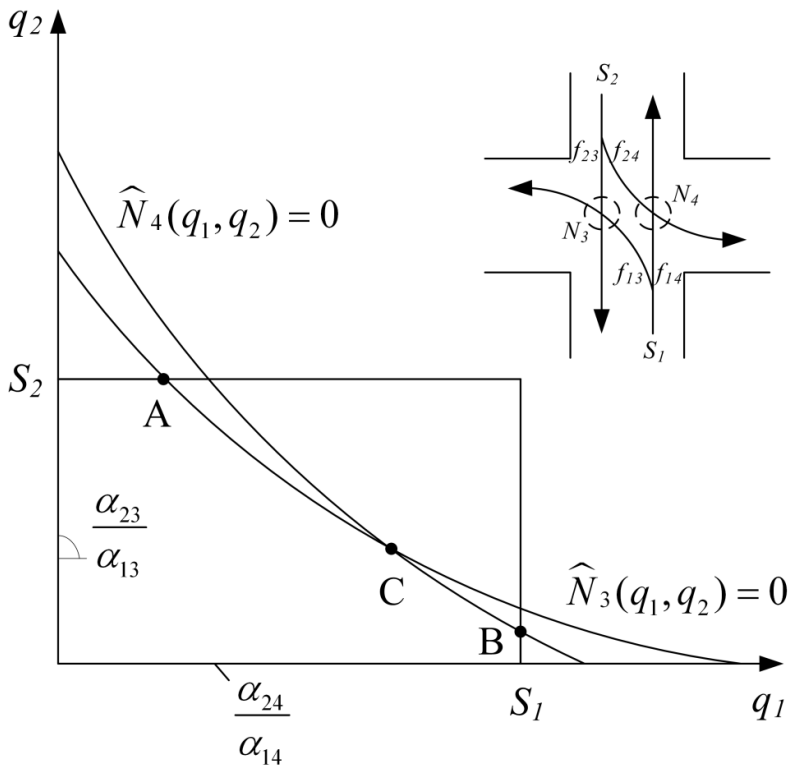

Figure 5-3: Multiple solutions of the intersection model

In this example, the constraints and SCIR definitions allow three solutions A, B, and C. In A, the resulting flow pattern is obtained from the distribution of internal supply N_3, for which q_2 has absolute priority (i.e. the priority ratio for this conflict coincides with the vertical axis). Hence, $q_2 = S_2$, leaving the remaining supply for q_1. Likewise, B results from distributing N_4. Point C also meets the model definitions, with q_1 and q_2 being constrained by $\widehat{N}_3$ and $\widehat{N}_4$ respectively, each leaving the remaining supply for the other flow. These solutions are now phrased in terms of the SCIR definitions:

$$A: \quad \begin{array}{c} q_1 < S_1 \\ q_2 = S_2 \\ U_3 = \{1\} \Leftrightarrow \widehat{N}_3 = 0 \;\&\; \frac{q_1}{q_2} \geq \frac{\alpha_{13}}{\alpha_{23}} \\ U_4 = \varnothing \Leftrightarrow \widehat{N}_4 < 0 \end{array} \tag{5.10}$$

$$B: \quad \begin{array}{c} q_1 = S_1 \\ q_2 < S_2 \\ U_3 = \varnothing \Leftrightarrow \widehat{N}_3 < 0 \\ U_4 = \{2\} \Leftrightarrow \widehat{N}_4 = 0 \;\&\; \frac{q_2}{q_1} \geq \frac{\alpha_{24}}{\alpha_{14}} \end{array} \tag{5.11}$$

$$
\begin{aligned}
& q_1 < S_1 \\
& q_2 < S_2 \\
C: \quad U_3 = \{1\} \Leftrightarrow \widehat{N}_3 = 0 \;\&\; & \frac{q_1}{q_2} \geq \frac{\alpha_{13}}{\alpha_{23}} \\
U_4 = \{2\} \Leftrightarrow \widehat{N}_4 = 0 \;\&\; & \frac{q_2}{q_1} \geq \frac{\alpha_{24}}{\alpha_{14}}
\end{aligned}
\tag{5.12}
$$

In summary, realistic behavioural assumptions (corresponding to the priority rules) have lead in this case to multiple solutions.

5.2.2 Analysis of solution non-uniqueness

We identify the source of the solution non-uniqueness in the intersection model as the fact that inflows q_i are faced with multiple, ambiguous priority ratios in the distribution of different supplies in outgoing links j or internal conflicts k. Revisiting the example in Figure 5-3, one observes that if the priority ratios in the competition for both internal supplies had been identical, only one intersection point with a constraint persists and a unique solution would result. This leads to the following sufficient condition for solution uniqueness, given arbitrary boundary conditions (demands, (internal) supplies and turning fractions):

$$
\begin{aligned}
\exists \alpha_i, \beta_j, \beta_k: \quad & \alpha_{ij} = \alpha_i \beta_j \quad \forall i, j \\
& \alpha_{ik} = \alpha_i \beta_k \quad \forall i, k \\
with: \quad & \alpha_i > 0 \quad \forall i \\
& \beta_j > 0 \quad \forall j \\
& \beta_k > 0 \quad \forall k
\end{aligned}
\tag{5.13}
$$

In words, (5.13) states that the priority parameters α_{ij} and α_{ik} can be written as the product of a single-valued α_i for each i and a factor β for each j and k. This implies that the same priority ratio is used in the distribution of any supply between any two i, thus ruling out ambiguity and non-uniqueness. Hereby, the factors β are scalable, as is the set of α_i's in its entirety. As such, all β could be set to one, so that the priority parameters reduce to single-valued α_i for each i to be applied in the distribution of all internal and external supplies. Specification (5.13) is also a necessary condition for the vast majority of real-world intersection topologies where the flows of at least two incoming links are mutually dependent[23] in at least two common (internal or external) supply constraints (as in the example in Figure 5-3). Proof is given in Appendix E.

[23] This usually requires CTF, since otherwise the q_{ij} are detached and the inflows q_i are no longer tied in the various supplies for which their movements ij compete. If each q_{ij} is then subject to only one supply constraint, it can be determined via its individual α_{ij} and hence the solution is unique. In Appendix E.2, however, an example shows that the q_{ij} may compete for several internal supplies such that a "circular mutual dependency of flows" exists, which can cause non-uniqueness regardless of CTF.

Of course, depending on the boundary conditions (demands, (internal) supplies and turning fractions), ambiguous priority ratios may or may not induce multiple solutions in a specific case. For instance, one supply constraint may dominate and define a unique solution. More specifically, non-unique solutions for two flows q_i and $q_{i'}$ are possible if the boundary conditions are such that a crossing or tangent point $(\bar{q}_i, \bar{q}_{i'})$ exists between (at least) two (internal) supply constraint functions in which q_i and $q_{i'}$ are mutually dependent and this point $(\bar{q}_i, \bar{q}_{i'})$ lies within the feasible domain bounded by the demand constraints and the other supply constraints. This condition (5.14) is written below for an external supply in j and an internal supply in k. The cases of two j or two k, or more than two supply constraints are equivalent.

$$\begin{aligned} &\exists \bar{q}_i, \bar{q}_{i'}, j, k : \hat{R}_j(\bar{q}_i, \bar{q}_{i'}) = \hat{N}_k(\bar{q}_i, \bar{q}_{i'}) = 0 \\ &with: \quad 0 < \bar{q}_i < S_i \\ &\qquad 0 < \bar{q}_{i'} < S_{i'} \\ &\qquad \hat{R}_{j'}(\bar{q}_i, \bar{q}_{i'}) < 0 \quad \forall j' \neq j \\ &\qquad \hat{N}_{k'}(\bar{q}_i, \bar{q}_{i'}) < 0 \quad \forall k' \neq k \end{aligned} \tag{5.14}$$

Condition (5.14) is sufficient but not necessary. It covers most realistic cases, but more far-fetched boundary conditions in which non-uniqueness may exist are conceivable (see Appendix E.2). Graphical clarification can be obtained from revisiting again Figure 3. Therein, point C is the intersection point $(\bar{q}_i, \bar{q}_{i'})$ between the two internal supply constraint functions. When shifting the constraints, multiple solutions remain possible as long as C lies between A and B.

Condition (5.14) specifies circumstances (i.e. model inputs) that may lead to non-uniqueness under a certain combination of α's. Multiple solutions will arise if an intersection point of the supply constraint functions exists as in (5.14) and the corresponding priority ratios point both "above" and "below" this intersection point as in Figure 5-3, i.e.:

$$\frac{\alpha_{ij}}{\alpha_{i'j}} \leq \frac{\bar{q}_i}{\bar{q}_{i'}} < \frac{\alpha_{ik}}{\alpha_{i'k}} \quad or\ vice\ versa \tag{5.15}$$

5.2.3 Solution non-uniqueness: retrospect and corollary

The problem of solution non-uniqueness in the DNL intersection model has now been theoretically discussed. In this section, we first inquire into existing models that are faced with this problem. Then, the observation that the solution non-uniqueness emanates from realistic behavioural assumptions is discussed.

5.2.3.1 Existing models

As argued before, the model framework presented in this chapter encapsulates and extends (the main principles of) most existing DNL intersection models[24]. While solution uniqueness is usually implicitly assumed in the state-of-the-art, it has been shown in the previous section that this is not trivially guaranteed[25]. Multiple-valued priority parameters that dictate the distribution of (internal) supplies through different priority ratios may lead to multiple solutions. It must be stressed that – although we have introduced multiple-valued priority parameters for internal supplies - also models that only consider external supply constraints are prone to non-uniqueness if they implement multiple-valued priority parameters α_{ij}. Most existing models implement single-valued priority parameters α_i and hence have unique solutions. For some models, however, this is not the case.

Firstly, some models intend to assume single-valued α_i, but fail to acknowledge the explicit role of the turning fractions in the distribution of supply. In such models, the competitive strength of a link *i* in the distribution of any supply is given by α_i, whereas our model framework assumes $f_{ij}\alpha_i$ (or $f_{ik}\alpha_i$). For non-CTF models that make this assumption, such as Ni et al. (2006), Taale (2008) and some models in Adamo et al. (1999), this has no impact on the solution uniqueness. For CTF models, however, not including f_{ij} in the competitive strength ironically leads to multiple-valued priority ratios and thus solution non-uniqueness. Indeed, (4.11) then translates to:

$$\frac{q_i}{q_{i'}} = \frac{\alpha_i}{\alpha_{i'}} \frac{f_{i'j}}{f_{ij}} \quad \forall i, i' \in U_j \tag{5.16}$$

This is the case for the model of Gentile et al. (2007).

Secondly, a few models assume multiple-valued α_{ij}, namely that of Gentile (2010) and one model in Adamo et al. (1999)[26]. The model of Gentile (2010) - and Gentile et al. (2007) - may exhibit multiple solutions as in the example in Section 5.2.1. For completeness, it is noted that the multiple-valued α_{ij} in the model in Adamo et al. (1999) lead to one inconsistent solution rather than multiple solutions. This is because this model does not redistribute excess supply. In such an inconsistent solution, for instance, two flows could be constrained by their own most restrictive supply constraint (not the same one for the two),

[24] Note, firstly, that considering internal conflicts in the form (5.3) is equivalent to assuming absolute priority in formulation (5.4). Secondly, the proposed model framework is general in the sense that it extends the universal approach of distributing external supply based on a proportionality rule to internal supply. The model of Gibb (2011) is – to our knowledge – the only realistic exception that does not distribute external supply according to such a proportionality rule. However, we did not see good opportunity to extend his model to internal conflicts (at least not with driver behaviour other than turn-taking).

[25] It should be noted that the non-uniqueness is not solved by releasing any of the seven requirements that we adhere in our model definition. For most violations, such as to the invariance principle or flow maximization, this is straight-forward to see. The CTF requirement is a factor in most cases of non-uniqueness, but not always (see Appendix E.2).

[26] Also, the model of Flötteröd & Nagel (2005) first considers α_{ij}, but ultimately resorts to the computation of one representative α_i.

while neither of the supplies would be fully utilized since the competitor is more constrained by the other supply.

Before the thorough analysis presented here was published (Corthout et al., 2012), solution non-uniqueness related to internal supply constraints had already been observed by Flötteröd & Rohde (2011). Furthermore, it also appears in the model of Yperman (2007). The solution non-uniqueness in these models arises as explained in Section 5.2.1. Hence, the observed non-uniqueness is not confined to the representation of internal supply constraints as in (5.4). Instead, our findings regarding the (conditions for) uniqueness apply to many existing models. Moreover, of the existing models that include internal supply constraints (see Section 5.1.2) these are the ones making the most realistic behavioural assumptions. Therefore, also in the development of future models, this issue requires careful consideration. This leads to the discussion on the connection between the non-uniqueness observed in the model and real-world traffic rules and behaviour in the next subsection.

5.2.3.2 The connection between non-uniqueness and reality

Most intersection models limited to external constraints assume single-valued α_i, ensuring uniqueness. However, solution non-uniqueness becomes a much more prominent problem when models are to be extended with internal supply constraints. To ensure solution uniqueness, uniform priority ratios are needed by adding condition (5.13) to the model definition. However, while other modelling assumptions are derived on a behavioural basis, (5.13) is not – its origin is to enforce uniqueness. Moreover, this technical condition (5.13) appears behaviourally unrealistic when introducing internal supply constraints. Indeed, it is (often) in contradiction with the priority rules, and thus with how one would naturally define the priority parameters to govern the distribution of internal supplies (see the example in Section 5.2.1). Blindly imposing single-valued priority ratios without any consideration of the ambiguity that seems inherent to reality is thus not advisable. Only if the expected driver behaviour is limited to turn-taking behaviour, as in AWSC intersections, reverting to this approach seems natural, e.g. with $\alpha_i = C_i$.

Preferably, the decision of how to treat the non-uniqueness in the model should be supported by empirical research. The empirical studies on intersection flows conducted in the past do not provide sufficient support since their focus is typically on various aspects of gap acceptance behaviour (e.g. gap distributions) and not on validating intersection models in DNL - except for simple merges; see e.g. Bar-Gera & Ahn (2010) and Ni & Leonard (2005). Flow non-uniqueness has never been reported in empirical data, but it has also never been looked for. Condition (5.14) identifies the rather specific circumstances in which the solution in the model may be non-unique. Thus, for empirical validation, these circumstances are to be sought in the field.
Firstly, such empirical research should identify whether or not real intersections indeed exhibit non-unique solutions, i.e. systematically different flow patterns under the same

boundary conditions[27]. On a microscopic scale, at the level of individual and small groups of vehicles, observations will inherently have a random nature. Very different situations can surely be identified on such a disaggregated level. To allow comparison to the result of a macroscopic DNL intersection model, which does not exhibit such randomness, the empirical data has to be analysed on an aggregated level (using a discretization of at least a few minutes).
Secondly, if multiple solutions can also be observed at this aggregated level, their properties are to be revealed. Some situations may be more stable than others. The probability of different flow patterns and the required perturbation to initiate a transition need to be studied. Most likely, this will be rooted in the microscopic level, since the decision of an individual driver might induce a transition. Other factors – some of which external to the intersection under observation – are expected to further increase the complexity of this analysis, e.g. the history of flows, the intersection geometry, neighbouring intersections, bus stops, etc.
However, it is also plausible that in reality, the inherently random and non-unique microscopic situations are aggregated into a mixture that can reasonably be considered unique[28]. If this is the case, then further empirical analysis of the (external) factors that enforce this unique flow pattern is very useful as well. This would allow developing stronger modelling guidelines to establish this unique flow pattern in the intersection model as well.

In conclusion, although we find that the solution non-uniqueness in the model emanates from realistic behavioural assumptions, it remains to be seen whether or not flow ambiguity can indeed be identified in reality. If it can, this may reinforce the need to further develop stochastic DNL models, which are currently still in a relatively early stage of development (e.g. Sumalee et al., 2011; Osorio et al., 2011). In stochastic DNL models, non-unique intersection flows could be resolved probabilistically. Just as a stochastic approach to the DTA problem yields a unique solution in distributional terms (e.g. Flötteröd & al., 2011) a stochastic DNL could replace non-unique intersection flows by a unique distribution.
Alternatively, perhaps (deterministic) chaotic models could be developed to describe traffic flows over intersections. Indeed, it seems plausible that an intersection can be modelled as a chaotic dynamical system, which is highly sensitive to the (history of) boundary conditions. Small details in the demands, supplies, intersection geometry, nearby traffic signals, bus stops, etc., would steer the solution(s) in such a chaotic model (likely including bifurcations).
Currently, however, state-of-the-art deterministic DNL models cannot deal with multiple solutions of the intersection model. To enable the further use of these traditional deterministic macroscopic DNL models, the solution non-uniqueness of the intersection model must therefore be resolved. A unique solution that is established in a well-considered, systematic and reproducible way is by far preferable over having the solution algorithm randomly or unknowingly come up with one solution. Therefore, we propose in the following section some possible approaches to remedy the non-uniqueness. Further (empirical) research as discussed above should determine whether these contributions have lasting value (if an intersection model producing a unique solution is shown to be sufficiently realistic for at least

[27] Microscopic simulation models cannot be used for this analysis instead of real data. Firstly, microscopic models are highly steerable through parameter calibration. Secondly, most existing microscopic simulation models do not realistically represent the vehicle interactions at intersections; especially not in congested conditions (see Chevallier & Leclercq, 2009).
[28] Of course, all observations will show some stochastic deviations.

some applications) or only intermediate value (if stochastic or chaotic models at some point would be to fully replace the traditional models).

5.2.4 Pragmatic approaches to establish a unique solution

In traditional deterministic DNL modelling, a transformation of the non-unique solutions of the intersection model into one prevailing flow pattern is needed, which is not straightforward and prone to subjective decisions of the modeller. It may be desirable to at first allow ambiguous priority parameters in the model definition, and then to alleviate the non-uniqueness by some kind of pre- or post-processing (which must be unambiguous given the model inputs). In the following, two different approaches are presented which we conceive to be possible and behaviourally plausible remedies for the observed solution non-uniqueness. We distinguish two types of approaches: (a) pre-processing the priority parameters so that the model produces a unique solution; and (b) computing non-unique solutions that result from ambiguous priority parameters and then post-processing these into one solution. In Chapter 6, we develop specific intersection models by further detailing the first approach.

1. Pre-processing of the priority parameters:
   For every i with multiple-valued α_{ij} and α_{ik}, a representative α_i is computed through (5.17). To allow a sensible weighting, both the priorities α_{ij} and α_{ik} and the weights w_{ij} and w_{ik} must be normalized.

$$\begin{aligned} &\alpha_i = \sum_j w_{ij}\alpha_{ij} + \sum_k w_{ik}\alpha_{ik} \quad \forall i \\ &with: \quad \alpha_{ij} > 0 \;\&\; \alpha_{ik} > 0 \quad \forall i,j,k \\ &\sum_i \alpha_{ij} = 1 \quad \forall j \\ &\sum_i \alpha_{ik} = 1 \quad \forall k \\ &w_{ij} \geq 0 \;\&\; w_{ik} \geq 0 \quad \forall i,j,k \\ &\sum_j w_{ij} + \sum_k w_{ik} = 1 \quad \forall i \end{aligned} \tag{5.17}$$

   The model of Flötteröd & Nagel (2005) – without internal supply constraints - resorts to this approach (with $w_{ij} = f_{ij}$).
   Different choices of the weights w_{ij} and w_{ik} define different pre-processing strategies:

   - If no further information is available, uniform weights can be chosen. Other averaging schemes are thinkable but require justification based on supplementary modelling assumptions. These assumptions could stem from one of the following considerations:

     - The weights may be assumed dependent on the history of flows. However, the analysis of Section 5.3 indicates that this may lead to undesirable results.

- The amount of competition that is present for each supply may be determinant for the weight of the corresponding priority parameter. That way, a supply that is more overloaded has a stronger impact on all flows. However, this competition or overloading cannot be directly determined from the demands S_i as this would not be compliant with the invariance principle. Rather, this ranking should be based on capacities. In Chapter 6, this approach is motivated and further elaborated.

- One representative α_{ij} (or α_{ik}) could be selected by setting the respective w_{ij} (w_{ik}) to one and all other weights to zero. Similar considerations as described above could motivate this approach. For instance, a ranking could be defined among all (internal) supply constraints so that the highest ranked active constraint determines the solution. This ranking could be made based on the history of flows, or the competition for each supply.

This pre-processing approach comes down to indirectly imposing condition (5.13) in the model, while (to some extend) accounting for driver behaviour. An advantage of this approach is that it is straightforward to implement and computationally efficient.

2. Post-processing of the flows:
Every possible solution resulting from the multiple, ambiguous priority ratios is computed, leading to separate flow patterns $\boldsymbol{q_r}$. These are then averaged into a unique resulting flow pattern $\boldsymbol{q}$ according to:

$$\begin{aligned} \boldsymbol{q} &= \sum_r w_r \boldsymbol{q_r} \quad \forall r \\ with: \quad w_r &\geq 0 \quad \forall r \\ \sum_r w_r &= 1 \end{aligned} \tag{5.18}$$

Again, different choices of the weights define different solution strategies:

- One could assume a uniform average solution to hold. This is in line with the assumption that a deterministic DNL represents average network conditions. A naïve averaging, however, raises considerable difficulties. The average flow pattern may not comply with the requirement of individual flow maximization, or it may violate the invariance principle or some non-linear internal supply constraint.

- One could also select a single solution as the most plausible or representative one for the given situation. (This again corresponds to setting the respective weight w_r to one and all others to zero). Again, this could be based on previous flows (assuming for instance that the smallest change in flow over time is the most plausible). Technically, this corresponds to selecting the priority

parameters that have led to this solution, but the selection criterion is now based on flows, which are not known a priori.

Although this flow post-processing approach makes sense intuitively, an important drawback is that an averaging of solutions can lead to flows that are inconsistent with the basic modelling assumptions. Also, multiple candidate solutions need to be evaluated, rendering it computationally intensive.

Determining with certainty which of the above approaches (or any other alternative) is most realistic or most appropriate is difficult or even impossible without further theoretical and empirical research. Moreover, the answer might differ under varying circumstances.

5.3 Spatial intersection modelling

Until now, we have exclusively considered DNL intersection models in the point-like, dimensionless form based on the arguments given in Section 2.2.1. In response to the suggestion made in Flötteröd and Rohde (2011) that the observed solution non-uniqueness could be resolved by reverting from a point-like to a spatial modelling approach, an analysis of a spatial intersection model is provided in this section. This analysis shows, however, that spatial models do not satisfactorily resolve the problem of non-uniqueness.

Recall that modelling intersections in DNL spatially implies that the intersections are essentially mini-networks, in which the conflict zones of crossing and merging flows are represented by dummy nodes, connected by dummy links. Figure 5-4 shows a spatial version of the example in Section 5.2.1, in which very short dummy links connect the dummy nodes[29]. The first dummy nodes at the end of links 1 and 2 are merely diverge points. The other two dummy nodes contain the internal crossing conflicts (the same as in Figure 5-3).
As before, the straight movements have absolute priority over the left-turning movement of the other incoming link and the left-turning movements do not hinder each other. For simplicity's sake, the internal supply constraint functions are modelled identically to external supply constraint functions, as a maximum flow (veh/h) that can cross the conflict zone:

$$\widehat{N}_k(q_1, q_2) = f_{1k} q_1 + f_{2k} q_2 - N_k \leq 0 \qquad k = 3,4 \tag{5.19}$$

[29] In fact, the experiments were conducted with a simpler configuration, in which the two straight dummy links connecting the diverges and the crossing dummy nodes are not present. (One could also consider these links to have zero length.) This way, only two dummy nodes are needed that govern both the diverging and the crossing actions. The graphical representation was modified to that of Figure 5-4, since this is easier to understand. Of course, the results and conclusions are entirely equivalent.

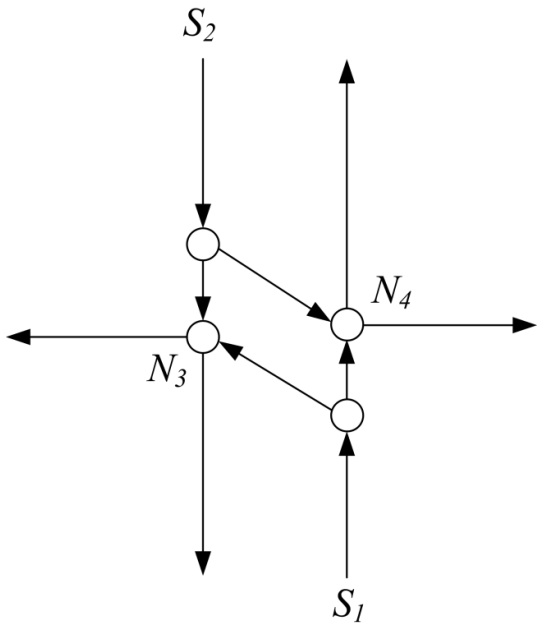

Figure 5-4: Spatial model of the example in Section 5.2.1

Consequently, it is possible to simulate this spatial intersection model with any state-of-the-art DNL model. Here, as in the rest of this thesis, LTM is chosen, with a minor modification to model the absolute compliance behaviour in the internal conflicts. For this simple example, this modification requires nothing more than granting the prioritized movement its maximal share (= S_{ik}) and passing the remaining internal supply N_k - S_{ik} to the minor movement. Apart from that, this modification does not have any implication for the usual working of the DNL model. For the simulation of this example, the conflict zone supplies are set to N_k = 1000 veh/h.

Firstly, two scenarios are considered in which the boundary conditions external to the intersection (i.e. the turning fractions and the demands) are identical at the end of the simulation, but their histories are not (see Figure 5-5 and Figure 5-6).

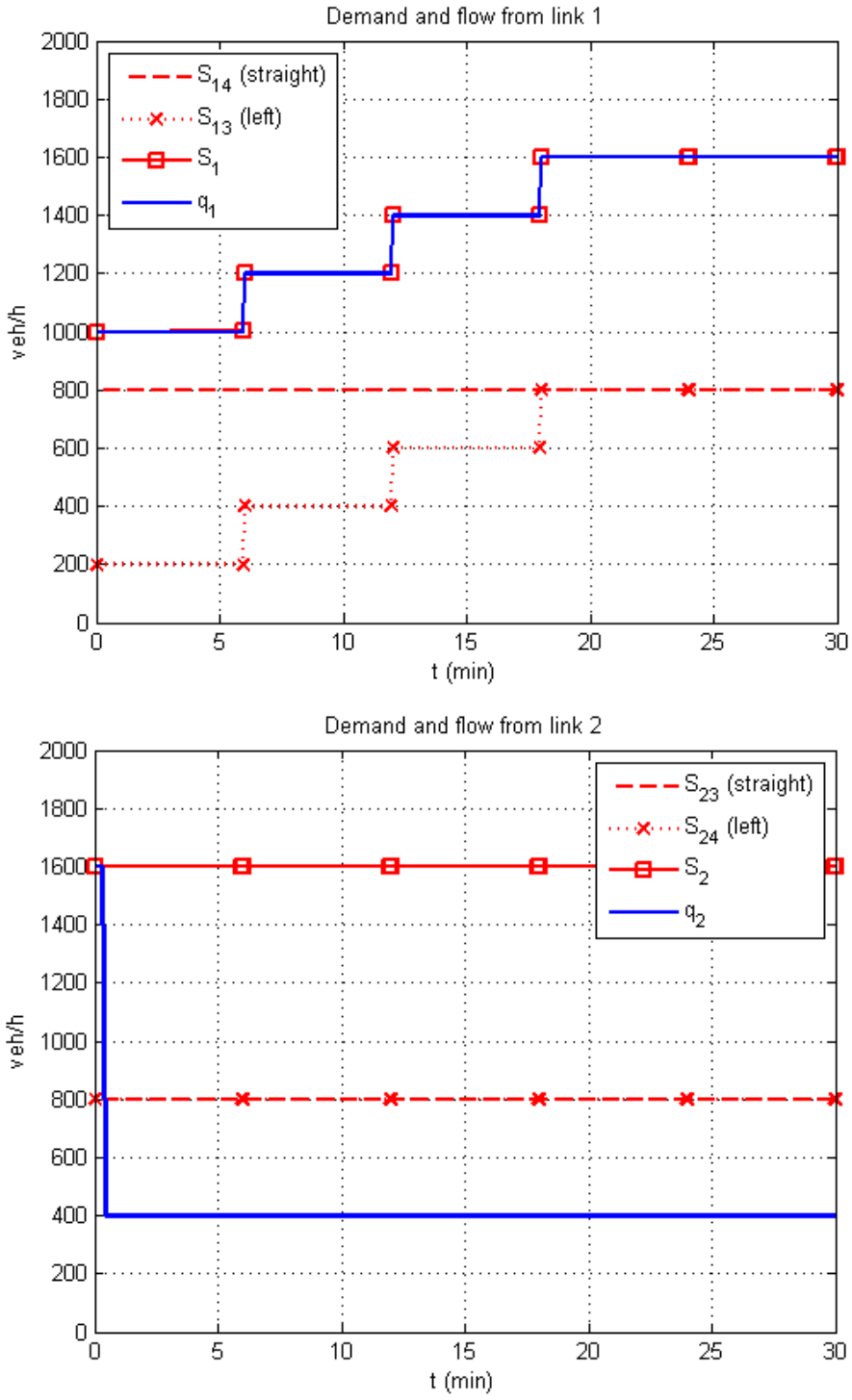

Figure 5-5: Scenario 1: S_{13} increases gradually

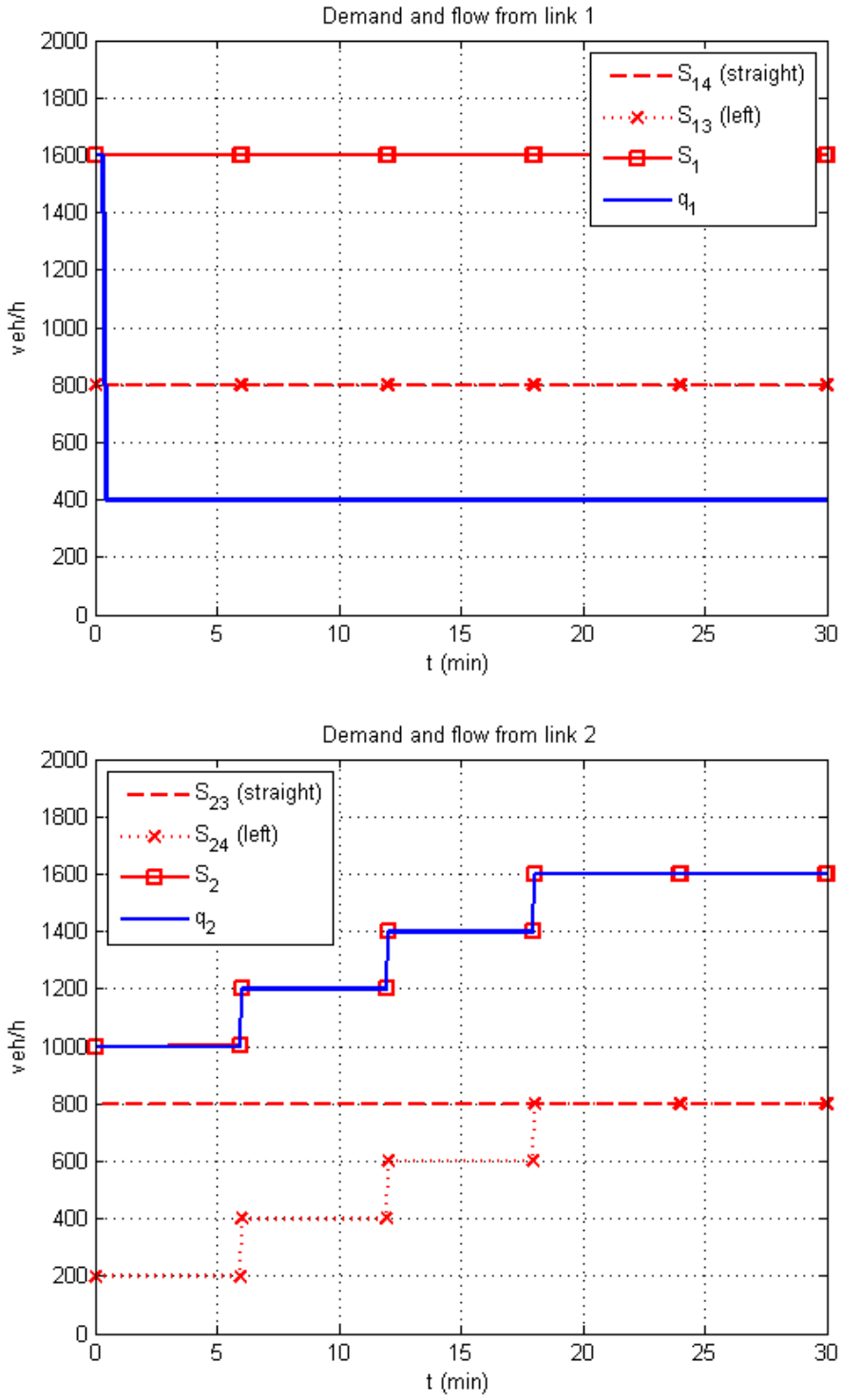

Figure 5-6: Scenario 2: S_{24} increases gradually

The straight demands are S_{14} (= $f_{14}S_1$) and S_{23} (= $f_{23}S_2$) from incoming links 1 and 2 respectively. The left-turning demands are S_{13} and S_{24} respectively. In scenario 1 and 2, the left-turning demands are initially different. During the simulation, the low left-turning demand is gradually increased until it reaches the same level as the high left-turning demand from the opposite link. In the initial phase of scenario 1, S_{13} is low, allowing link 1 to send its full demand, since there is enough remaining supply ($N_3 - S_{23}$ = 200 veh/h) for its minor left-turning movement 13. Meanwhile, the left-turning movement 24 is obstructed by $\widehat{N}_4$, so that a queue forms on the dummy link that spills back onto link 2 very quickly. This renders link 2 congested; also the straight movement 23 is thus held back. While S_{13} is gradually increased, this does not change the flows, as link 2 is now unable to claim its maximal share of N_3 (leaving $N_3 - q_{23}$ = 800 veh/h for link 1) due to the activated constraint $\widehat{N}_4$. As a result, q_1 stays dominant throughout the simulation, while q_2 remains constrained. Scenario 2 is the exact opposite.

Thus, although the boundary conditions at the end of the simulation are identical in both scenarios, the resulting flow patterns are very different. This demonstrates that also in a spatial model, solution uniqueness is not guaranteed given only the instantaneous boundary conditions. The spatial model inherently determines the solution on the basis of history. While dependency of the solution on history is not unrealistic as such, Figure 5-5 shows that in scenario 1, only a few seconds during which $S_{13} < S_{24}$ suffice for q_1 to take the upper hand and hold it for the entire simulation. (Figure 5-6 of scenario 2 is the opposite.) Clearly, this cannot be considered a realistic dependency on history. Indeed, a time period equal to the spillback time over the left-turning dummy link is enough to block the straight flow. Since a DNL simulation starts from an empty network, this means a spatial intersection model could steer the results solely based on which flows happen to reach an intersection first in the beginning of the simulation.

Finally, a nearly symmetric demand pattern is chosen for scenario 3 (Figure 5-7). This results in an oscillating flow pattern, caused by back and forth propagating waves on the very short dummy links, which stabilizes very slowly. In the converged solution, q_2 dominates q_1 thanks to a slightly smaller left-turning fraction. The frequency of the oscillations and the speed of convergence depend not only on the boundary conditions. Also the exact spatial representation of the intersection (for instance the length of the dummy links) determines the spillback dynamics. This type of oscillations clearly is an artefact and not an interpretable model output.

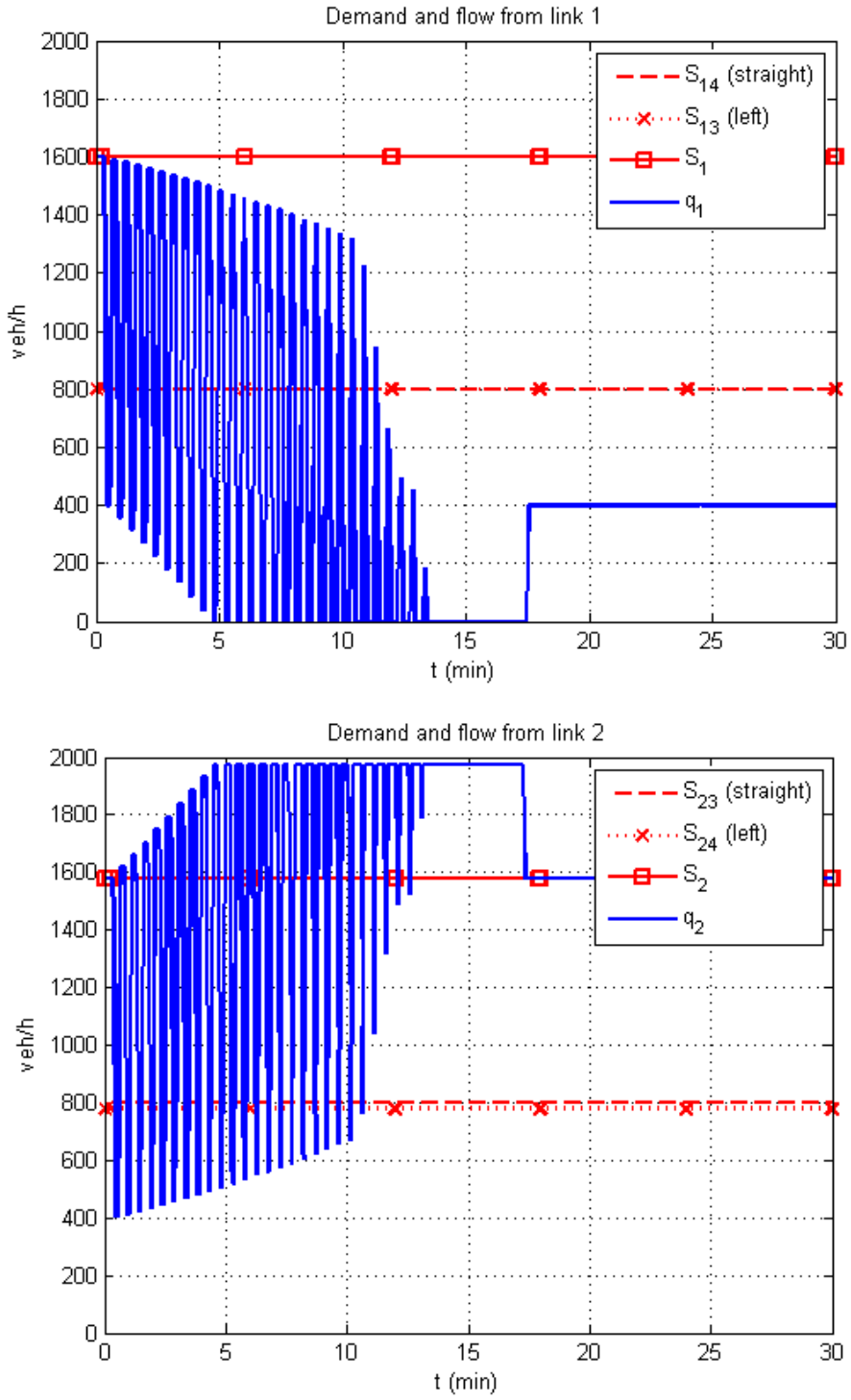

Figure 5-7: Scenario 3: nearly symmetric demands engender unstable flows

In summary, the current problems of point-like intersection models cannot be bypassed in a spatial way. While at first sight a spatial model physically detaches the mutual dependencies that lie at the origin of the ambiguity in point-like models, the wave propagation over the dummy links rejoins these dependencies. This rejoined ambiguity may result in an oscillating flow pattern, but also in an irreversible dependency on a history of only a few seconds. Both effects are caused by spillbacks internal to the spatial intersection model and therefore the result of the (unjustified) application of macroscopic propagation theory designed for long links on very short dummy links that fit at most a few vehicles. This model behaviour depends on the spatial representation (e.g. the dummy links' length and characteristics) and is clearly unrealistic and undesirable. Because of this, and due to the three disadvantages of spatial modelling discussed in Chapter 2 (namely simultaneous driver decisions, model-induced gridlock and the possible computational consequences of the CFL condition), we see better potential in continuing research to further develop traditional point-like models rather than abandoning them in favour of spatial models. On a side note, this also means it is not advisable to blindly adopt networks for DNL or DTA purposes directly from static models, in which – in our experience - complex intersections are often modelled spatially.

5.4 Conclusion

This chapter discusses the inclusion of internal supply constraints, arising from conflicts within the intersection itself, in DNL intersection models. It is shown that this may lead to multiple solutions of the intersection model. This solution non-uniqueness is thoroughly analyzed and possible approaches to obtain a unique solution are suggested. As such, this discussion is conceptually similar to Carey (1992), who shows that the FIFO behaviour of traffic can lead to non-convexity in DTA and also proposes how to practically deal with this problem. Moreover, our findings contribute to the analysis of solution non-uniqueness in the DTA problem (see also, e.g., Daganzo, 1998).
Building on the previous chapters - and Flötteröd & Rohde (2011) -, the main contributions of this chapter are:

- the general definition of internal supply distribution through proportionality of priority parameters (analogous to external supply), which can be varied to model different driver behaviour
- the identification of (the cause of) solution non-uniqueness within this model framework, which is found to result from realistic behavioural assumptions
- the formulation of a sufficient (and necessary) condition for the priority parameters that ensures solution uniqueness, which, however, conflicts with a realistic representation of driver behaviour
- the description of the boundary conditions (input to the intersection model) that may induce multiple solutions
- the suggestion of pragmatic approaches to remedy the non-uniqueness in deterministic point-like intersection models; one of which will be detailed in the following chapter
- the elaboration on the spatial modelling approach, revealing unrealistic behaviour of such models

Since the solution non-uniqueness is argued to follow from realistic behavioural assumptions, it poses a severe challenge for modellers. This chapter provides technical conditions to identify this problem to the research community and practical guidelines on how to establish a unique solution, which is necessary in deterministic DNL models. To support the decision of how to deal with the solution non-uniqueness in the model, foremost empirical studies would be valuable. This chapter aids to that by helping to understand the phenomenon (and when it occurs) in the model, so that these specific circumstances can be sought in the field. If the existence of multiple flow patterns under identical boundary conditions is indeed observed, their characteristics – e.g. probability, frequency of switches, duration of stable periods - and the (external) factors that lead to these characteristics – e.g. history, neighbouring (signalized) intersections, intersection geometry – should be identified. Capturing the non-uniqueness would require a paradigm shift towards stochastic or chaotic intersection models. If on the other hand this empirical work disproves non-uniqueness in reality, it can hopefully support the development of more stringent modelling guidelines to correctly identify the unique solution.

In conclusion, we see better potential in further developing traditional point-like models rather than abandoning them in favour of spatial models. In the next chapter, intersection models for different types of intersections are formulated. For this, two specifications are required to the general model presented here. Firstly, a definition for the internal supply constraint functions is suggested based on conflict theory after Brilon & Wu (2001). Secondly, the suggested approach of pre-processing multiple-valued priority parameters into a single α_i is further detailed. This way, solution uniqueness is ensured while accounting for different driver behaviour in different (internal) conflicts.

6
SPECIFIC INTERSECTION MODELS

This chapter presents specific intersection models for all types of intersections discussed in Section 2.2.2. These models specify the general model framework of Chapter 5. Firstly, the internal supply constraint functions are defined. Conflict theory (Brilon & Wu, 2001) is chosen as a basis, which is more general and flexible in a DNL environment than gap acceptance theory. Secondly, solution uniqueness is ensured by following the first approach suggested in Section 5.2.4. Separate priority parameters are specified for each conflict, according to the expected driver behaviour. These are then weighted into single-valued priorities to meet condition (5.13).
The advantage of this approach is that it renders the intersection models sensitive to internal conflicts – which are usually ignored in state-of-the-art models – while ensuring solution uniqueness and compliance with the requirements in Section 3.4.1. However, ensuring these consistencies limits the range of action of the modeller. The models presented here are fully consistent and unambiguous but (partly in result) still simplified. The development of consistent but in some ways basic models is considered to be the most logical next step in the process towards developing macroscopic DNL intersection models that are both well-defined and realistic.

The above discussion is continued in Section 6.5. The preceding sections are dedicated to the specification of the intersection models. Section 6.1 defines the internal supply constraint functions, the priority parameters per conflict and the weighting procedure into unambiguous priorities. These ingredients are used in Section 6.2 to present different types of intersection models. Additional modelling specifications for signalized intersections are included. In Section 6.3, a numerical example is presented to demonstrate the functionalities of the presented modelling approach. Finally, in Section 6.4, some practical guidelines are given for the use and future empirical validation of the proposed models.

6.1 Further model specifications

Summarizing the previous chapters, (6.1)-(6.4) defines the general intersection model framework that complies with all requirements of 3.4.1 and ensures a unique solution. In this model framework and in all its specific instances presented in Section 6.2, each flow q_i is determined by its most restrictive demand or (internal) supply constraint.

- Demand constraints:

$$q_i \le S_i \quad \forall i \tag{6.1}$$

- External supply constraints:

$$\widehat{R}_j(\mathbf{q}) = \sum_i f_{ij} q_i - R_j \le 0 \quad \forall j \tag{6.2}$$

- Internal supply constraints:

$$\widehat{N}_k(\mathbf{q}) \le 0 \quad \forall k \tag{6.3}$$

- SCIR definitions:

$$\begin{gathered}
U_j = \{i \mid i \mapsto \widehat{R}_j\} \quad \forall j \\
U_k = \{i \mid i \mapsto \widehat{N}_k\} \quad \forall k \\
q_i < S_i \Leftrightarrow \exists j \mid i \in U_j \text{ or } \exists k \mid i \in U_k \quad \forall i \\
U_j = \varnothing \Leftrightarrow \widehat{R}_j < 0 \quad \forall j \\
U_j \neq \varnothing \Leftrightarrow \widehat{R}_j = 0 \quad \forall j \\
U_k = \varnothing \Leftrightarrow \widehat{N}_k < 0 \quad \forall k \\
U_k \neq \varnothing \Leftrightarrow \widehat{N}_k = 0 \quad \forall k \\
\frac{q_i}{q_{i'}} \ge \frac{\alpha_i}{\alpha_{i'}} \quad \forall i,i' \mid i \in U_j \;\&\; f_{i'j} > 0 \;\&\; \forall i,i' \mid i \in U_k \;\&\; f_{i'k} > 0 \\
\frac{q_i}{q_{i'}} = \frac{\alpha_i}{\alpha_{i'}} \quad \forall i,i' \in U_j \;\&\; \forall i,i' \in U_k
\end{gathered} \tag{6.4}$$

To obtain clearly defined, specific intersection models, two components of (6.1)-(6.4) need further detailing. Firstly, definitions of the internal supply constraint functions $\widehat{N}_k$ are needed. Secondly, the derivation of the priority parameters α_i is to be specified.

6.1.1 Internal supply constraint functions derived from conflict theory

This section motivates and explains the use of conflict theory as the basis for defining internal supply constraint functions in the DNL intersection model. Here, internal constraints resulting from crossing conflicts and merging conflicts on the arcs of a roundabout are considered. Internal constraints due to traffic lights will be discussed later, when the signalized intersection model is outlined, as these constraints obviously only apply to this type of intersection.

6.1.1.1 Motivation for using conflict theory

Most formulations of intersection supply constraint functions in DNL intersection models in the state-of-the-art stem from gap acceptance theory. However, we find the newer and less known conflict theory (see Wu, 2000; Brilon & Wu, 2001 and Brilon & Miltner, 2005) more promising for implementation in DNL intersection modelling for the following reasons.

- Gap acceptance theory does not carry over to situations in which the prioritized streams are congested. Indeed, the assumption that vehicles from the minor streets accept or reject gaps that are supposed to exist in the prioritized streams – for which a distribution is assumed – does not correspond well to reality in such cases. To our knowledge, modifications of gap acceptance theory for congested prioritized streams do not exist. While also conflict theory has not been validated in congested conditions, at least the underlying assumption – i.e. conflict points having a maximum occupancy of 100 % - is intuitively consistent with free flowing as well as congested conditions.

- Most gap acceptance procedures are designed for situations with only one prioritized and one minor movement. It is stated explicitly in Gartner et al. (2000) that no rigorous analytical solution is known for the capacity of the lower ranked movements at intersections with more than two ranks (which is commonly the case for 4x4 intersections). In fact, Gartner et al. (2000) suggest approximate formulations for these capacities that resemble conflict theory.

- Contrary to gap acceptance theory, an extension to include conflicts with non-motorized traffic (pedestrians and cyclists) is rather straight-forward in conflict theory. In an urban environment, this may be an important advantage.

- Despite some modifications that exist to accommodate cases of limited priority (e.g. Troutbeck & Kako, 1999 and Chevallier & Leclercq, 2007), the underlying assumption of gap acceptance theory is that of a prioritized movement having (absolute) priority over a minor movement. Formulations of conflict theory on the other hand, are more easily convertible to other behavioural assumptions such as turn-taking.

- Last but not least, conflict theory has the advantage over gap acceptance theory of being less complicated. This is even more so in a DNL context, since it reasons directly from the flows instead of deriving flows via (distributions of) gaps.

In summary, conflict theory appears more general and flexible than gap acceptance theory regarding driver behaviour, multiple conflicts and saturation levels. In addition, it is simpler. Empirical validation – in uncongested conditions - of conflict theory is presented in Wu (2000) for AWSC intersections and Brilon & Miltner (2005) for priority-controlled intersections. The latter also shows a reasonable agreement with a gap acceptance approach. In the following, conflict theory and how to it translates to internal supply constraint functions is explained.

6.1.1.2 Conflict zones

Conflict theory is based on the addition-conflict-flow procedure of Gleue (1972). It is first proposed in Wu (2000) for AWSC intersections. Brilon & Wu (2001) use it for priority-controlled intersections. Brilon & Miltner (2005) propose an extension with an exponential factor adopted from gap acceptance theory. Since the assumption of exponentially distributed gaps is only realistic for very low traffic volumes (see e.g. Heidemann & Wegmann, 1997), this addition is omitted here. Future research must identify the possibilities for introducing more realistic extensions to account for the effect of random arrivals and gap distributions, by means of some factor with diminishing effect as the saturation level rises. In absence of such a factor, the internal supply constraint functions as defined in the next section correspond best to bunched arrivals and congested conditions.

Conflict theory groups the conflicts at an intersection into conflict groups or zones. These conflict zones represent a part of the intersection infrastructure that is to be shared by the participating movements. A zone can only be occupied by one vehicle at a time, so the movements have to make alternate use of it. For roundabouts, the conflict zones are obviously situated at the merging point on the arcs of the roundabout. For standard intersections, it may be more difficult to determine the location of the conflict zones and which movements to include in which zone. The intersection design - the size of the intersection, possible storage space in the middle, the configuration of turning lanes, etc. - plays an important role. This, however, falls outside the scope of this thesis. Additional information can be found in Wu (2000), Brilon & Wu (2001) and Brilon & Miltner (2005). Still, further research complementing these studies on this matter would be useful. Figure 6-1 provides an example. This configuration is suggested by Brilon & Wu (2001) for the motorized conflicts at a 4x4 intersection (with separate turning lanes for all movements). In the outer zones, one recognizes the external, merging conflicts. We will continue to model these conflicts as external constraints, as in the previous chapters. The inner, crossing conflicts are to be translated in internal supply constraint functions.

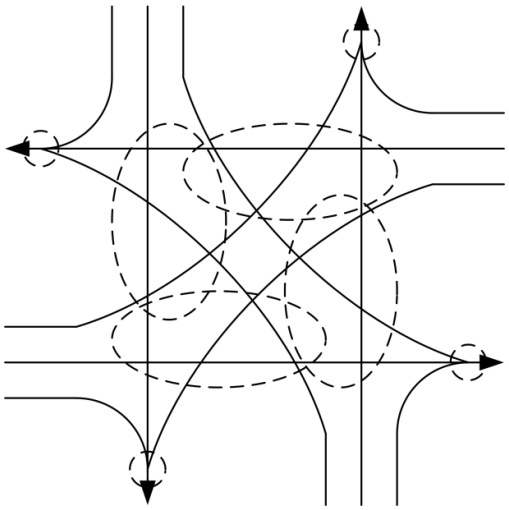

Figure 6-1: Conflict zones at a 4x4 intersection (Brilon & Wu, 2001)

6.1.1.3 Definition of the internal supply constraint functions

Conflict theory assumes a consumption time t_{ik} (in seconds) during which one vehicle originating from i occupies conflict zone k. The total available time for all movements is 3600 s/h[30]. This is thus the internal supply N_k in (6.5). Each partial flow $f_{ik}q_i$ occupies k for a total time of $f_{ik}q_it_{ik}$, which constitutes the consumption of the internal supply N_k.

The possibility of queues blocking these internal conflicts zones is not considered. The effect of traffic lights will be accounted for in Section 6.2.5. In conflict theory, a conflict determines only the flow of the minor movements, while the prioritized flow is considered given. A slight reformulation yields the following internal supply constraint functions:

$$\begin{aligned} &\widehat{N}_k(\mathbf{q}) = \sum_i f_{ik} q_i t_{ik} - N_k \le 0 \quad \forall k \\ &with: \quad N_k = 3600 \text{ s/ } h \end{aligned} \tag{6.5}$$

where the vector $\mathbf{q}$ is composed of all $q_i|f_{ik} > 0$. Calibrations of the consumption times t_{ik} are included in Wu (2000), Brilon & Wu (2001) and Brilon & Miltner (2005). In absence of the appropriate empirical data and/or time for calibration (which is very likely), it makes sense for DNL applications to assume that t_{ik} can be derived from the capacity of a (turning) lane $C^{L,ik}$ from which the movement ik originates:

$$t_{ik} = \frac{3600}{C^{L,ik}} \tag{6.6}$$

In formulation (6.5), no priority is expressed of some movements over others. The priority rules and their effect on the internal supply distribution are entirely governed by the priority

[30] We limit the discussion to the standard case in which one vehicle at a time may occupy the conflict zone. Constraints for zones that accommodate more vehicles at a time could be defined analogously.

parameters, which will be defined in the next section. Finally, note that internal supply constraint functions in the form of (6.5) imply that the conflict zones are considered as independent constraints, and that only the most restrictive of all constraints ultimately takes effect on a flow. This is in line with the assumptions throughout the preceding chapters. However, this also implies that the dependencies that exist at the level of a single driver are not considered. Simultaneous driver decisions when passing through multiple conflict zones may cause this assumption to underestimate the restriction on minor flows (i.e. higher flows than in reality). We refer to the discussion in Section 6.5 on the reasons, consequences and future intentions regarding the neglect of the influence of this microscopic level on the macroscopic solution of flows.

6.1.2 Priority parameters

As mentioned in Section 5.1.2.2, one could consider defining priority functions $\alpha(\mathbf{q})$ - rather than constant parameters - that depend on the flows to take into account (at least to some extent) the effect of saturation levels on driver behaviour. However, additional research is needed before an increase in realism of the model can be expected that would justify the accompanying increase in complexity. The same statement can be made regarding the definition of priorities that depend on the history of flows rather than the instantaneous flows. Moreover, this introduces dynamics over which the modeller has little or no control, since the solution is no longer instantaneously unique (see the discussion in Section 5.3). For these reasons, priority parameters independent on the (history of) flows are implemented in the models of this chapter. According to uniqueness condition (5.13), single-valued priority parameters are required to ensure a unique solution[31] of the intersection model. Simply defining single-valued α_i without additional considerations seems behaviourally plausible only for AWSC intersections. In Section 5.2.4, two approaches to derive single-valued α_i from initially multiple-valued α_{ij} and α_{ik} have been proposed, namely pre-processing of the priority parameters α_{ij} and α_{ik} into one α_i for each i and post-processing of multiple flow solutions. The former approach is adopted, as this seems more theoretically sound and algorithmically efficient than the latter.

$$\begin{aligned} &\alpha_i = \sum_j w_{ij}\alpha_{ij} + \sum_k w_{ik}\alpha_{ik} \quad \forall i \\ &with: \quad \alpha_{ij} > 0 \;\&\; \alpha_{ik} > 0 \quad \forall i,j,k \\ &\sum_i \alpha_{ij} = 1 \quad \forall j \\ &\sum_i \alpha_{ik} = 1 \quad \forall k \\ &w_{ij} \geq 0 \;\&\; w_{ik} \geq 0 \quad \forall i,j,k \\ &\sum_j w_{ij} + \sum_k w_{ik} = 1 \quad \forall i \end{aligned} \tag{6.7}$$

[31] More specifically, it is the priority ratios that are required to be single-valued to ensure solution uniqueness. However, this is equivalent to defining single-valued α_i (see Appendix E).

In (6.7), different priority parameters need to be defined for each (internal or external) conflict that express (the compliance to) the priority rules in that conflict. Then, these are weighted according to their expected influence on the resulting flows over the intersection. The specification of the priority parameters α_{ij} and α_{ik} on the one hand, and the weights w_{ij} and w_{ik} on the other is separately discussed in the following two subsections.

6.1.2.1 Priority parameters

As explained in Section 5.1.2.2, different driver behaviour - ranging from absolute compliance to the priority rules to turn-taking - can be modelled through an according specification of the priority parameters. In the following, a practical overview is given of how to define the priority parameters for some reference cases.

- Turn-taking:
  The relative strength of each i that competes for an external supply in j or an internal supply in k of which the distribution is resolved by turn-taking can be defined by the link capacities:

$$\alpha_{ij} = \frac{C_i}{\sum_{i' \setminus f_{i'j} > 0} C_{i'}}$$
$$\alpha_{ik} = \frac{C_i}{\sum_{i' \setminus f_{i'j} > 0} C_{i'}} \qquad (6.8)$$

- Absolute compliance:
  In this case, an ordering exists that grants priority to some movements over others. Let us denote the rank of each i that competes for an external supply in j or an internal supply in k by m (m = 1..M). Hence, i_1 has absolute priority over all other i (in the considered j or k), i_2 must give way to i_1 but has priority over $i_3..i_M$ and so on. This situation can be described by priority parameters as defined by (6.9). The definition is given for α_{ik}. Of course, if an external supply in j is to be modelled with absolute compliance, α_{ij} can be defined entirely analogously to (6.9).

$$\alpha_{i_1 k} = 1 - \sum_{m=2..M} \varepsilon^{m-1}$$
$$\alpha_{i_2 k} = \varepsilon$$
$$\alpha_{i_3 k} = \varepsilon^2 \qquad (6.9)$$
$$\ldots$$
$$\alpha_{i_M k} = \varepsilon^{M-1}$$

  where ε is an arbitrarily small number.

- Limited compliance:
  Limited compliance to the priority rules can be modelled by any definition of the priority parameters that lies in between the two extremes (6.8) and (6.9).

With the above described reference cases, the model user can define the priority parameters for each (internal) supply according to the expected driver behaviour for that conflict.

6.1.2.2 Weights

Now, the weights w_{ij} and w_{ik} are to be defined that translate the multiple α_{ij} and α_{ik} into one resulting α_i so that solution uniqueness of the intersection model is ensured. On the one hand, it may be conceivable to derive the weights w_{ij} and w_{ik} based on data. However, as this would require tremendous data gathering and calibration efforts, this is more than likely infeasible in macroscopic DNL applications. Rather, we suggest deriving the weights w_{ij} and w_{ik} from the input parameters of the intersection model. This implies that – contrary to the priority parameters α_{ij} and α_{ik} - these weights do not have to be specified by the model user.
Since compliance to the invariance principle dictates that the weights must be independent of the demands S_i and the supplies R_j, this leaves only the turning fractions f_{ij} and the incoming, outgoing and internal capacities (C_i, C_j and N_k). The following definition (6.10) of starting weights is proposed, which are then normalized to w_{ij} and w_{ik} (6.11).

$$\begin{aligned} w_{ij}^{start} &= \frac{f_{ij}C_i}{C_j}\sum_{i'}\frac{f_{i'j}C_{i'}}{C_j} \quad \forall i,j \\ w_{ik}^{start} &= \frac{f_{ik}C_i t_{ik}}{N_k}\sum_{i'}\frac{f_{i'k}C_{i'}t_{i'k}}{N_k} \quad \forall i,k \end{aligned} \tag{6.10}$$

$$\begin{aligned} w_{ij} &= \frac{w_{ij}^{start}}{\sum_j w_{ij}^{start} + \sum_k w_{ik}^{start}} \quad \forall i,j \\ w_{ik} &= \frac{w_{ik}^{start}}{\sum_j w_{ij}^{start} + \sum_k w_{ik}^{start}} \quad \forall i,k \end{aligned} \tag{6.11}$$

The rationale behind (6.10) is the following. A supply that could potentially be more overloaded has a stronger impact on all flows. In (6.10), each w_{ij}^{start} and w_{ik}^{start} consists of two factors. The first expresses the maximum relative consumption of the supply in j or k by i. Indeed, the maximum consumption is obtained if $q_i = C_i$. Similarly, the second factor is the maximum relative consumption by all other i' that compete with i for the same supply. This formulation exhibits the following properties:

- If f_{ij} (f_{ik}) = 0 then w_{ij} (w_{ik}) = 0. This implies that α_i is not influenced by conflicts in which i is not engaged.

- If $\sum_{i'} f_{i'j}$ ($\sum_{i'} f_{i'k}$) = 0 then w_{ij} (w_{ik}) = 0, i.e. if there are no competitors for i in a conflict; this conflict does not influence α_i.

In summary, w_{ij} and w_{ik} are measures of the maximum potential overload of each (internal) supply and the role of a particular i therein. It is not proven or even claimed that this is the best possible definition of these weights, but it is based on reasonable assumptions. Further research should be conducted to confirm these formulations or provide better alternatives.

6.2 Specific intersection models

Using the components presented in the previous section, specific intersection models for different types of intersections can be composed.

6.2.1 AWSC intersection model

At AWSC intersections, turn-taking is prescribed for all merging and crossing conflicts. Consequently, the AWSC intersection model is fully defined by (6.1)-(6.4), with (6.5)-(6.6) for the $\widehat{N}_k$ and the following single-valued α_i (see Section 5.1.2.2):

$$\alpha_i = C_i \quad \forall i \tag{6.12}$$

6.2.2 Priority-controlled intersection model

This model is defined by (6.1)-(6.7), (6.10)-(6.11) for the weights and by (user) specifications for the priority parameters. For internal, crossing conflicts, the priority parameters α_{ik} can be defined as in (6.9). Hereby, the prioritized movement i_1k (as dictated by the signs or markings at the intersection) is granted priority in k over all other movements ($\alpha_{i_1k} = 1 - \sum_{m=2..M} \varepsilon^{m-1}$). Formulation (6.9) is followed for lower ranked movements, up to the movement i_Mk who has to give way to all others ($\alpha_{i_Mk} = \varepsilon^{M-1}$). Of course, if desirable, less extreme values can be appointed to model limited compliance to these priority rules. For external merging constraints, we advise to define capacity-based α_{ij} as in (6.8), because these conflicts – when decisive – are typically resolved by turn-taking. Of course, the imposed priority rules can be followed in the definition of α_{ij} as well. In this case, (6.9) is followed equivalently as for α_{ik} .

6.2.3 PTTR intersection model

This intersection model is in essence identical to the previous one. Of course, the priority parameters α_{ik} are defined such that a movement that approaches all competitors for conflict

k from the right is granted priority[32]. This reasoning is continued up to the movement approaching from the left who has to give way to all others.

6.2.4 Roundabout model

Again, (6.1)-(6.7) is complemented with (6.10)-(6.11) for w_{ij} and w_{ik} and specific detailing of the priority parameters α_{ij} and α_{ik}. The internal conflict zones are located at the merging point on the arcs of the roundabout, where vehicles enter the roundabout. Generally, the entering vehicles have to give way to vehicles already on the roundabout. This can be modelled by appointing $\alpha_{i_M k} = \varepsilon$ to the entering movement. Streams already on the roundabout originating from other i are obviously of equal priority, since they are already mixed at that point. Still, it seems reasonable to assume that the chance of a flow occupying this conflict zone is - among other factors, which are neglected for now - proportional to its entering capacity. Consequently, α_{ik} for streams already on the roundabout are given by a combination of (6.8) and (6.9), i.e.:

$$\alpha_{ik} = \frac{(1-\varepsilon)C_i}{\sum_{i' \setminus \{i_M\} | f_{i'k} > 0} C_{i'}} \tag{6.13}$$

Following the same philosophy, (6.8) can be applied for external merging constraints as always.
As stated before, the possibility of queues spilling back over the roundabout itself is not explicitly considered. If such a situation is to be expected, it is advisable to define all priority parameters of conflicts that are affected by the congestion spillback according to (6.8), since this is the more plausible behavioural assumption for congested merging.

6.2.5 Signalized intersection model

This intersection type requires some additional modelling efforts that complement and adapt (6.1)-(6.11) where necessary. As motivated in Chapter 2, adaptive control is not considered.

Green phases p are introduced. A green phase is defined as a period during which all signals remain the same. Hereby, only effective green and red lights are considered (no amber). We write $i \in p$ to indicate that the light is green for i during p. Analogously, $j \in p$ and $k \in p$ indicates that flow is sent to j (k) during p, i.e. $\exists i \in p : f_{ij}$ (f_{ik}) > 0. Only full green lights are considered, that allow all movements ij to pass. This is in line with the fact that turning lanes are not explicitly modelled. At intersections with green arrows for movements separated in turning lanes, we advise to introduce a diverge upstream of the intersection that physically separates the turning lanes into different incoming links.

[32] Obviously, the conflict zones must be defined small enough. Otherwise, a circular priority could exist within one conflict zone, in which no highest ranked movement can be identified.

When macroscopically modelling signalized intersections, two approaches can be followed. Either the green and red phases are explicitly simulated, or the effects of the traffic lights are averaged to allow larger time steps in a DNL simulation (as in Yperman & Tampère, 2006). If the first option is chosen, the priority-controlled intersection model of Section 6.2.2 can be applied unaltered for every green phase p separately, considering only $i \in p$. This is alternated with periods of zero flow to simulate the red phases. In the remainder, the second option is chosen. For that purpose, not only additional specifications for the internal supply constraints due to the traffic lights are needed, but also their effect on other conflicts has to be accounted for.

A green phase p has a green time equal to a fraction G_p of the cycle time ($\sum_p G_p = 1$). The green fraction for i is simply:

$$G_i = \sum_{p|i \in p} G_p \tag{6.14}$$

The traffic control manages the use of most conflict points by alternately blocking the flows. The isolated effect of the traffic light is then a constraint that further reduces the most stringent part of the infrastructure, so that:

$$q_i \leq \min_{j,k|f_{ij},f_{ik}>0} (C_i, \frac{N_k}{f_{ik} t_{ik}}, \frac{C_j}{f_{ij}}) G_i \tag{6.15}$$

This demand-like constraint is added to the model definition (6.1)-(6.4). Priority parameters do not have to be defined for this constraint, as it can only affect i itself (no distribution is required).

In addition to imposing a constraint of its own in the form of (6.15), the signal control also influences the distribution of other supplies. The traffic control removes many conflicts by separating them into different green phases. Movements that are entirely separated from other movements usually do not have to be considered conflicting anymore. The constraints on these movements are taken care of by (6.15). Likewise, the remaining conflicts between movements that share a green phase p can usually be treated without consideration of other green phases p' during which these movements are blocked. This is so because excess supply in j or k that is unused during p' – whether due to insufficient flow to j or k during p' or because j or k is not served during p' – is lost, and does not have to be redistributed to movements during p. The only exception to the above is when congestion spillback reduces the supply of some j[33], i.e.:

$$\begin{aligned} & R_j < G_j C_j \\ & with: \quad G_j = \sum_{p|j \in p} G_p \end{aligned} \tag{6.16}$$

[33] Recall that the possibility of congestion spillback on internal conflict points k is not considered.

Indeed, if a queue spilling back onto to the intersection constrains the outflow into *j*, supply that is unused during *p'* will result in a temporary shortening of the queue. However, during *p* (when R_j is stringent) the excess supply that has become available during *p'* will be used up, causing the queue to grow again. For congested *j* for which condition (6.16) holds, the second approach discussed below is followed. Other *j* and *k* – for which still a conflict exists at the signalized intersection – are treated according to approach 1.

1. Distribution of partial supply during shared green phases
   a) During one single green phase
      If two or more conflicting movements are confined to one single green phase *p*, the partial supply that is available during *p* is simply:

$$\begin{aligned} R_{j,p} &= G_p C_j \\ N_{k,p} &= G_p N_k \end{aligned} \tag{6.17}$$

      The priority parameters can be defined as in (6.8) or (6.9), dependent on which type of behaviour is to be modelled. For the weights, G_p is introduced in the numerator and denominator of all factors and thus cancels out so that (6.10)-(6.11) can be applied unaltered.

   b) During partially overlapping green phases
      It may occur that some movements are in conflict during mutual green phase(s) *p*, but that one or more movements are allowed during additional green phase(s) *p'* in which all or some competing movements are blocked. In this case, the combination *P* of these green phases is jointly considered:

$$\begin{aligned} &P = \{p, p' \mid \exists i \in p \cap p'\} \\ &with: \quad \#i \in p \geq 2 \end{aligned} \tag{6.18}$$

      Then, the supply during *P* that is to be distributed among all $i \in P$ is defined by:

$$\begin{aligned} &R_{j,P} = G_P C_j \\ &N_{k,P} = G_P N_k \\ &with: \quad G_P = \sum_{p,p' \in P} G_p \end{aligned} \tag{6.19}$$

      The above implies that supply that is not used up by some movement is redistributed among every other $i \in P$. Note that it is conceivable that this redistribution appoints more supply to some *i* than this *i* can consume within its green time. Constraint (6.15) ensures in this case that the flow q_i is still bounded by its green time.
      Now, since the competition for the supplies (6.19) is not the same during all green phases that are combined in *P*, this has to be accounted for in the

priority parameters and weights. Let us define $\alpha_{ij,p}$ ($\alpha_{ik,p}$) - according to (6.8) or (6.9) - as the priority parameters that apply to *j* (*k*) during *p*, considering of course only $i \in p$. The $\alpha_{ij,p'}$ ($\alpha_{ik,p'}$) are defined analogously. Note that if *i* is the only element of *p'*, obviously $\alpha_{ij,p'}$ ($\alpha_{ik,p'}$) $= 1$. Then, the priority parameters that apply to the distribution of the supplies in (6.19) are given by:

$$\begin{aligned} \alpha_{ij,P}^{start} &= \sum_{p|i \in p} G_p \alpha_{ij,p} + \sum_{p'|i \in p'} G_{p'} \alpha_{ij,p'} \quad \forall i, j \in P \\ \alpha_{ik,P}^{start} &= \sum_{p|k \in p} G_p \alpha_{ik,p} + \sum_{p'|i \in p'} G_{p'} \alpha_{ik,p'} \quad \forall i, k \in P \end{aligned} \tag{6.20}$$

$$\begin{aligned} \alpha_{ij,P} &= \frac{\alpha_{ij,P}^{start}}{\sum_{i' \in P} \alpha_{i'j,P}^{start}} \quad \forall i, j \in P \\ \alpha_{ik,P} &= \frac{\alpha_{ik}^{start}}{\sum_{i' \in P} \alpha_{i'k,P}^{start}} \quad \forall i, k \in P \end{aligned} \tag{6.21}$$

Finally, the starting weights are given by (6.22), which are again normalized according to (6.11):

$$\begin{aligned} w_{ij}^{start} &= \frac{f_{ij} G_i C_i}{G_P C_j} \sum_{i'} \frac{f_{i'j} G_{i'} C_{i'}}{G_P C_j} \quad \forall i, j \in P \\ w_{ik}^{start} &= \frac{f_{ik} G_i C_i t_{ik}}{G_P N_k} \sum_{i'} \frac{f_{i'k} G_{i'} C_{i'} t_{i'k}}{G_P N_k} \quad \forall i, k \in P \end{aligned} \tag{6.22}$$

2. Distribution of external supply over separated green phases
   For an outgoing link *j* on which congestion spills back onto the intersection (6.16), largely the same approach as in 1.b. is followed. The difference is twofold. Firstly, the full supply R_j is distributed at once and not split up into separate (groups of) green phases. As a logical result, secondly, all $i|f_{ij} > 0$ are included in the distribution. Analogously to (6.20)-(6.22), the following formulations define the priority parameters and weights. Hereby, the $\alpha_{ij,p}$ are defined according to (6.8), to model turn-taking between all $i \in p$.

$$\alpha_{ij}^{start} = \sum_{p|i \in p} G_p \alpha_{ij,p} \quad \forall j \text{ for which (6.13) holds} \tag{6.23}$$

$$\alpha_{ij} = \frac{\alpha_{ij}^{start}}{\sum_{i'} \alpha_{i'j}^{start}} \quad \forall j \text{ for which (6.13) holds} \tag{6.24}$$

$$w_{ij}^{start} = \frac{f_{ij} G_i C_i}{G_j C_j} \sum_{i'} \frac{f_{i'j} G_{i'} C_{i'}}{G_j C_j} \quad \forall j \text{ for which (6.13) holds} \tag{6.25}$$

The above is a simplification, since the composition and order of the green phases is not fully accounted for in the redistribution of supply. Let us illustrate this with an example. Suppose that some R_j is active and four flows compete for it, two of which share the same green phase. If one of the latter flows is demand constrained, then in reality the supply that is not consumed by the demand constrained link will be available first for the other flow in the shared green phase. When the latter cannot fully consume its increased share, the opportunity goes to the third flow in the subsequent green phase. Only if this flow also cannot fully consume its increased share, the fourth flow will profit. In the above described model, all competing flows will receive an increased share of supply as soon as one of the flows cannot fully claim its share. It is envisaged that this simplification has minor impact in the majority of applications.

More so than in the previous intersection model types, some simplifying assumptions are introduced to keep the signalized intersection model manageable. The possibility of building a more complex model that avoids some of the current simplifications should be investigated in future research.

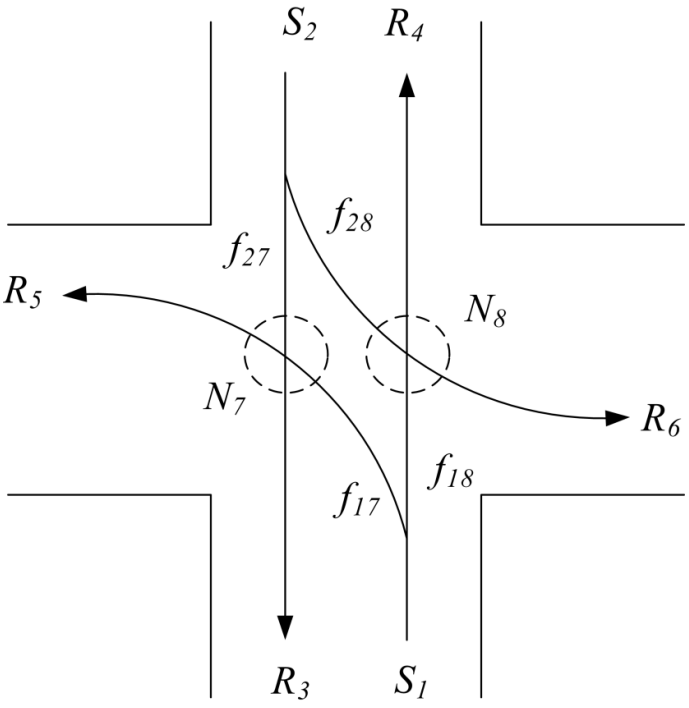

Figure 6-2: Revisiting the example of Section 5.2.1

6.3 Numerical example

In this section, a numerical example is presented to create a better understanding of the solution provided by the presented intersection models. For this, the priority-controlled example used throughout Chapter 5 is revisited (see Figure 6-2). This intersection type is used as a reference since it has the most communal properties with other intersection types.
As can be seen in Figure 6-2, each turning fraction f_{ik} (to the internal conflicts) corresponds to one turning fraction f_{ij} (to the outgoing links). In the following, we refer only to the f_{ik}.

The solution produced by the model in Section 6.2.2 is compared to that of the model in Section 4.2, which is limited to external supply constraints, to demonstrate the added value of including internal constraints. Also, its solution is compared to that of a model that persists with multiple-valued α_{ik} instead of averaging into one representative α_i. Let us denote the latter model as 'model_non-unique', the model of Section 6.2.2 as 'model_unique' and the model of Section 4.2 as 'model_ext'. The solution is compared in two scenarios for different values of the turning fractions. Indeed, the turning fractions are particularly influential for the solution of the presented intersection models. This is so because they largely determine the weights (6.10)-(6.11), and with that the single-valued α_i that appoint the solution. (Of course, the capacities are an important factor in (6.10) as well, but these are – at least to a much larger extent – fixed for a given intersection.) The other boundary conditions that serve as input to the intersection model are considered fixed and equal to:

- $C_1 = C_2 = 2000$ veh/h
- $S_1 = S_2 = 1600$ veh/h
- $R_3 = R_4 = 2000$ veh/h
- $R_5 = R_6 = 1000$ veh/h
- $N_7 = N_8 = 3600$ s/h
- $t_{17} = t_{18} = t_{27} = t_{28} = 3.6$ s

Naturally, the internal constraints and all corresponding variables do not apply to model_ext. For the other two models, the priority parameters α_{ik} are defined for the internal conflicts. In this example, they express absolute priority for movements 14 and 23:

- $\alpha_{17} = \varepsilon;\ \alpha_{27} = 1-\varepsilon$
- $\alpha_{18} = 1-\varepsilon;\ \alpha_{28} = \varepsilon$

Two scenarios are considered. The first is a symmetric example, with all f_{ik} changing together. In the second, f_{27} and f_{28} are fixed while f_{17} and f_{18} fluctuate.

6.3.1 Scenario 1: symmetric example

In this scenario, the turning fractions for both minor movements are increased simultaneously from 0 to 100% of the total demand; i.e. $f_{17}: 0 \to 1$ and $f_{28}: 0 \to 1$. Obviously, f_{18} and f_{27}

change reversely. The results of the models are depicted in the following figures as a function of f_{17}.

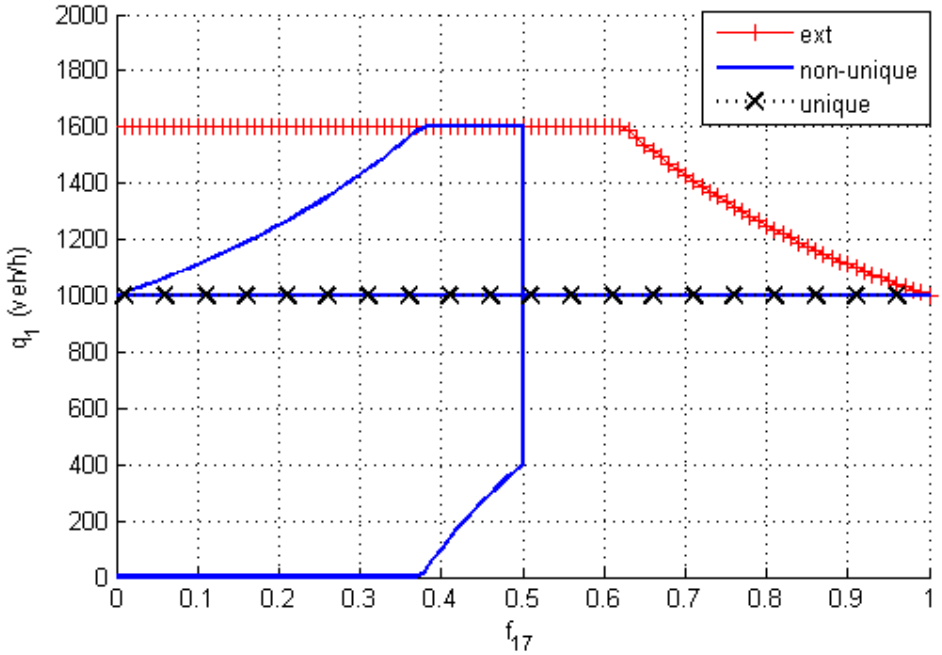

Figure 6-3: Scenario 1: q_1

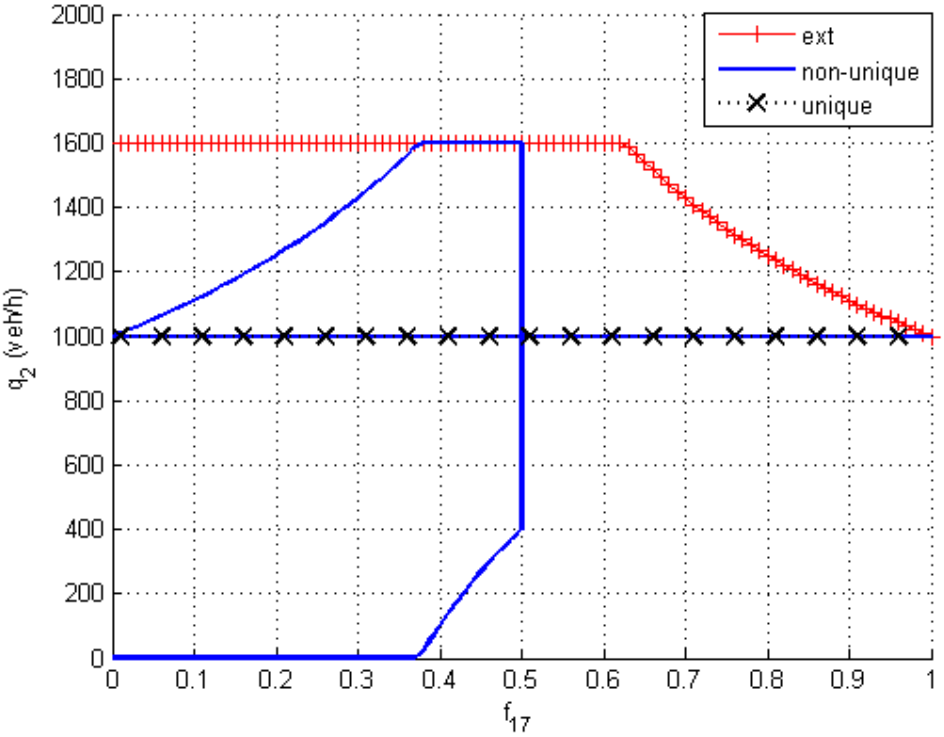

Figure 6-4: Scenario 1: q_2

It can be seen from Figure 6-3 and Figure 6-4 that neglecting the crossing conflicts causes model_ext to significantly overestimate the flows. The results of model_ext are indicated by the striped red line ('ext' in the legend). Only for high f_{17} and f_{28}, the supplies R_5 and R_6 respectively constrain q_1 and q_2.

Model_non-unique (indicated by 'non-unique'; the full blue line) produces three solutions for $f_{17} = f_{28} =]0, 0.5[$. Of course, each high solution for q_1 corresponds to a low q_2 and vice versa. The symmetric solution line ($q_1 = q_2 = 1000$ veh/h) corresponds to the crossing point of the two internal supply constraint functions (point C in Figure 5-3). Moreover, for $f_{17} = f_{28} = 0.5$, the two internal supply constraints entirely coincide, so all flow combinations for which $\widehat{N}_7 = \widehat{N}_8 = 0$ (and that fall within the demand constraints) take the characteristics of point C in Figure 5-3. Thus, for $f_{17} = f_{28} = 0.5$, an infinite number of solutions arises. This example clearly shows the difficulty of algorithmically finding all possible solutions.

Model_unique (indicated by 'unique'; the dotted black line) always produces the same symmetric solution of $q_1 = q_2 = 1000$. Indeed, (6.7) and (6.10)-(6.11) always lead to $\alpha_1 = \alpha_2 \Leftrightarrow \frac{\alpha_1}{\alpha_2} = 1 \Rightarrow q_1 = q_2$. This seems plausible because both links always send the same amount of flow into their prioritized and minor direction.

6.3.2 Scenario 2: non-symmetric example

In this scenario, f_{27} and f_{28} are fixed to 0.5. f_{17} is again varied from 0 to 1. For $f_{17} = 0$, there is no conflict for N_7. This means that q_1 (which is prioritized in N_8) is dominant in this case. As f_{17} rises, the conflict for N_7 gains in importance, so that the benefit shifts to q_2.

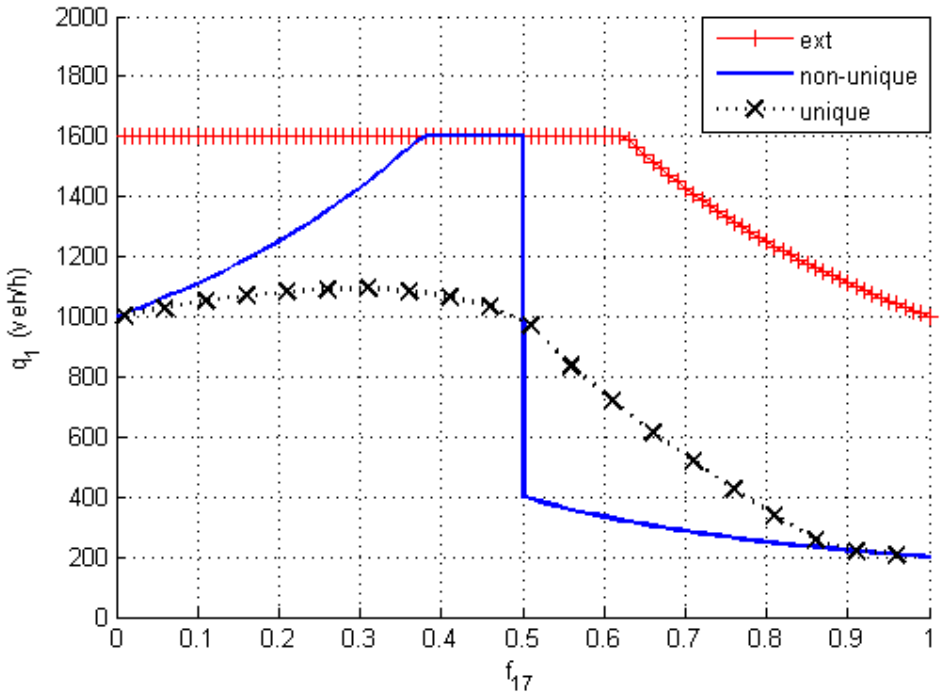

Figure 6-5: Scenario 2: q_1

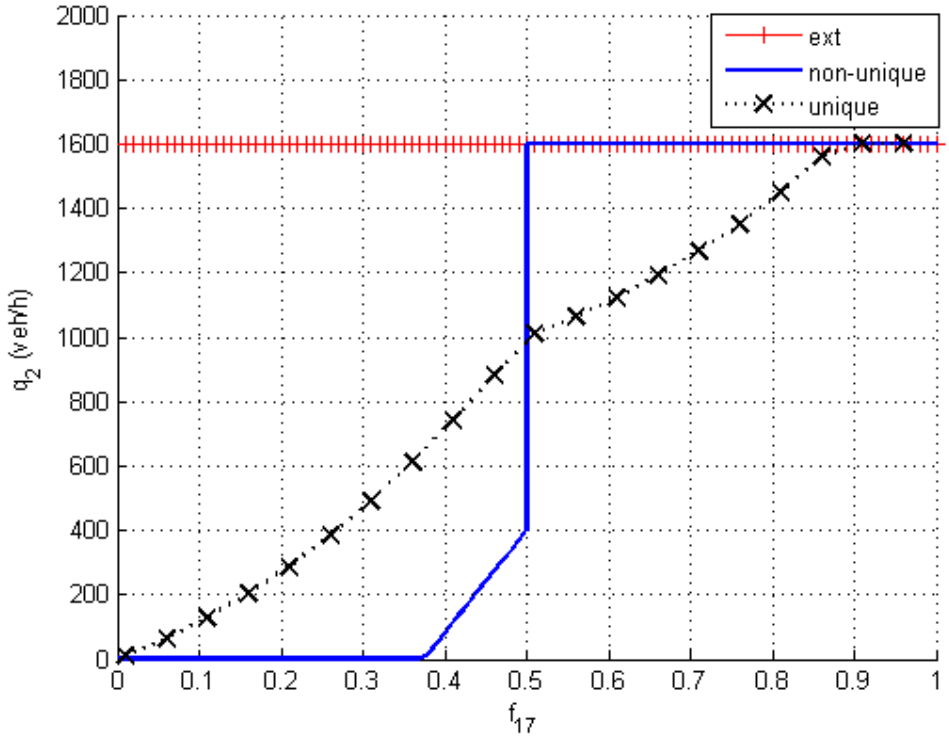

Figure 6-6: Scenario 2: q_2

Again, it is clear that model_ext consistently overestimates the flows. The solution of model_non-unique either favours q_1 (when $f_{17} < 0.5$; and the conflict for N_8 is dominant) or q_2 (when $f_{17} > 0.5$; and the conflict for N_7 is dominant). At $f_{17} = 0.5$, analogously to the previous example, an infinite number of possible solutions arises.
The solution of model_unique progresses continuously; with q_2 systematically increasing while q_1 first slightly increases before starting to diminish. This can be explained as follows. Initially, the rise in f_{17} causes q_1 to increase[34], since more flow is sent in the unconstrained direction. Meanwhile, q_1 is still strongly prioritized. Figure 6-7 depicts the priority ratio between the two incoming links, which appoints the solution of model_unique. As f_{17} rises further, link 1 loses priority to link 2 so that q_1 decreases.

In conclusion, the above two scenarios show in simple examples how the intersection models of this chapter – by means of the priority-controlled model of Section 6.2.2 – solve the problem of solution non-uniqueness. While - in absence of empirical validation - it is not postulated that this guaranteed unique solution is always more realistic than potentially non-unique ones (see the discussion in Section 5.2.3.2), it is a necessary property in deterministic macroscopic DNL models. However, to ensure this solution uniqueness, several aspects of reality have been omitted from the intersection model. This is extensively elaborated on in the following concluding discussion. In any case, the above examples visually demonstrate

[34] Note that for $f_{17} < 0.5$, both q_1 and q_2 are constrained by N_8 (which only allows 1000 veh/h to pass).

the obvious fact that neglecting the internal constraints – as is usually done in existing DNL intersection models – may cause a significant overestimation of the flows.

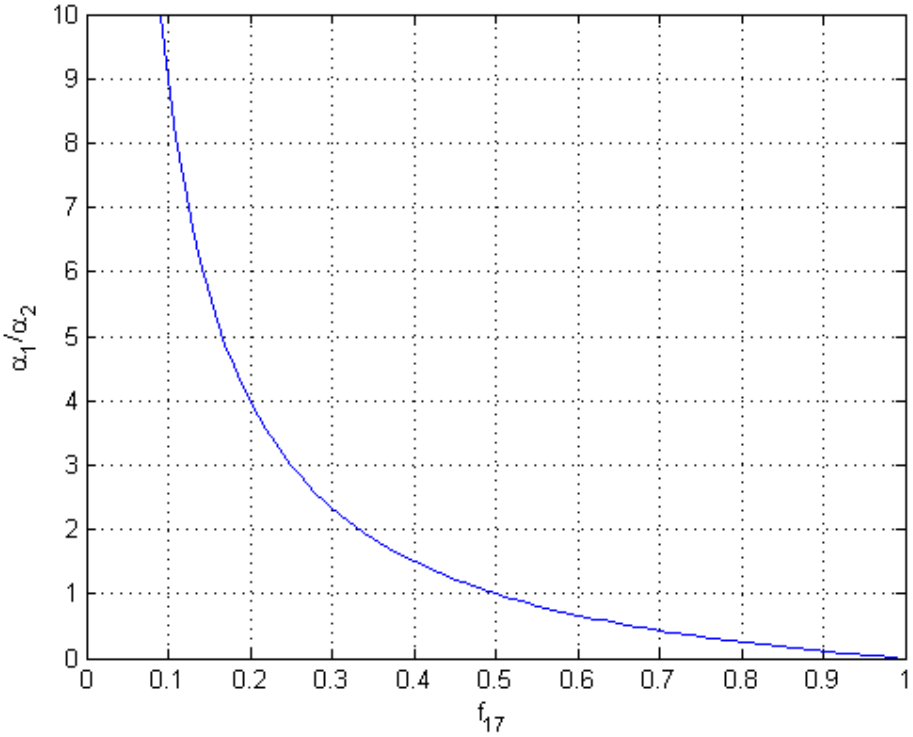

Figure 6-7: Scenario 2: priority ratio α_1 / α_2

6.4 Practical guidelines for the use and validation of the proposed models

The intersection models of this chapter are conceptually but also computationally more complex than the model in Chapter 4. Therefore, especially for large-scale networks, it may not be feasible to model every intersection in such detail. Moreover, for quite some intersections (at which flow restrictions or significant delays do not occur) this will not be necessary. Hence, one should be somewhat cautious to determine for which intersections in the network to use the more complex intersection models including internal supply constraints. We suggest two steps to make this selection (see further on). In any case, if the DNL model is used iteratively, e.g. in a DTA model or in optimal control applications, it is wise to use only simple intersection models in the first few iterations. When the simulation results have stabilized to some extent, another level of accuracy can be added by making a selection of intersections that are to be modelled in more detail.

1. Based on knowledge of the network and/or available traffic data, make a selection of intersections on which the internal conflicts are expected to have a non-negligible influence on traffic flows or delays. Hereby, intersections with a low degree of saturation can be excluded.

2. Run a DNL simulation (or in iterative procedures, e.g. in DTA, a few more iterations) with complex intersection models for the selection made in the previous step and simpler intersection models (as in Chapter 4) for the remaining intersections. Then, extend the selection with intersections for which for some link(s) the demand exceeds a certain percentage of the maximum outflow (determined by the link capacity or the share of the most stringent supply) in the solution of the simple intersection model. This percentage depends on the required accuracy and the available time budget. In any case, it should be no higher than 0.8, since above this degree of saturation the delay increases rapidly (see chapters 8 and 9 in Gartner et al., 2000).

Now, although including internal constraints inevitably renders the intersection model more complex, the proposed models are relatively parsimonious. In addition to link capacities and (for signalized intersections) green fractions, which are also necessary in simpler models, the intersection models of this chapter require the following input to be specified by the model user[35]:

- Grouping of movements into conflict zones:
  As explained in Section 6.1.1.2, the movements are to be grouped in conflict zones that correspond to a shared part of the intersection infrastructure that is consecutively used by vehicles of these movements. This must be established from observation of the exact routes that vehicles follow when crossing the intersection, and where the routes of different movements intersect. For additional information we refer to Wu (2000), Brilon & Wu (2001) and Brilon & Miltner (2005).

- Internal supplies N_k:
  These can simply be set to 3600 s/h (6.5). Only if a conflict zone is defined that could be occupied by two vehicles at a time, a value of 7200 s/h should be chosen.

- Consumption times t_{ik}:
  These can be derived from the (turning) lane capacities (6.6), which are easier to estimate or measure. Alternatively, the consumption times could be directly measured from (video) observations of the conflict zones.

- Priority parameters α_{ij} and α_{ik} :
  One way to define these is to follow the priority rules of the intersection and assume absolute compliance. In this case, (6.9) is used ((6.13) for roundabouts). For

[35] Recall that demands and supplies are provided by the link model and do not have to be specified by the model user. In multi-commodity DNL models, this holds for the turning fractions as well. In single-commodity models, however, the turning fractions do have to be predefined, or can be derived from the anticipated arrival order (Blumberg & Bar-Gera, 2009).

(over)saturated merging conflicts into outgoing links, assuming turn-taking (6.8) is more advisable. The priority parameters could be more rigorously defined (assuming limited compliance) if empirical observations are available. Then, for each internal and external conflict, all situations in which vehicles from competing movements simultaneously want to advance over the intersection have to be analyzed. It should then be identified in how many cases each movement takes priority (as is done Brilon & Miltner, 2005) in order to specify the α_{ij} and α_{ik}.

From (6.10)-(6.11), the weights w_{ij} and w_{ik} are determined from other variables and thus do not have to be specified by the model user.

In future research, the proposed intersection models should be empirically validated (and possibly refined). First of all, we recall that also an inquiry into the possibility of non-unique flow patterns should be included in this empirical research (see Section 5.2.3.2), as this might disprove the validity of intersection models that yield a unique solution (for some situations).
When empirically validating the models, of course all input parameters should first be calibrated as described above. Also, data describing the boundary conditions (demands, supplies, turning fractions) is needed as input to the intersection model. The flows from each incoming to each outgoing link of the intersection must be measured and aggregated in time segments of a few minutes to allow comparison to the outcome of the macroscopic intersection model. Of course, data should be gathered under varying degrees of saturation. Also, ideally, cases should be included of congestion spillback from an outgoing link as well as queue formation due to internal conflicts. This will allow validating the intersection model to various traffic conditions.
If model refinements turn out to be necessary, we advise to calibrate the weights w_{ij} and w_{ik} until the model produces good results. Then, an alternative formulation for these weights should be sought that produces values that satisfactorily match the calibrated values. If such an adjustment proves insufficient, the internal supply constraint functions may have to be adjusted, so that the effects of microscopic dependencies that exist at the level of individual vehicles (e.g. simultaneous driver decisions when crossing multiple conflicts) are included. This is further elaborated on in the following, concluding section.

6.5 Conclusion

In Section 6.1, we have further detailed the internal supply constraint functions (based on the conflict theory of Brilon & Wu, 2001) and the weighting of multiple-valued priority parameters into a single value (as suggested in Section 5.2.4). The latter is a pragmatic approach to ensure solution uniqueness of the model. Hereto, multiple-valued priority parameters are defined per conflict, reflecting the expected driver behaviour regarding the priority in that conflict. Then, sensible weights are suggested that depend on the maximum amount of competition for a conflict. Further research is necessary, however, to validate and if necessary refine both the suggested definition of the internal supply constraint functions as the pre-processing approach of the priority parameters (as is discussed in Section 6.4).
Combination of the appropriate components defined in Section 6.1 leads to intersection models specific for AWSC, priority-controlled, PTTR, roundabout and signalized

intersections in Section 6.2. To our knowledge, these are the first models that combine both functions of the intersection model (as discussed in Section 2.1) into a consistent, unique solution. Since the intersection models presented in this chapter detail the (distribution of) internal supply constraints while preserving the analogy with the model of Chapter 4, the solution algorithm of Section 4.3 can straightforwardly be extended to implement these models.

The additional modelling assumptions for the intersection models of this chapter - necessary to ensure a unique solution - are made as realistically as possible within what we now perceive to be the boundaries allowed by the requirements put forward in Chapter 3 and the uniqueness condition of Chapter 5. An important conclusion is that these boundaries are tight. Foremost, simultaneous decisions are now neglected in the presented models. Simultaneous decisions occur when drivers from a minor movement have to cross several conflicts with prioritized movements. In the example in Figure 6-8, the minor flow from the eastbound link has to give way to bi-directional traffic on the major street (unless some storage space would be present in the middle of the intersection).

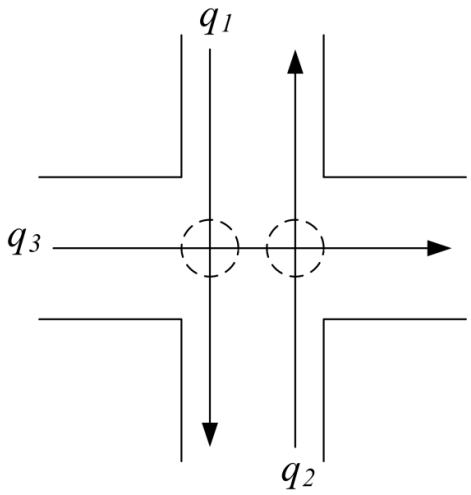

Figure 6-8: Simultaneous driver decisions involving multiple conflicts

Since a vehicle from the minor street will only cross if no prioritized vehicle is approaching from either side, a constraint in the form of (6.26) is needed.

$$q_3 \leq \widehat{N}(q_1, q_2) \tag{6.26}$$

In (6.26), q_1 and q_2 each individually consume supply that is in result no longer available for q_3. Hereby, however, q_1 and q_2 do not hinder each other. This corresponds to how most existing models include internal supply constraints. As explained in Section 5.1.2, this formulation corresponds to absolute compliance to the priority rules. This implies that this correspondence is violated when including constraints in the form of (6.26) and solving them through α_i that weigh α_{ij} and α_{ik} from individual conflicts. Holding on to the absolute compliance in (6.26) on the other hand inevitably reintroduces solution non-uniqueness if other conflicts are solved according to different priorities.

Now, except for priority-controlled intersections (and possibly some signalized intersections) neglecting the above described dependency is defendable. This may be because driver decisions are limited to a single location (at roundabouts); or because no simultaneous decisions are to be made since drivers do not (consistently) have to give way in subsequent conflicts (AWSC, PTTR).

However, even for these types of intersections, other dependencies exist at a microscopic level that may be decisive for the macroscopic flows. For example, the order in which a movement encounters conflicts obviously influences the competitive strength of that movement in subsequent conflicts. For instance, a right-turning movement has an advantage in a merging conflict (regardless of priorities) over the opposite left-turning movement that is first hindered by one or more crossing conflicts. Including this effect into the model requires the definition of priority functions $\alpha(\mathbf{q})$ instead of constant priority parameters. Realistically defining such functions while ensuring uniqueness condition (5.13) seems very difficult.

Finally, future research aiming to include separate turning lanes into the intersection model would be useful. Also, further analysis of the influence of the previous history on instantaneous flow solutions might be interesting, although the (possibly seminal) exploration in Section 5.3 suggests that this introduces undesirable and uncontrollable dynamics into the model.

In conclusion, further research is needed to investigate the possibilities and consequences of model extensions as described above. It is envisaged, however, that introducing these microscopic effects into macroscopic DNL intersection models cannot be fully reconciled with all of the desirable model properties acquired in this thesis (namely solution uniqueness and the requirements of Section 3.4.1). In this case, it may be necessary to (partially) release some of the postulated model properties. The insight provided in this thesis will help to make well-considered decisions on this matter. As such, the models presented in this chapter are another important intermediate step, rather than the end of the line. They include sensitivity to internal conflicts – which are usually ignored in state-of-the-art models – in a theoretically consistent, but still somewhat simplified way.

PART II:
MARGINAL SIMULATION

7

MARGINAL SIMULATION: INTRODUCTION AND SCOPE

This chapter introduces the novel concept of marginal DNL simulation and thus forms the onset to the second main part of this PhD thesis. Marginal simulation algorithms only compute variations to a base scenario instead of performing full simulations. Their purpose is to overcome the problem of high computation time that comes with using sophisticated DNL models with realistic congestion dynamics (e.g. first-order models such as LTM) for repeated simulations. While our research of marginal simulation originates from aiming to support Travel Time Variability (TTV) studies, it gradually became clear that the applicability of marginal DNL simulation surpasses this original aim. Indeed, it may be of aid in any study that involves repeating a large number of consecutive simulations with large overlap. Other possible applications for road traffic networks include (robust) network design (see Snelder, 2010) and determining (flow) sensitivities (e.g. Jacobian matrices) that can be used to find optimization directions, for instance in optimal control and in gradient-based dynamic OD estimation (see Frederix et al., 2011).

Section 7.1 provides a general overview of how the computational inefficiency that comes with the need for a large number of quasi-identical simulations is treated in the state-of-the-art. Section 7.2 proposes marginal simulation as a general solution to such problems.

7.1 Overcoming the computational inefficiency due to repeated simulations: literature overview

Many simulation-based studies of road traffic require a large number of simulations to be performed on the same network. This includes for instance TTV studies and applications in which one may wish to numerically derive[36] the gradient of the objective function (e.g. minimizing total time spent or the deviation between simulated link flows and detector counts) to the input variables (e.g. route demands or link capacities). Examples of the latter are OD estimation, (robust) network design and optimization of traffic control measures such as dynamic pricing, ramp metering, signal planning and route guidance. As explained in Section 1.1.2, macroscopic simulation-based DNL models based on traffic flow theory are well suited for the above mentioned applications, especially on congested networks. On the downside, the relatively high computation times of these models render performing a large number of simulations troublesome. Particularly real-time or large-scale applications may be infeasible. In many studies, the computational burden currently limits the scope to few scenarios and small networks[37].

Several ways to reduce the computational burden are reported in the literature. A first way is to limit the number of simulation runs. For example in optimization problems, imposing a maximum number of runs is often unavoidable. Obviously, this is undesirable, since reaching an optimum is not guaranteed. Also in TTV studies a small set of simulations does not allow to correctly construct probability distributions of travel times. Preferably, the number of required simulations should be reduced by smart sampling, i.e. focusing on those samples for which a large impact on the (network) performance or the objective function is expected (Tampère et al., 2009). For instance vulnerability analyses are sometimes defined as two player games, in which one player – the 'evil entity' – aims at maximizing the damage by striking only the most vulnerable links (e.g. Murray-Tuite & Mahmassani, 2004). Although smart sampling is advisable when tackling large-scale problems, additional computation time savings are often necessary.

A second option is to revert back to simpler, faster tools like (static or analytical) models with link exit or performance functions. Examples of studies applying such models are plentiful in the literature. Schrijver (2004) introduces the SMARA model, which quantifies TTV due to demand and capacity variations through static Monte-Carlo simulation. Also Chen et al. (2002) use a static assignment in their Monte-Carlo framework to assess reliability under correlated variations of link capacities. Noland et al. (1998) and Lo & Tung (2003) apply a model based on link performance functions to obtain probabilistic distributions of travel times under stochastic link capacities. Snelder et al. (2007) do not consider congestion effects for their optimal re-design of the Dutch road network, and neither do Jenelius et al. (2006) and Jenelius (2010) when assessing network vulnerability by imposing link closures. Although the application of simple, fast models can be justified for networks with relatively low traffic volumes, proper modelling of congestion spillback is vital to obtain credible results on congested networks (Knoop et al., 2007).

[36] Numerical derivation may be desirable if an analytical expression is not available due to the non-linearities and discontinuities that are introduced by (foremost) congestion spillback.

[37] For example in Ukkusuri et al. (2010), who present an approach for robust signal control under varying demand with an embedded CTM, this is explicitly stated.

Another strategy that is often applied to lower the computational burden is to develop approximate methods so that repeated simulations are (largely) avoided. For example in real-time control applications, feedback approaches are often used instead of potentially more effective, but troublesome iterative procedures (e.g. Pavlis & Papageorgiou, 1999). Others use analytical approximations to avoid Monte-Carlo simulation; e.g. Bell et al. (1999), who describe the relation between demand variations and route TTV. Also Clark & Watling (2005) analytically estimate TTV under stochastic demand. Qian & Zhang (2011) approximate the sensitivity of link Cumulative Vehicle Numbers (CVN) to the addition of one unit of flow. Furthermore, Ukkusuri & Waller (2006) propose several approximation schemes to determine the solution to the PUE problem through evaluation of one single-point demand pattern. Following the same philosophy, Ng & Waller (2009) transform a range of stochastic link capacities to one deterministic value. By doing so, they account for the effect of TTV due to supply variations on route choice. Some other work avoids cumbersome simulations by developing simple quick scan methodologies for vulnerability analysis (Scott et al., 2006; Tampère et al., 2007).
Furthermore, in dynamic gradient-based OD estimation (e.g. Cascetta & Postorino 2001, Bierlaire & Crittin 2004), the gradient of the objective function - usually expressing the deviation between simulated link flows and detector counts - to the demand is typically not determined numerically through repeated simulations (finite differences). Rather it is approximated by a so called assignment matrix to speed up the optimization procedure. Frederix et al. (2011) and Frederix et al. (in press) show, however, that the assignment matrix does not suffice on congested networks and discuss the importance of using an improved gradient in dynamic OD estimation.
Finally, an alternative to repeated simulation is proposed by Sumalee et al. (2011) and Osorio et al. (2011) in the form of a stochastic DNL model, propagating stochastic network flows under uncertain demand and supply. The mean and standard deviation of densities are calculated and propagated through the network, as are the probabilities of various traffic states. However, the independence assumptions that need to be made between traffic states and flows in adjacent cells or links form a severe restriction on the realism of such models. Further research of stochastic DNL modelling to overcome this current issue is valuable, however. Potentially, stochastic DNL models may become a valuable tool for many of the aforementioned applications (e.g. TTV studies).

In conclusion, approximate methodologies as described above typically lack general applicability and/or a sufficiently realistic consideration of congestion dynamics. The latter is true for simple, fast (e.g. static) models as well. Therefore, while some can definitely be valuable for certain applications or in certain conditions, e.g. with little or no congestion, they are inapplicable or inappropriate in other cases.
In this dissertation, marginal DNL simulation is proposed. This is also an approximation aiming at computational gain, but not by avoiding repeated simulation. Rather, the simulation runs themselves are approximate. In that sense, conceptually similar approaches can be found in Chiu et al. (2007) and AbdelFatah & Mahmassani (1998). The former propose a method to transform the network and OD-matrix to allow more efficient simulation of mass evacuation. In the latter, an approach is described to optimize signal green times through repeated simulation of only a part of the network, namely a manually defined local area around the

considered signal control. The authors do not use this approach in their experiments, however.
The following paragraph discusses the philosophy, concept and benefits of marginal simulation.

7.2 Rationale and scope of marginal simulation

When repeated simulation runs with large overlap[38] are to be performed, marginal DNL simulation exploits the fact that the input - both in terms of supply (network topology, link capacities and other characteristics) and demand (route choice, OD matrix) - is to a large extent identical in the successive simulations; and therefore also the output. Instead of performing each case as a full explicit simulation, marginal simulation aims at calculating only the changes to the outcome of a base simulation. This base simulation is performed (in full) by the maternal base model, from which the marginal simulation algorithm should be derived (see below). These changes arise from (local) variations to the base input and are tracked as they propagate in time and space. As such, only the part of the input that differs from the base scenario invokes new calculations, and the focus is expanded as the marginal simulation progresses - but only when and where necessary. Hence, no or few identical calculations are performed in consecutive simulations. By focusing only on the (most) relevant part of the network and time period, the simulation effort is limited in space and time. Of course, (variations to) several base scenarios could be considered if necessary.
It should be noted that, while this concept is innovative in the field of transportation, highly similar approaches have been proposed in the design of digital hardware circuits (Hwang et al., 1988; Salz & Horowitz, 1989). In fact, marginal simulation may provide a solution in many research domains or applications with computationally demanding simulations. If the core of the computational inefficiency is that a lot of identical calculations are executed throughout repeated explicit simulations, marginal simulation can yield considerable advantages, especially for large-scale problems.

The marginal simulation algorithm should be drafted from that of the maternal base model. As such, marginal simulation does not require the development of new theories to describe the process that is simulated. In fact, if there is a discrepancy between the base model and the marginal derivation, the accuracy of the results (and possibly also of the computational performance) are negatively affected. Indeed, a marginal algorithm is designed to calculate the changes that arise from imposed variations to a base scenario (which are typically local and quite small). A discrepancy compared to base model, however, may create changes in the marginal simulation results compared to the base reference. Such changes do not result from the imposed variations and hence are approximation errors. Moreover, such errors may be propagated by the marginal simulation in the same manner as the justified changes. If this happens, computation time is wasted for tracking errors.

[38] This implies that a large part of simulation results are identical. This is the case for many applications. Of course, other applications require repeated simulations that exhibit many (albeit small) changes, such as dynamic process simulations or many iteration schemes for equilibrium DTA - e.g. Method of Successive Averages (MSA). Marginal simulation cannot be applied to this kind of repeated simulations.

Anyhow, when applying marginal simulation, some approximation errors are probably unavoidable. To some extent, marginal simulation could be regarded as the simulation-counterpart of an analytical Taylor approximation: locally a good approximation is obtained; further away from the base scenario the error will grow larger (see Section 9.3).

Since marginal simulation is an approximation designed for fast iterative, finite difference or Monte-Carlo simulation, its application should be limited to problems where the outcome of each individual simulation is not directly of interest. Rather, the aim should be to correctly quantify the properties of the set of simulations, for instance its probabilistic characteristics (in TTV studies) or the optimization direction it produces.

7.3 Conclusion

This chapter generally describes the concept of marginal simulation. By performing marginal simulations as variations to a base simulation, calculations are limited to a small part of the network and time period. By doing so, the computational restrictions of repeating a large number of quasi-identical simulations can be relieved to a large extent, while still considering congestion dynamics in a realistic way. In the following two chapters, two marginal simulation algorithms are developed that fulfil this aim with (relatively) small approximation errors, namely the Marginal Incident Computation (MIC) algorithm and the Marginal Computation (MaC) algorithm. For both, the maternal base model is LTM (see Appendix A). The main discrepancy with LTM is that the Multi-Commodity (MC) representation of traffic is approximated by a computationally more efficient Single-Commodity (SC) representation (see Appendix B).

A wide range of potential applications of marginal simulation is identified. Firstly, it can be used as an efficient tool in Monte-Carlo TTV studies. As this has been the context in which the idea of marginal DNL simulation has originated, TTV studies and the current challenges therein are more thoroughly discussed in Appendix F.1. Also, a proof-of-concept case study of a stochastic DTA is explored there, applying MIC to quantify TTV due to incidents.
Other applications for road traffic include optimization problems such as dynamic OD estimation, optimal control (e.g. ramp metering) and evacuation planning, in which marginal simulation could be used to numerically derive the gradient of the objective function[39] (as in Frederix et al., 2011). In robust network design, marginal simulation could be applied to simulate a wide range of possible scenarios to quantify TTV (as in Snelder, 2010) and on the optimization level to determine the sensitivity of the network's performance to changes in the network design. Furthermore, marginal simulation may also support (real-time) DTM decisions, for instance by predicting a probabilistic range of the possible impact of an incident.
Finally, it is anticipated that marginal simulation could also be beneficial in other domains than road traffic, for instance optimal factory design or supply chain management. In fact, it

[39] Especially in the first iterations - when the distance to the optimum is still large – fast marginal simulation probably suffices to determine the optimization direction. In the last iterations, one could divert to explicit simulation if higher accuracy is required to reach the optimum.

has already been applied to design digital hardware circuits (Hwang et al., 1988; Salz & Horowitz, 1989).

8

MARGINAL INCIDENT COMPUTATION (MIC)

This chapter presents the Marginal Incident Computation (MIC) algorithm, a computationally highly efficient marginal simulation tool. It is specifically designed for approximately quantifying congestion effects and the corresponding travel time increases due to incidents. MIC superimposes the effect of an incident onto a single base simulation run instead of carrying out a complete DNL with the incident. Being the first of the two marginal DNL algorithms developed in this thesis, it tends somewhat towards an efficient post-processing method rather than a 'pure' marginal algorithm in the philosophy of Section 7.2. This leads to larger approximation errors compared to the later developed MaC, but also to an even greater computational gain. Therefore, MIC can still be more appropriate for large-scale, coarser or quick-scan studies where a large number of scenarios are more important than detailed results for each specific case (such as robust network design and vulnerability to incidents).

Section 8.1 discusses the MIC algorithm. Afterwards, a vulnerability analysis of the Sioux Falls benchmark network is presented (Section 8.2), comparing the results of MIC simulations to explicit simulations with LTM. The aim of this case study is not to present a profound vulnerability analysis, but rather to evaluate MIC's performance in terms of accuracy and computation time.

This chapter is an edited version of Corthout, R., Tampère, C.M.J. & Immers, L.H. (2009). Marginal Incident Computation, an Efficient Algorithm to Determine Congestion Spillback due to Incidents. *Transportation Research Record, 2099, pp. 22-29.*

8.1 MIC outline

MIC performs marginal simulations that approximate the congestion effects caused by an incident by superimposing it onto a base simulation (without incidents). This is done by only calculating the additional congestion on the affected links upstream of the incident (not downstream; see Section 8.1.4.2). MIC can be applied for incidents of varying severity (from a complete blocking to no effect on the capacity) and to any base scenario.
Compared to explicit simulation – which means re-evaluating the entire network and time period for each incident case – this provides considerable computation time savings when a large set of incident cases is to be simulated (e.g. in vulnerability analysis or robust network design). While the explicit approach involves many calculations largely identical to the base situation (e.g. prior to the incident, far away from the incident), MIC reduces the computational effort to a fraction of all links during a fraction of all time intervals.

Section 8.1.1 discusses the required input, followed by a description of the working of MIC in general terms in Section 8.1.2. Afterwards, in Section 8.1.3, the details of the algorithm are explained. Finally, Section 8.1.4 provides more insight into the approximation errors of MIC compared to explicit simulation.

8.1.1 Required input

First, a full, explicit DNL simulation is to be performed of the base scenario, on which incident cases will be superimposed. This is called the base simulation, obtained from the base model. The maternal base model of MIC is LTM. As such, MIC applies basically the same intersection model (that of Chapter 4), and the link model is based on first-order traffic flow theory with a triangular fundamental diagram as in LTM. In principle, any existing DNL model[40] could provide the base simulation for MIC. However, this is likely to increase the approximation errors as the discrepancy with the model used for the base simulation grows.
Then, the network is loaded into MIC. Also the base Cumulative Vehicle Numbers (CVN) are required as input to MIC. The total CVN at upstream and downstream link boundaries, as well as the CVN separated per turning direction at the downstream link boundaries – equivalently, turning fractions could have been used for the same purpose – are needed. Of course, also the characteristics of each incident have to be specified, i.e. incident location, starting time, duration and severity (a reduction factor that is to be applied to the original capacity). For convenience, it is assumed that an incident always takes place at either the upstream or downstream boundary of a link, where the CVN are available. (If required, any location within a link could be chosen. Indeed, Newell's theory (1993) allows constructing CVN at any location within a link, which could subsequently serve as a reference for the algorithm.)

[40] This would, however, require transforming the outputs of this model into CVN at link boundaries before MIC can be applied.

8.1.2 MIC procedure

Each incident is consecutively evaluated. An incident is imposed by reducing the capacity at a link boundary, according to the incident characteristics specified in the input. This forms the local variation to the base scenario. This variation causes changes that will be calculated and tracked by MIC. More specifically, MIC performs an approximated DNL of the influence zone of an incident. The influence zone is a set of affected links, starting from the incident location and growing in the upstream direction. The affected links are links on which the incident spillback wave imposes a constraint on the original, base link flow and thus causes (extra) congestion. New calculations are only carried out for the affected links. Since in large networks the influence zone of an incident may only cover a few percent of the entire network, a significant reduction of computation time is possible. Hereby, drivers are assumed to make the same trip (no changes in departure time, destination or route) in case of an incident as they make in the base situation. Rerouting effects that can occur after an incident are thus neglected in MIC. In Appendix G, the enhanced possibilities due to the inclusion of an en-route rerouting model (in the later developed MaC) are demonstrated.

8.1.3 MIC algorithm

The MIC algorithm described here modifies the base CVN at the upstream and downstream boundaries of affected links in accordance with the flow constraint imposed by the incident. This procedure is repeated for every incident case.

From the incident location on, the congestion spillback wave due to the incident, as well as the acceleration wave of the discharging queue after the incident is cleared, are tracked upstream. We refer to the constraint that this queue propagation and dissipation imposes on the flows on a link as the incident flow constraint.
The main MIC algorithm employs two submodels: a link and an intersection model. These are approximations of the link and intersection model embedded in the maternal base model LTM. Using the base CVN, these models calculate the new CVN for the incident scenario for every affected link. This is not done by dividing the simulation into time intervals and updating the CVN of the links in the network chronologically per time interval. Instead, entire CVN curves at link boundaries are calculated at once, without considering or updating non-adjacent parts of the network. By doing so, MIC provides a coarse representation of traffic flow and congestion dynamics, but at a significantly reduced computational cost. In summary, when a link becomes affected, the link and intersection model are run once[41]. The simplified link model passes the incident flow constraint from the downstream to the upstream boundary of an affected link. If the incident spillback reaches the upstream boundary of the link, the incident flow constraint is passed on to the upstream intersection. Then, the intersection model subjects the incoming links of that intersection to the incident flow constraint. If the incident flow constraint is restrictive to the flow of an incoming link, this link becomes affected by the incident and thus the influence zone expands. When for all remaining affected links the incident spillback does not reach the upstream link boundary anymore, the simulation is finished. The output, namely the changed CVN of the affected

[41] This means that gridlock (whether realistic or model induced) cannot occur in MICs, since each affected link is only considered once.

links, is stored. Other output such as the VHL due to an incident and travel times can later be derived from these CVN. All variables are restored to the base values so that the analysis of the next incident can start from the same base input.
The link and intersection model of the MIC model are described below.

8.1.3.1 Link model

The total CVN of the base simulation (represented by the dashed line in Figure 8-1) are reduced to fit the incident flow constraint on the affected link. Since MIC is SC, only the total CVN need to be considered, not CVN separated by route as in LTM.

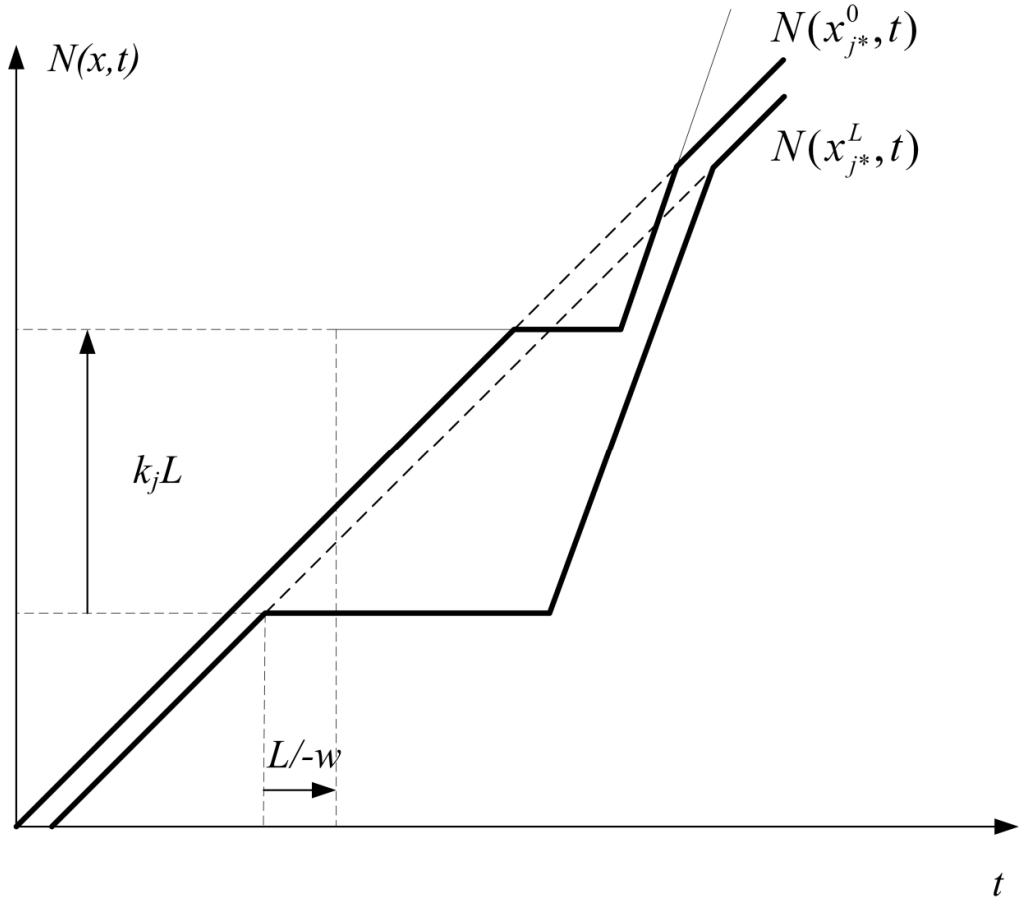

Figure 8-1: The incident flow constraint is passed on to the upstream link boundary

First, the downstream CVN are lowered according to the incident flow constraint. The incident flow constraint is visible in Figure 8-1 as the slope of the altered cumulative curve (see also Figure 8-2 for a more detailed example). Then, the changes in the downstream curve are transferred to the upstream link boundary. This is done, as in LTM, according to the simplified CVN procedure of Newell (1993). The incident flow constraint is propagated to the upstream link boundary by shifting it $L/-w$ in time and k_jL in vehicle numbers (L being the length of the link, w the negative spillback wave speed and k_j the jam density). See Appendix A for more details on this procedure. If the incident flow constraint intersects with the

upstream base curve, these CVN are lowered as well. The incident flow constraint at the upstream link boundary is then input to the upstream intersection.

8.1.3.2 Intersection model

The intersection model in MIC is essentially equivalent to that of Chapter 4. Therefore, instead of presenting the entire model here, this discussion is limited to the differences with the model in Chapter 4. The main difference is that the model is run in two distinct phases, as explained below.

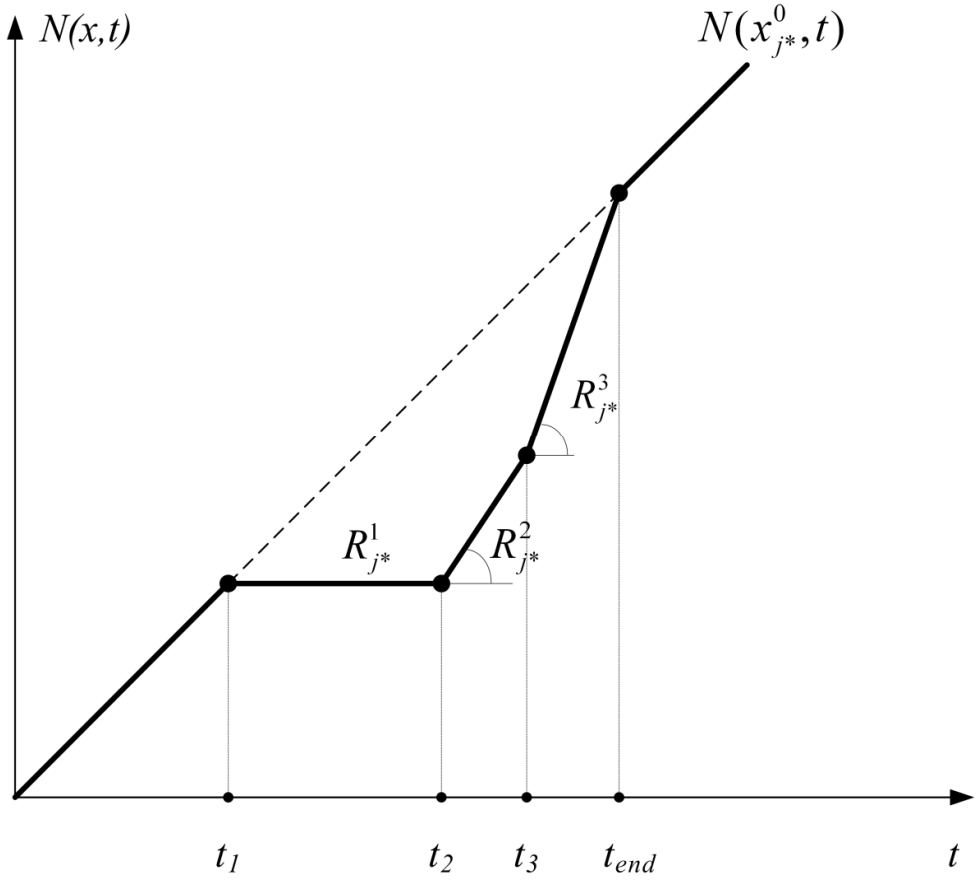

Figure 8-2: The incident flow constraint at the upstream boundary of link *j forms the supply of R_{j*}**

If the incident flow constraint on an affected link *j** reaches the upstream intersection, it serves as the supply R_{j*} (i.e. the maximum flow that can enter *j**) to be distributed among the incoming links *i*. This constraint is represented by the slope of the CVN curve at the upstream boundary of j^* (see Figure 8-2), which has been determined in the link model. This constraint is split up into two phases. The first phase, the queue propagation phase, starts when the incident congestion reaches the intersection (at t_1). The second phase, starting when the congestion begins to dissolve (at t_2), is the queue dissipation phase, which can be split up further into *k* sub-phases. (In Figure 8-2 this second phase consists of two sub phases.) This fragmentation of the second phase results from the fact that queues do not dissolve at the

same time on all incoming links of an intersection (see the explanation of the queue dissipation phase).

When calculating the flows of the incoming links of the intersection, a distinction is made between the two mentioned phases.

1. Queue propagation:
   First, the average base flows during the queue propagation phase (between t_1 and t_2) from each link i to link j^* are calculated from the base CVN. This is equivalent to following the base turning fractions; hence, the SC nature is clear. The total base flow of link i during the queue propagation phase is denoted as $q^1_{i,base}$; the partial base flow from i to j as $q^1_{ij,base}$. From this, the turning fractions to be used by the intersection model are calculated:

$$f_{ij} = \frac{q^1_{ij,base}}{q^1_{i,base}} \quad \forall i, j \tag{8.1}$$

   The supply $R^1_{j^*}$ of the affected link is then distributed among the incoming links i proportional to oriented capacities (see Section 4.2) derived from the capacities C_i and the turning fractions f_{ij^*} (from (8.1)). The supply constraints from other outgoing links j are not yet considered in this phase since they have (probably) not changed compared to the base scenario. Also, instead of the demand constraints S_i that are normally considered in an intersection model, the base flows $q^1_{i,base}$ apply as constraints:

$$q^1_{i,MIC} \leq q^1_{i,base} \tag{8.2}$$

   In other words, in the queue propagation phase flows cannot exceed the base flows, as this is not to be expected as a result of an incident.
   Then, the intersection model calculates the new flows $q^1_{i,MIC}$ during the queue propagation phase. If $q^1_{i,MIC} < q^1_{i,base}$, link i becomes affected by the incident. Otherwise, the flow on i is not restricted by the incident flow constraint and thus $q^1_{i,MIC} = q^1_{i,base}$ is assumed (i is not affected).

2. Queue dissipation:
   The flows of unaffected links i ($q^1_{i,MIC} = q^1_{i,base}$) are assumed to remain unchanged also in this phase ($q^k_{i,MIC} = q^1_{i,base} \quad \forall k$). This is a simplification to increase the computational speed. This leaves only the flows $q^k_{i,MIC}$ of the affected incoming links to be calculated, i.e. those for which a reduced outflow due to the incident flow constraint was found in phase 1 ($q^1_{i,MIC} < q^1_{i,base}$).

As the supply value of R_{j*} changes, calculations are carried out consecutively for each sub phase k (with k starting at 2). Now, however, the supply constraints of all outgoing links j are taken into account, since in this phase the flows are no longer bound to their base values. The increased flows that result from the queues discharging may activate any outgoing link's supply constraint. For link j^*, R_{j*}^k is the incident flow constraint in sub phase k (see Figure 8-2). For all other j, R_j^k is assumed to be equal to the capacity C_j. In the queue dissipation phase, the flows may be higher than the base flows, but never higher than the link capacities. Hence, instead of (8.2) holds that:

$$q_{i,MIC}^k \le C_i \quad \forall k \tag{8.3}$$

The intersection model then determines the flows $q_{i,MIC}^k$ via an oriented capacity proportional distribution - using f_{ij} as in (8.1) - as in Section 4.2. Note that the turning fractions of phase 1 are maintained throughout the queue dissipation phase. This simplifying assumption is further explained in Section 8.1.4.1.
Once the flows during sub phase k are calculated, the smallest time t_{k+} (before the end of phase k) is sought for which the estimated new CVN ($N'_{MIC}(x_i^L, t_{k+})$) at t_{k+} are equal to the base CVN (and would exceed the base CVN for $t > t_{k+}$):

$$\begin{aligned} &N'_{MIC}(x_i^L, t_{k+}) = N_{base}(x_i^L, t_{k+}) \\ &N'_{MIC}(x_i^L, t_{k+} + \varepsilon) > N_{base}(x_i^L, t_{k+} + \varepsilon) \\ &with: \quad N'_{MIC}(x_i^L, t_{k+}) = N_{MIC}(x_i^L, t_k) + q_{i,MIC}^k (t_{k+} - t_k) \\ &\qquad t_k \le t_{k+} \le t_{k+1} \end{aligned} \tag{8.4}$$

In (8.4), $N_{MIC}(x_i^L, t_k)$ is known from the flows in earlier sub phases. If such a t_{k+} is found, this means that the incident congestion dissolves (the base CVN are reached again) on the corresponding link(s) i at t_{k+}. For this particular link(s) i, the flow is assumed equal to the base flow for $t > t_{k+}$. Since this may also affect the flows from other incoming links i', sub phase k is split up into two sub phases at t_{k+} and all flows $q_{i,MIC}^k$ (between t_k and t_{k+}) are stored. Then, the intersection model calculates the flows for the remaining affected links i' for the new sub phase $k + 1$, starting at t_{k+}. This is how the multiple-phased queue dissipation part of the ICF originates (see Figure 8-3).

The flows $q_{i,MIC}^k$ calculated in each (sub) phase by the intersection model form the incident flow constraint at the downstream boundaries of the affected incoming links. For each of these links, the link model is run to store the new CVN and check whether or not the incident congestion reaches the upstream intersection, in which case the intersection model is called again. This alternation between the link and intersection model continues until there are no more affected links on which the incident flow constraint reaches the upstream boundary.

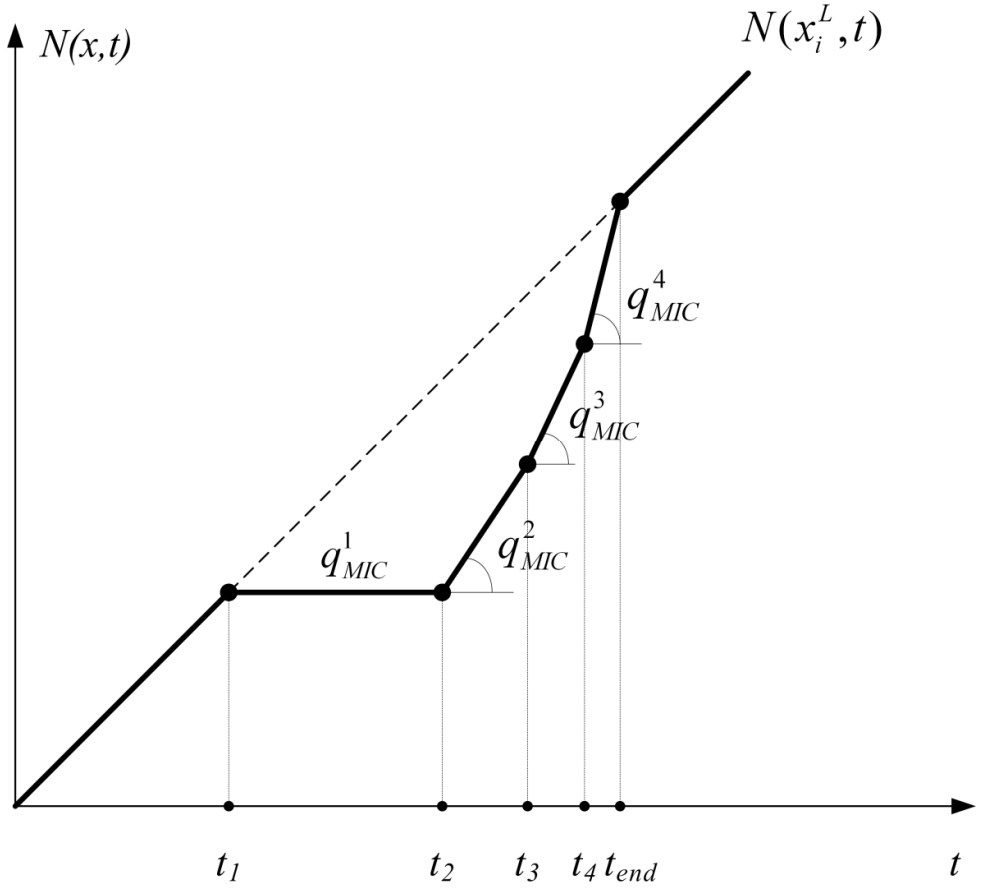

Figure 8-3: Multiple-phased queue dissipation

8.1.4 Sources of error

Compared to MaC (Chapter 9), the discrepancy between MIC and the base model LTM is quite large. As a result, MIC provides a rougher approximation. However, since MIC only calculates upstream moving effects, the spatial propagation of these errors is quite limited. The simplifications that are made in the MIC algorithm are elaborated on below.

8.1.4.1 Averaging demand

The intersection model determines the reduced flows of the incoming links using the base flows and turning fractions averaged between t_1 and t_2. Small, high-frequency fluctuations in these flows are ignored and will not appear in MIC. Also the flows in phase 2 (queue dissipation phase) are determined using the average turning fractions between t_1 and t_2 (8.1). Changes in the turning fractions after t_2 are discarded. However, due to the encountered delay, an important part of the traffic flows in the queue dissipation phase consist of vehicles that passed between t_1 and t_2 in the base situation. Moreover, the turning fractions usually do not change drastically in a short amount of time. Another simplifying assumption is that links that are unaffected during the queue propagation phase (when the incident flow constraint is

most severe) remain unaffected during the queue dissipation phase. Changes in these flows in this second phase are thus not taken into account.

8.1.4.2 Neglecting downstream effects

MIC only tracks changes in the upstream direction. Links downstream of the incident are not examined. Usually, only a shift in the CVN is expected there, not additional delay. Possibly, however, secondary effects such as relieving or overloading of downstream bottlenecks may occur. An example is provided in Figure 8-4. Say that link 1 has insufficient capacity, causing a queue to form on link 2 (a). Now, if an incident should occur on link 3, a downstream wave with a lower flow rate originates (b). This relieves the bottleneck, enabling a higher flow from link 2 to link 1. When the incident is cleared and the congestion starts dissolving (c) a downstream wave at capacity originates (high flow rate). This overloads the bottleneck and thus restricts the flow from link 2 to link 1, so that again a (possibly even longer) queue forms on link 2. Effects like this are not taken into account by MIC.

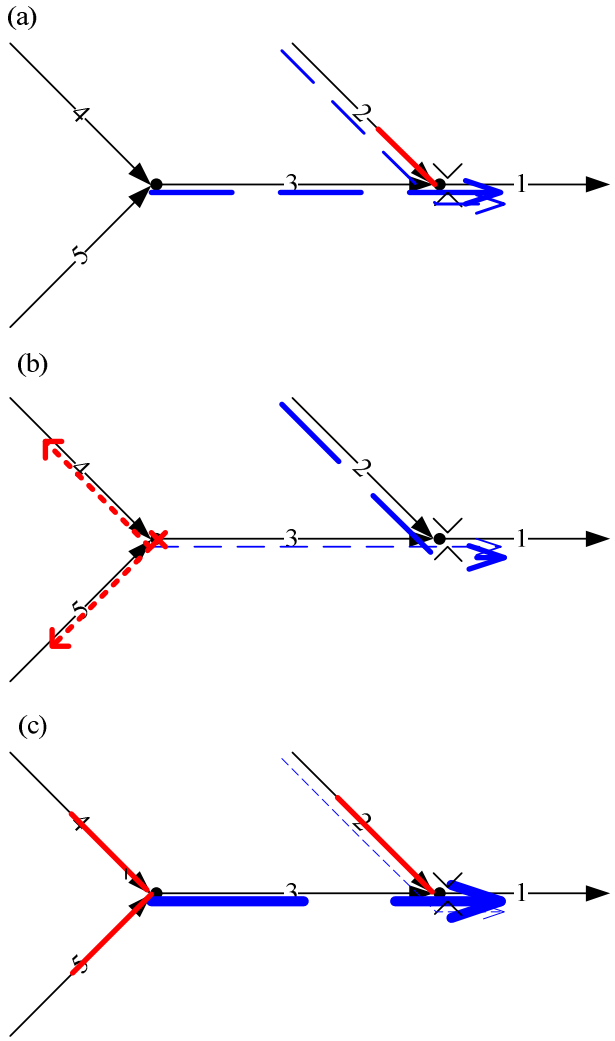

Figure 8-4: Downstream effects of an incident

8.1.4.3 Neglecting delayed spillback

The neglected downstream effects might not only influence other queues and bottlenecks such as in Figure 8-4, but also the incident queue that MIC aims to simulate. Indeed, when the incident flow constraint reaches an intersection via an affected link *j**, it (usually) decreases the flow over this intersection in several directions *j*. This may cause 'delayed spillback' of the incident flow constraint on other downstream affected links. This is not accounted for in MIC. Delayed spillback occurs if multiple spillback waves travel towards the diverge n via different routes, which is illustrated in Figure 8-5. If an incident occurs on link 1, congestion spills back via link 2 and link 3. The spillback wave via link 2 reaches n first, thus decreasing the flow into link 4. This causes the spillback wave on link 4 to slow down (or even to be reversed), since there is less flow to build up the queue.

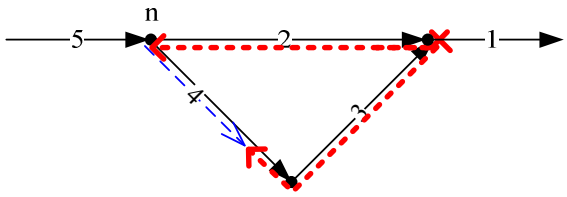

Figure 8-5: Delayed spillback

By neglecting the delayed spillback effect in MIC, the arrival time of the second spillback wave is not correctly calculated. Tests have shown that this can lead to a significant error on isolated circuits such as in Figure 8-5, while on a realistic traffic network notable errors are expected to be quite rare. Because of this approximation, multiple spillback waves over the same affected link are not simulated, meaning that gridlock effects are implicitly neglected.

8.1.4.4 Neglecting up- and downstream bottlenecks during queue dissipation

As explained in Section 8.1.3, the entire new flow profile of an affected link is calculated in one run of the link and intersection model. MIC simulations are thus ordered by link instead of by time. Since the simulations are not entirely chronological, it is not possible to account for interactions other than those with adjacent links. This can cause errors mainly in the queue dissipation phase of the intersection model. Indeed, since only the capacities of the links of that intersection are considered, bottlenecks further up- or downstream that could limit an incoming link's flow or an outgoing link's supply are disregarded. Figure 8-6 and Figure 8-7 illustrate this simplification on two simple linear networks.

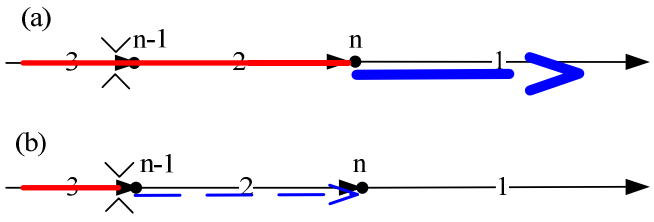

Figure 8-6: Upstream bottleneck

In Figure 8-6 (a), a queue has formed on links 2 and 3. The outflow of link 2 is calculated considering only C_1 and C_2 (i.e. of the adjacent links of n), yielding a high flow rate. However, after the queue on link 2 has dissolved, the queue on link 3 will dissolve at a lower flow rate C_3 - the capacity of upstream bottleneck link 3 (b). From then on, for vehicles that started queuing on link 3, the queue dissipation rate of link 2 is overestimated. Link 3 does not suffer from this overestimation, since at n-1 the bottleneck capacity C_3 is accounted for.

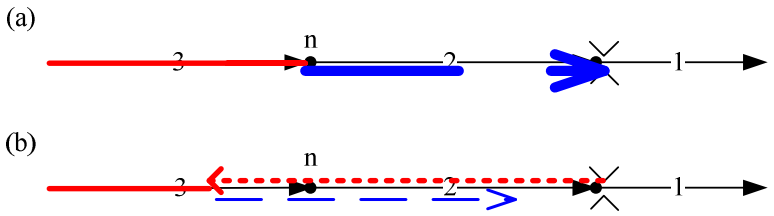

Figure 8-7: Downstream bottleneck

Consider on the other hand a queue dissolving from link 3, as in Figure 8-7 (a). As the downstream discharge wave (at high flow rate) reaches the downstream bottleneck link 1, a new congestion wave will spill back, limiting the flow rate on link 3 to C_1 (b).

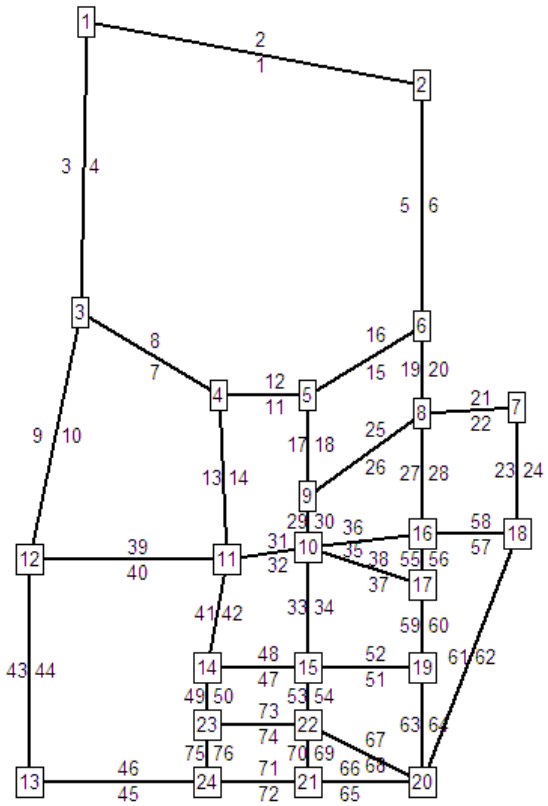

Figure 8-8: Sioux Falls network

8.2 Case study Sioux Falls network

A case study of the Sioux Falls network is presented to test the performance of MIC in comparison to explicit simulation of incidents using LTM. The Sioux Falls network (Figure 8-8) consists of 76 links and 1752 predefined routes between 24 origins and destinations.
The simulation time period is four hours. 76 incident scenarios are simulated, one for each link of the network. An incident completely blocks the link by reducing its capacity at the upstream link boundary to zero. All incidents occur after one hour into the simulation and last one hour, after which the link capacity is restored to the original value. The traffic demand is chosen considerably low (all links well below capacity) to avoid gridlock situations in the explicit simulations[42].

MIC is evaluated on a rather aggregate level, by comparing the vehicle hours that are lost (VHL) on the affected links of the network due to each incident (compared to the base scenario) in MIC and LTM. This suits the intended scope of MIC best, namely to provide a somewhat rough gauge of the impact of incidents for a very low computational cost, not to produce highly accurate results (e.g. in terms of traffic flows). The VHL are calculated from the base and incident CVN for every affected link and added up to determine the total VHL due to an incident, identifying the most vulnerable links. Table 8-1 shows that the same vulnerable links are found by MIC and by explicit simulation, with limited deviation in VHL.

Table 8-1: Vehicle hours lost on most vulnerable links

Blocked link	VHL exp (h)	VHL MIC (h)	dev (%)
37	1289.4	1295.5	0.5
35	1189.2	1194.8	0.5
33	1175.8	1179.7	0.3
30	1156.9	1160.4	0.3
59	1141.3	1139.7	-0.1
38	1122.2	1123.2	0.1
28	1056.9	1018.2	-3.7
57	996.8	946.0	-5.1
31	989.8	989.4	0.0
55	928.5	926.1	-0.3

Of course, not only the global effect of an incident is important, but also how the delay is distributed over the road users. A correct distribution of the VHL over the various affected links is therefore required. To show that MIC satisfies this requirement, the results of the most severe incident (on link 37) are presented in detail in
Table 8-2. (Links 85, 86 and 87 are origin links not depicted in Figure 8-8.) MIC identifies the same set of affected links as the explicit simulation with LTM.
Table 8-2 shows that the total deviation in VHL is distributed quite evenly over the affected links and does not result from overestimation on some links being compensated by

[42] Which traffic demand levels are chosen is not that important, since the case study only serves as a comparison between the performance of MIC and explicit simulation; and MIC can be applied to any base reference case.

underestimation on other links. Time instants t_1 and t_{end} (see Figure 8-2) for the downstream link boundaries are presented, showing that the maximum deviation between MIC and LTM is two minutes (for link 32).

Table 8-2: Results of incident scenario on link 37

Affected link	Explicit (h)			MIC (h)			dev (%)		
	VHL	t_{start}	t_{end}	VHL	t_{start}	t_{end}	VHL	t_{start}	t_{end}
29	143.5	1.10	2.22	143.4	1.10	2.22	0.0	0.0	0.0
32	236.0	1.10	2.34	236.5	1.10	2.30	0.2	0.0	-1.7
34	168.2	1.10	2.22	168.3	1.10	2.21	0.1	0.0	-0.5
36	159.7	1.10	2.21	159.7	1.10	2.20	0.0	0.0	-0.5
86	419.3	1.10	2.29	419.4	1.10	2.29	0.0	0.0	0.0
17	10.9	1.81	2.20	11.6	1.81	2.20	6.0	0.0	0.0
25	7.5	1.81	2.20	7.5	1.81	2.20	0.0	0.0	0.0
85	25.2	1.81	2.20	25.2	1.81	2.20	-0.1	0.0	0.0
13	16.9	1.84	2.29	17.1	1.84	2.28	1.1	0.0	-0.4
40	26.0	1.84	2.28	25.5	1.84	2.27	-1.8	0.0	-0.4
42	37.8	1.84	2.32	43.5	1.84	2.35	15.2	0.0	1.3
87	38.5	1.84	2.28	37.8	1.84	2.27	-1.7	0.0	-0.4

This case study shows that the results from MIC and explicit simulation match well. The weighted average of the absolute value of deviation between the VHL calculated by MIC and by explicit simulation is only 0.9 %. Comparison of computation times illustrates that MIC is highly efficient. Simulation of the 76 incident scenarios on the Sioux Falls network takes about 20 minutes[43] for LTM versus 11 seconds with MIC. This means that for each explicit simulation of one incident on a link, MIC could perform a set of over 100 incidents. Moreover, this computational advantage increases with the size of the network (assuming that the relative size of the affected area compared to the whole network reduces with increasing network size).

8.3 Conclusion

This chapter presents MIC, a marginal DNL algorithm designed for fast Monte-Carlo simulation of incidents on road networks. MIC applies similar link and intersection models as its maternal base model LTM. Hence, it adopts a realistic representation of queue propagation and congestion spillback, consistent with first order traffic flow theory. The MIC algorithm superimposes the congestion effects caused by an incident onto the results of a base simulation run. The CVN are altered, according to the congestion resulting from the incident. Calculations are only carried out for the affected links, not for the entire network. A case study of the Sioux Falls network shows that MIC approximates the results (in terms of VHL)

[43] This computation time is an estimate for how long it would take for the more recent version of the LTM code that is used later in the comparative case study with MaC in Chapter 9. With the older, less efficient code that was used for this case study, the actual computation time was 3 hours and 32 minutes.

of sequential explicit simulation well. A significant computational advantage is achieved compared to explicit simulation, where countless identical traffic flows (e.g. prior to or far away from the incident) are recalculated. In large networks, the computation time can be reduced to less than 1 % of the explicit simulation time. Due to this computational advantage, calculation time can instead be devoted to fine sampling of incident scenarios, using extensive data on incident occurrence, duration and severity. Possible applications (whether or not as a quick scan tool) include identifying vulnerable links, real-time incident management and TTV studies (see the proof-of-concept case study of Appendix F.2). Moreover, MIC has been effectively applied for robust network design by Snelder (2010).

However, it should be noted that MIC is not really a marginal derivation of the maternal base model LTM in the philosophy of Section 7.2. Rather, it is an efficient method based on the same modelling principles as LTM. As such, the discrepancy between MIC and LTM is considerable. Therefore, it exhibits larger approximation errors than its successor MaC, leaving it less suited for detailed modelling (e.g. analysis of link flows) and less widely applicable. The need for a different approach, which lead to MaC, stems mainly from the fact that the MIC methodology proved not to be extendable to the modelling of demand variations. Still, the development, evaluation and (unsuccessful) attempts to extend MIC contributed significantly to our understanding of marginal simulation. Also, thanks to its computational speed it may still be preferred in large-scale applications that require only a somewhat coarse estimation of the impact of incidents (e.g. in terms of VHL).

9
MARGINAL COMPUTATION (MAC)

The second marginal DNL algorithm, Marginal Computation (MaC), follows the general concept of marginal simulation as discussed in Section 7.2 more closely than MIC. As a consequence, MaC is much more versatile. Foremost, it can simulate both supply and demand variations. This significantly broadens its applicability. Furthermore, the approximation errors are reduced because MaC is algorithmically much closer to its maternal base model (LTM) than MIC. While slightly lower than that of MIC, the computational gain of MaC compared to explicit simulation is still considerable.

Section 9.1 introduces MaC. Then, Section 9.2 describes a case study on a medium-scale network (around Ghent, Belgium) to illustrate the performance of MaC compared to explicit simulation. Section 9.3 provides a sensitivity analysis to several parameters. We also refer to Appendix G, in which a proof-of-concept implementation of an en-route rerouting module in MaC is presented. This demonstrates an important future research direction for marginal simulation, which will enhance their realism and broaden their applicability.

This chapter is an edited version of Corthout, R., Tampère, C.M.J., Frederix, R. & Immers, L.H. (forthcoming). Improving the efficiency of dynamic network loading applications through marginal simulation. Submitted to *Transportation Research Part C*.

9.1 MaC outline

MaC is a marginal DNL simulation algorithm that approximates full, explicit DNL simulations in the form of partial (marginal) simulations. It only calculates the traffic flows and states that change due to a local variation (in demand or supply) to a base simulation. Hence, the computational efficiency is substantially increased; enabling applications that are infeasible with an explicit DNL simulation approach. For examples of such applications, we refer back to Section 7.2. Moreover, while the existing approximate methodologies discussed in that section – and this is equally true for MIC (presented in Chapter 8) - typically exploit some of the specific features of the problem at hand, rendering them not generally applicable, MaC can be used for a wide range of applications.

The modelling assumptions and simulation algorithm of MaC are directly adopted from the maternal base model LTM. Only the MC representation of traffic flows is replaced by a computationally more efficient SC approach. Naturally, some additions have to be made in MaC to detect and track the changes to the base simulation in each marginal sample. How MaC is derived from LTM is explained in Section 9.1.1. Note that marginal DNL algorithms could be derived from other existing DNL models analogously. Then, Section 9.1.2 discusses possible sources of approximation errors.

9.1.1 MaC algorithm

MaC performs marginal DNL simulations as variations to a base simulation. For each variation, calculations are only performed for the active part of the network and simulation period, i.e. that part where the traffic flows (are expected to) differ from the base flows. The links on which different flows indeed occur are the affected links, constituting the affected area. MaC can simulate variations in both demand and supply (i.e. link capacities) to the base input.

First, the MaC procedure is presented in a stepwise scheme. Afterwards, each of these steps is explained in detail. Although the explanations further on specifically explain how MaC is derived from LTM, the general stepwise method below can be followed to derive marginal algorithms from other existing DNL models.

(0. Run base simulation and store state variables)
1. Read input; initialize variation $v = 1$
2. Impose change according to variation v to base input variables
3. Activate part of network and time period for simulation
4. Simulate traffic propagation in active part of network
   Check if affected area grows and if so, activate additional part of network
5. Post-process:
   Write output
   Restore base input variables
   Set $v = v + 1$ and return to step 2

9.1.1.1 Read input

MaC requires the following input:

- Base input:
  By base input, we mean the traditional input that also the base DNL model (in this case LTM) requires. This consists of the network characteristics, dynamic OD matrix and route choice. These are stored in MaC as the base input variables.

- Characteristics of variations:
  For each (demand or supply) variation, it needs to be known which of the base input variables change and to what extent. In case of a supply variation (typically an incident on a link), the incident location, starting and ending time and the fraction of the base capacity that remains are specified. For demand variations, the OD-pair(s) or route(s) on which the demand in- or decreases are needed, as well as starting and ending time of the change and the fraction by which the base demand changes. Currently, variations to other input variables (e.g. free flow speed or jam density on links) are not considered.
  In the remainder of this chapter, we focus on demand variations since these are more intricate to model than supply variations.

- Information from base simulation:
  From the base simulation results, the following information is derived as input to MaC:

  - Total CVN at upstream and downstream link boundaries:
    This is a direct output of the base simulation run with LTM. These curves represent the total traffic flows (not disaggregated by route) in the base simulation. Also the traffic states and travel times of the base scenario are given by the CVN curves.

  - Turning fractions at downstream link boundaries:
    Since MaC is SC, the turning fractions are needed as inputs to propagate traffic through the network. The turning fractions are dynamic, or more precisely: piecewise constant with a fixed turning fraction interval T.

  - Numerical approximation of the dependency of the base turning fractions on the base demand:
    To simulate demand variations with MaC, it needs to be known how the turning fractions in the network change with a variation in demand. The following is limited to a brief explanation. More details are provided in Appendix B and, e.g., Blumberg & Bar-Gera (2009); since also in SC DNL models this derivation is needed to construct the turning fractions from the route demands and travel times.
    Let us denote the arrival time at the downstream boundary of link a by t. The travel time from the origin of a route r to the end of link a is then written as

$tt_{r,a}(t)$. Then, the dynamic turning fractions $f_{ab}(t)$ from a to downstream links b are a function of the dynamic route demand profile $q_r(t-tt_{r,a}(t))$ for each route r that passes a ($a \in r$):

$$f_{ab}(t) = \frac{\sum_{r' | a,b \in r'} q_{r'}(t - tt_{r',a}(t))}{\sum_{r | a \in r} q_r(t - tt_{r,a}(t))} \tag{9.1}$$

Not only do the route demands q_r directly influence f_{ab}, also the travel times $tt_{r,a}$ depend on q_r. However, we neglect the dependency of $tt_{r,a}$ to q_r, so that an approximation is obtained, using the demands and travel times from the base simulation. Hereby, again a numerical discretization with interval T is used. We refer to Section 9.1.2 for an elaboration on the error this approximation induces.

This information allows MaC to propagate a demand variation through the network by adapting the turning fractions along the way, as is explained in the next section.

9.1.1.2 Impose change

According to the characteristics of the variation, the proper change(s) are made to the base input variables. For a supply variation, this implies for example to reduce a link's capacity for the duration of an incident. A demand variation is imposed by changing the demand at the origin of a certain route or OD pair. Consequently, the turning fractions along the route must be updated to guide the demand variation through the network to its destination. For example, consider the network depicted in Figure 9-1. A variation is considered in the form of an increased route demand $q_r(t^0)$ with $1..6 \in r$.

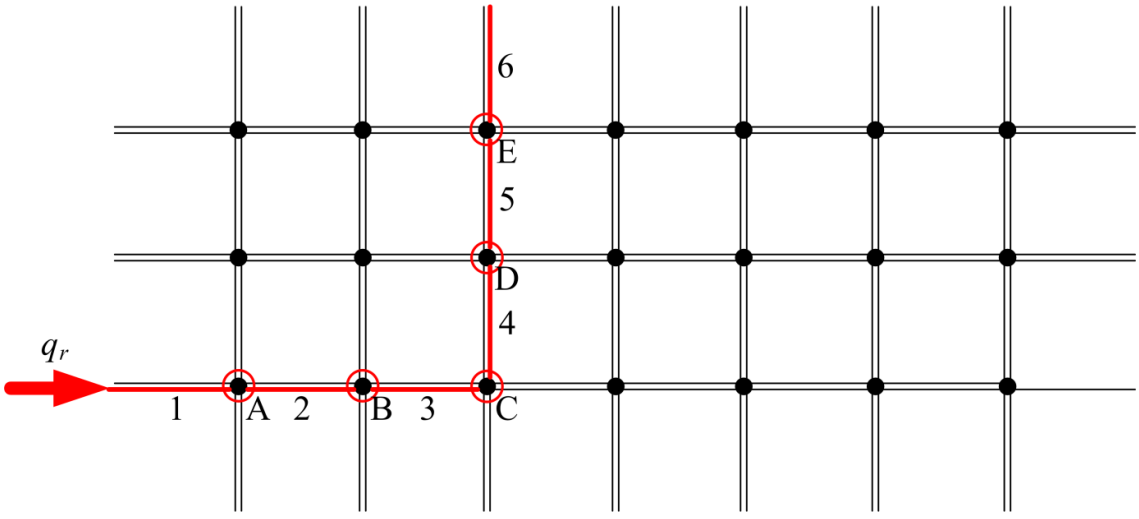

Figure 9-1: Demand variation - imposing changes and initial activation

The turning fractions $f_{ab}(t^0 + tt^0_{r,a})$ are updated by recalculating equation (9.1) with the altered demand $q_r(t^0)$ and the base demand for other routes *r'*. This is done for links a = 1..5 with $tt^0_{r,a}$ being the base travel time for route r until the end of link a for a vehicle that departed at t^0. For example, the increased route demand q_r causes f_{34} to be increased from $t^0 + tt^0_{r,3}$ on, so that more traffic on link 3 turns left towards link 4. An approximation lies in the fact that all travel times are hereby assumed not to differ from the base value (see Section 9.1.2).

9.1.1.3 Activate part of network

The marginal simulation of the current variation is initialized by activating that part of the network where changes to the input variables, e.g. a decreased link capacity or altered turning fractions, have been imposed. Since simulations in MaC (as in LTM) are governed by intersection updates, it is the intersections that need to be activated. For a supply variation this would imply that the intersection adjacent to a changed link capacity is to be activated. Thus, if the ingoing capacity of a link is changed, the upstream intersection of this link is activated; a change in outgoing capacity activates the downstream intersection. For the case of a demand variation we refer again to the example in Figure 9-1. There, all the intersections n = A..E that are traversed by the changed route flow on r are activated from t_n on. The intersection activation time t_n is the base arrival time at intersection n (i.e. the downstream boundary of link a) of the first vehicle of the increased route demand $q_r(t^0)$. This coincides with the time $t^0 + tt^0_{r,a}$ at which the turning fractions at that intersection were altered in the previous step. Also, the status of the links of route r is set to 'affected'.
Intersections, once activated, are not deactivated in MaC before the simulation of the current variation is finished. This would necessitate introducing additional checks to detect the end of the change. This may constitute a future improvement.

9.1.1.4 Simulate

In the previous step, some of the intersections in the network have been activated. For this active part of the network, intersection updates are performed to calculate the flows crossing the intersection from each incoming link to each outgoing link (both affected and unaffected). In each intersection update, MaC derives the local demand and supply at the intersection level from the CVN at the adjacent link boundaries in the same way LTM does (see Appendix A). For affected links, this information is derived from the new CVN inherent to this variation; for unaffected links from the base CVN. The only difference with LTM is that the MC approach is replaced by a computationally more efficient SC representation. As explained earlier, an approximation lies in the fact that the turning fractions in MaC are in correspondence to the base travel times and not the travel times experienced in MaC. (This means that the iterations to achieve consistency in the DNL simulation as explained in Appendix B are omitted.)
At each intersection update, the calculated flows of the affected links are used to update the new CVN. For unaffected links, the calculated flows are compared to the base flows of the

same time step. Hereby an accuracy threshold ε is used in the check so that very small changes are not needlessly tracked. If $\frac{|q_{a,MaC} - q_{a,base}|}{q_{a,base}} > \varepsilon$ for a certain link a, this link becomes affected. The threshold ε is defined as a percentage. Depending on the application, a definition of ε in absolute rather than relative terms might be more appropriate. For the case study presented in this paper, there was no significant difference in performance using absolute or relative thresholds. A different threshold is used for links that are incoming to the intersection (ε_{up}; controlling changes that propagate upstream) and outgoing links (ε_{down}; for downstream changes). Caution is needed to select a proper value for these thresholds. Particularly if ε_{down} is set too low, insignificant changes (that propagate downstream fast) could activate large parts of the network and thus reduce MaC's efficiency. Different values for ε will be compared in Section 9.3.

If a difference in flow is detected and a link becomes affected, the active part of the network needs to be expanded. For incoming links, the upstream intersection of this link a is activated (if currently inactive) at $t^c + \frac{L_a}{-w_a}$ to allow the change to propagate upstream. Herein, t^c is the current time, L_a the length of link a and w_a the maximum spillback speed from the link's triangular fundamental diagram. Hence, $t^c + \frac{L_a}{-w_a}$ is the earliest time the change could reach the upstream intersection (see Appendix A). Analogously, for outgoing affected links the downstream intersection is activated at $t^c + \frac{L_a}{v_{f,a}}$, with $v_{f,a}$ being the (maximum) free flow speed on the link.

To illustrate how the affected area and the active part of the network can expand, consider again the example network (see Figure 9-2).

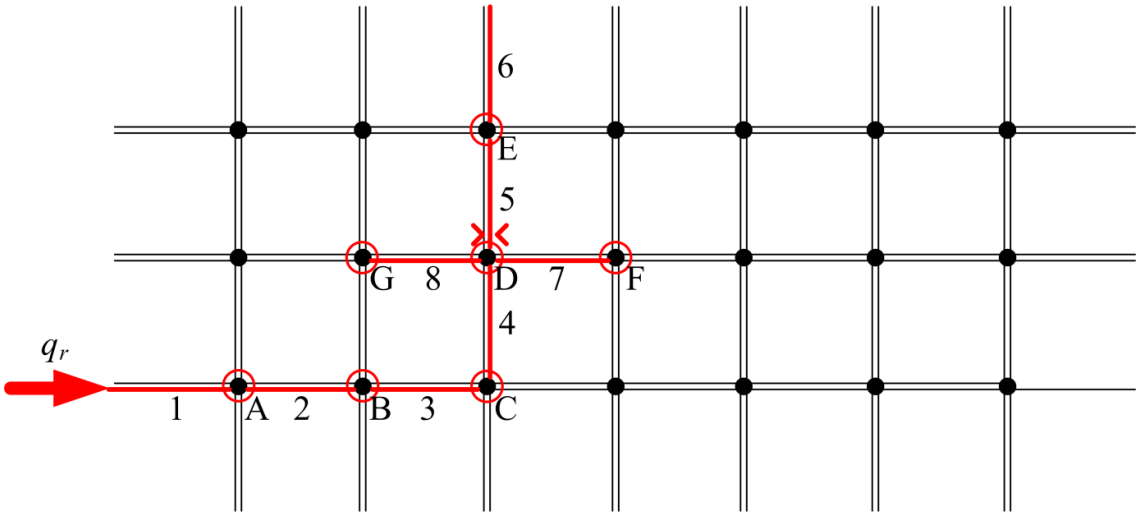

Figure 9-2: Demand variation – expanding affected area

Suppose a bottleneck has activated due to the increased route demand q_r, constraining the flows from intersection D into link 5. Some incoming and outgoing links of intersection D may remain unaffected, i.e. the change in flow (if any) is smaller than ε_{up} or ε_{down} respectively. Meanwhile, congestion forms on link 8, affecting it and activating intersection G at $t_G = t^c + \frac{L_8}{-w_8}$. Also, the decreased outflow into link 7 is tracked as a downstream effect, affecting link 7 and activating intersection F at $t_F = t^c + \frac{L_7}{v_{f,7}}$. The change in turning fractions generated by this secondary downstream effect is neglected, however.
It is clear from Figure 9-2 that the computational effort is limited in space and time. Outside the active part of the network and time period, no calculations are performed and the flows and traffic states are assumed identical to those of the base simulation.

9.1.1.5 Post-processing

The main output of MaC consists of the changed CVN of affected links. From this, traffic flows, TTS, VHL and link and route travel times can be derived. Afterwards, all variables (demands, link capacities, CVN and turning fractions) are reset to the base values and the next variation is simulated (starting at step 2).

9.1.2 Sources of error

As mentioned in Section 7.2, a discrepancy between the base model and the marginal derivation leads to approximation errors. In this case, the SC approach that is adhered in MaC (different from the MC LTM) causes small deviations. Before discussing these deviations themselves, it is explained how the propagation of these errors is confined. Finally, the introduction of accuracy thresholds naturally leaves room for errors that fall below these thresholds.

9.1.2.1 Propagation of errors due to discrepancy with base model

Clearly, when the MaC algorithm checks for changes in the flow of unaffected links of active intersections ($\frac{|q_{MaC} - q_{base}|}{q_{base}} > \varepsilon$), any discrepancy between MaC and the maternal base model (producing q_{base}) that exceeds ε would cause unjustified link affections and intersection activations. In such case, the marginal algorithm would be tracing errors due to a model discrepancy rather than changes due to the imposed variation. In other words, it would consume computation time for actually propagating and hence increasing the approximation error. Unlike in MIC, errors can propagate downstream in MaC and may thus quickly activate large parts of the network. Loss of not only accuracy but also computation time is therefore a potential hazard.
This undesirable effect can be avoided as follows. First the base scenario is run with the base model LTM, yielding the MC base results B_{LTM}. Then, the base scenario is re-run with MaC

with the entire network and simulation period activated[44]. This produces adapted SC base CVN B_{MaC} that now include the model discrepancy. Then, these adapted base results B_{MaC} are used as the input and comparing ground to marginally simulate the variations, thus avoiding the tracing and propagating of errors due to the model discrepancy. Indeed, the purpose of MaC is to approximate as closely as possible the difference between the results of the variations and base scenario as explicitly simulated by LTM (V_{LTM} - B_{LTM}). The difference $V_{MaC} - B_{MaC}$ is a more correct approximation of the impact of the variations than the difference $V_{MaC} - B_{LTM}$, in which the model discrepancy is mistakenly considered as being part of the output change caused by the variation.
Even though the discrepancy between MaC and LTM is very small, this operation reduces the computation time of MaC by more than 50 % in the case study of Section 9.2 while the explicit flows are (slightly) better approximated.

9.1.2.2 SC approximation

The computational gain that is achieved by adopting a SC approach in MaC comes at the cost of some approximation errors:

- Discretization of the turning fractions:
  MaC operates with discretized turning fractions at intersections with a fixed time interval T. This in itself may cause small errors compared to MC DNL models that calculate the flows disaggregated by route (or destination) in each simulation time step (which is typically smaller than T). Although the length of T influences the computation time (see Section 9.3.3) it can be chosen reasonably small, in the order of a few minutes. Therefore, large errors are unlikely to arise from the discretization. Naturally, the error increases with the size of the turning fraction interval and with the volatility of the real (non-discretized) turning fractions.

- Dependency of the turning fractions on the travel times:
  As explained in Section 9.1.1.1, the numerically approximated dependency of the turning fractions on the base demand and base travel times are input to MaC. To impose a demand variation, this relationship is used to update the turning fractions. However, the demand variation may cause travel times to change and thus alter this relationship. This can cause a time shift in the turning fractions in MaC compared to LTM. This error increases as the marginal travel times deviate more from the base travel times, and with the volatility of the turning fractions. In case of unacceptable errors, MaC could be run in an iterative loop to update the relation between turning fractions and demand according to the new rather than base travel times (as explained in Appendix B). Of course, the computation time would increase accordingly.

[44] Note that MaC is then actually a SC version of LTM. Hence, this comes down to performing a SC simulation of the base scenario, i.e. the turning fractions are fixed during time interval T instead of changing at each intersection update according to the composition of route flows on a link as in LTM.

- Neglecting the change of the turning fractions for secondary downstream effects: Whereas prior to simulating a demand variation, MaC adjusts the turning fractions along the way, this is not the case when tracking secondary downstream effects, as on link 7 in the example in Figure 9-2.

For applications in which the errors in MaC due to the SC approximation are unacceptable, a MC marginal DNL algorithm could be developed. Alternatively, MaC could be run in an iterative loop to update the relation between the turning fractions and the demand according to the new travel times (as explained in Appendix B). This way, the above approximations are avoided at the cost of an increase in computation time.

9.1.2.3 Accuracy thresholds

All flow changes that are below the chosen accuracy thresholds (ε_{up} and ε_{down}) are disregarded in MaC. Clearly this invokes some approximation errors, but this measure is needed to preserve the computational efficiency of MaC. At least for ε_{down}, a careful consideration is needed to select a value that allows to capture the important changes as much as possible, but is not so low that large parts of the network become activated needlessly. Since upstream moving changes not only propagate slower but also occur much less frequently, ε_{up} can be set very low. We refer to Sections 9.3.1 and 9.3.2 for more details on the sensitivity of MaC's performance to these thresholds.

9.2 Case study on the network around Ghent

In this section, a case study on the network around Ghent (Belgium) is presented to compare the performance of MaC with that of explicit simulation.

A four hour simulation period is considered on the Ghent network, which consists of 326 intersections, 992 links and 2032 routes connecting 32 origins and destinations. For simplicity's sake, only one base scenario is considered, which is calibrated to match the daily observed traffic patterns on the network. Then, a set of demand variations to this base scenario is evaluated with MaC on the one hand and explicit simulation (with LTM) on the other. The set consists of demand increases on each route separately with 100 veh/h (2032 variations) during one of the four hours of the simulation period (which is randomly chosen). The output that is compared are the link flows, or rather the difference between the flows resulting from a variation and the base flow. Since the flows are the most basic output of MaC and LTM[45], this comparison serves as a validation of MaC for various applications. Indeed, if the flows are approximated well, also other, derived outputs (e.g. travel times, VHL) will be accurate. The specific set-up of this case study can be seen as an investigation into the sensitivity of the link flows to each route demand; for example to numerically determine the gradient of an objective function in optimization procedures. In Frederix et al. (2011) a preliminary version of MaC is used in this set-up for dynamic OD estimation. Of

[45] More precisely, the CVN are the most basic output. However, the flows are the derivatives of the CVN and thus time-profiles of the flows provide the same information as the CVN.

course, other types of studies (e.g. reliability studies) may require a different set-up, with for example several simultaneous route changes or multiple base scenarios.

The simulation results of the base scenario with LTM, B_{LTM}, forms the comparing ground for the explicitly simulated variations V_{LTM}. As explained in Section 9.1.2.1, the base simulation results are re-derived with MaC, yielding B_{MaC}, to avoid the loss of accuracy and computation time due to tracing of errors due to the model discrepancy. These new base results B_{MaC} form the input and comparing ground for the marginal simulations of the variations, producing V_{MaC}.
The parameters of MaC are the up- and downstream accuracy thresholds ε_{up} and ε_{down}, and the interval for the turning fractions T. In this section, we present the results obtained with the set of parameters for which MaC was found to perform best in this case study in terms of results and computation time (see (9.2)). This set is derived from the sensitivity analysis in Section 9.3. For the reader's convenience, the detailed results of the optimal set are presented first, followed by the more summarized description of the sensitivity analysis in the next section.

$$\begin{aligned} \varepsilon_{up} &= 0.5\ \% \\ \varepsilon_{down} &= 2\ \% \\ T &= 3\ \text{min} \end{aligned} \tag{9.2}$$

It should be noted that the optimal settings for the parameters (particularly for ε_{down}) are probably case-dependent.

The explicit flow changes $\Delta q_{\exp}$ are the differences between the link flows in each explicitly simulated variation and the LTM base simulation. MaC approximates these flow differences (Δq_{MaC}). If Δq_{MaC} matches $\Delta q_{\exp}$ well, then also other, derived outputs (e.g. travel time (losses) and VHL) will be accurate.
The link flows and link flow differences Δq are averaged over 15 min intervals. For all 15 min Δq, 2032 values - one for each variation - are obtained from MaC and from the explicit simulations with LTM. This totals 32251904 link flow evaluations (2032 variations x 992 links x 16 time intervals). This number will be the reference point when percentages of changed flows with respect to all link flows are given. Unless stated otherwise, $\Delta q \geq 0.001$ veh/h are included in the results and graphs. Smaller changes are discarded; flows are considered unchanged in this case. As such, 1.6 % of all link flows have changed in the explicit simulations. Some of these Δq have also changed in MaC. Hence - as these flows are 'affected' in MaC as well as in the explicit simulations – these are called the Rightfully Affected Flows (RAF). The Wrongfully Unaffected Flows (WUF) have not changed in MaC (while they have in the explicit simulations. These cover 0.5 % and 1.1 % of all link flows respectively. Figure 9-3 shows, however, that this one third of all explicit Δq that are also detected in MaC (RAF) are in fact the most important changes; whereas the two thirds that are ignored (WUF) are indeed largely negligible. Note that the Cumulative Distribution

Function (CDF) in Figure 9-3 shows the explicit results $\Delta q_{\exp}$, which MaC tries to approximate as closely as possible.

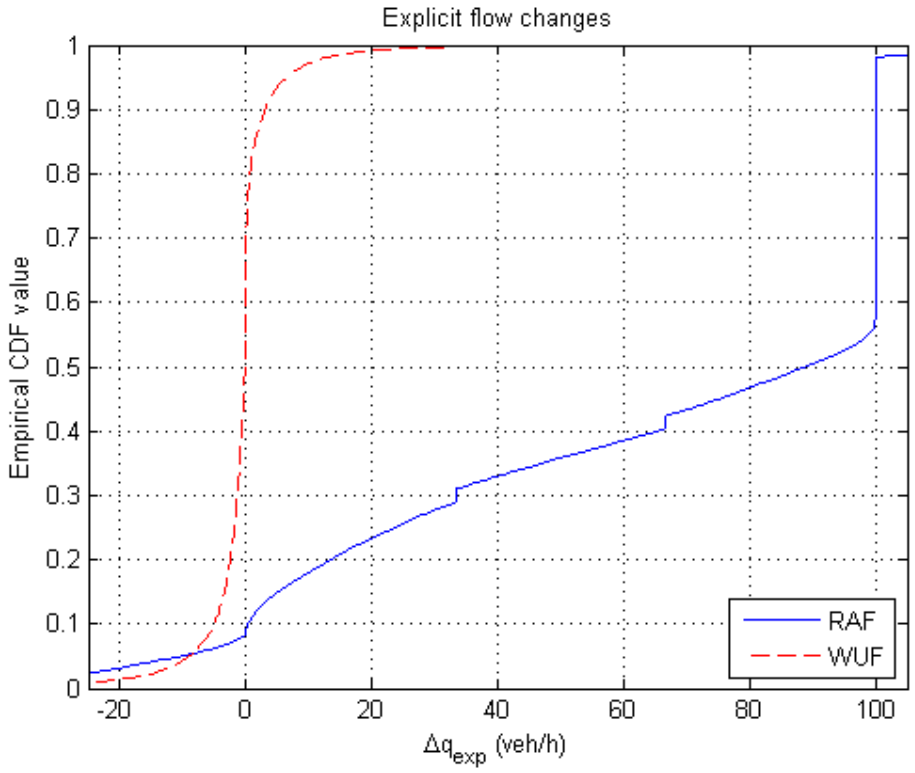

Figure 9-3: The explicitly simulated changes in link flows

Furthermore, a small percentage of all link flows (0.13 %) have changed in MaC but not in the explicit simulations; these are the Wrongfully Affected Flows (WAF).

Figure 9-4 demonstrates that MaC approximates the explicit results well. For only 7.2 % of the $\Delta q_{\exp}$ (RAF + WUF) – this is 0.11 % of all link flows – the error of the corresponding Δq_{MaC} is larger than 10 veh/h. For the flows that have only changed in MaC (WAF), this is only 3.1 % - or 0.004 % of all link flows.

Finally, Figure 9-5 depicts the absolute value of the relative error for the flow changes in MaC vs. explicit simulation, i.e. $\left|\frac{\Delta q_{MaC} - \Delta q_{\exp}}{\Delta q_{\exp}}\right|$. Of course, only the RAF are included, since for the WUF and WAF, the relative error is always 1 and infinite respectively. Different precisions are compared to separate the larger, more important flow changes in the results. It is clear that larger flow changes are better approximated.

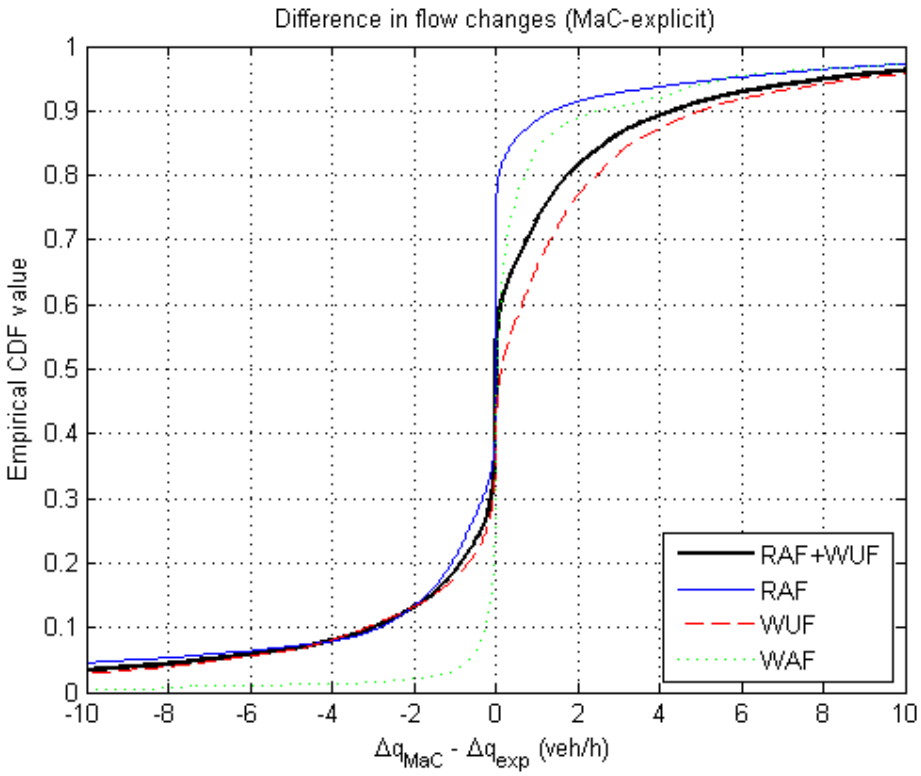

Figure 9-4: The error between MaC and explicit results

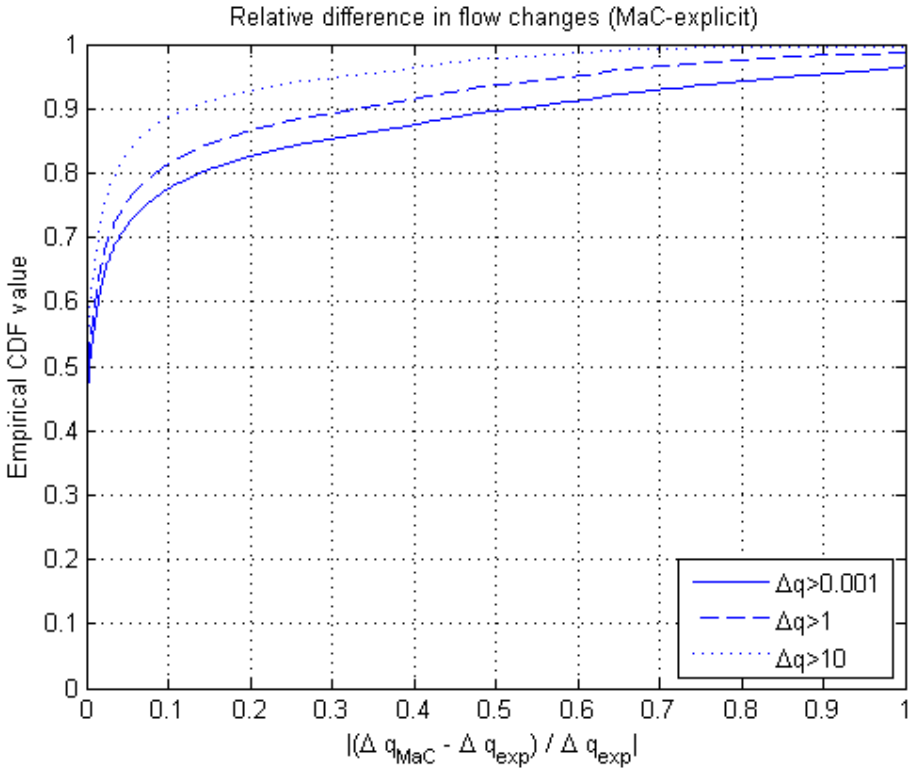

Figure 9-5: The relative error as a function of the size of the flow changes (only RAF)

Finally, let us compare the computation times. The explicit simulations take almost 27 hours, whereas MaC needs only 68 minutes to run the 2032 demand variations. This constitutes a computational gain by a factor of almost 25. This gain is largely due to the reduction of the number of required intersection updates: MaC performs only 3.9 % of the updates that are carried out explicitly. The reduction in number of intersection updates is thus comparable to the computational gain. This indicates that for this case study the overhead in MaC - to compare the new flows to the base flows at run time - is compensated by the computationally more efficient SC approach of MaC.

9.3 Sensitivity analysis

In this section, a sensitivity analysis of the accuracy and computational efficiency to the MaC parameters and to the size of the demand variations is presented. The same network and base scenario as in the previous section is used. The parameter settings in (9.2) and demand variations of +100 veh/h for each route are used as the reference and each parameter is varied separately from its reference value. As previously, only $\Delta q \geq 0.001$ veh/h are included in the results unless stated otherwise.

9.3.1 Upstream accuracy threshold (ε_{up})

The results of MaC for different values of ε_{up} are shown in Table 9-1.

Table 9-1: Sensitivity of $|\Delta q_{\exp} - \Delta q_{MaC}|$ to ε_{up}

ε_{up} (%)	RAF + WUF			WAF			
	50-pct	90-pct	99-pct	%	50-pct	90-pct	99-pct
10	0.8	8.0	43.4	0.12	0.1	2.4	15.7
2	0.8	7.6	31.4	0.13	0.2	3.6	18.6
0.5	0.8	7.6	31.2	0.13	0.2	3.6	18.6

The 50-, 90- and 99-percentile of the absolute value of the difference between Δq_{MaC} and $\Delta q_{\exp}$ are given. Firstly, Table 9-1 indicates the accuracy of the approximation of MaC of the explicit results, i.e. of the flows that have changed in the explicit simulations (RAF+WUF). Secondly, the additional errors on flows that should have remained unchanged (WAF) are given. Also, the percentage of all link flows that are WAF (%) contributes to the evaluation of the MaC results.

Table 9-2 illustrates the sensitivity of the computational performance of MaC to ε_{up}. The computation time and the number of intersection updates of MaC are given as a percentage to the time and updates needed for the explicit simulations.

Table 9-2: Sensitivity of computational gain to ε_{up}

ε_{up} (%)	% updates	% comp time
10	3.8	3.9
2	3.9	4.2
0.5	3.9	4.2

It is clear from Table 9-1 that the results of MaC only deteriorate for quite high values of ε_{up}. Since a too high ε_{up} mainly causes large errors (increasing the 99-pct) and because the computation time is not very sensitive to ε_{up}, this parameter should be set to a low value out of precaution.

9.3.2 Downstream accuracy threshold (ε_{down})

Table 9-3 and Table 9-4 show the sensitivity to ε_{down}. It is immediately clear that selecting a proper value for ε_{down} is a more difficult task. On the one hand, the explicit changes are better approximated with low ε_{down}. On the other hand, the number of WAF increases. This is logical, since lowering the threshold ε_{down} allows tracking more and smaller changes, but also more and smaller errors are recorded. Consequently, also the computation time increases. An appropriate value for ε_{down} should be selected based on the available time budget and the required accuracy of the results. Since secondary downstream effects are in general smaller than the imposed variations, it is advisable to set ε_{down} smaller than the size of the variations (relative to the link flows) - at least if these secondary downstream effects are of interest.

Table 9-3: Sensitivity of $|\Delta q_{\exp} - \Delta q_{MaC}|$ to ε_{down}

ε_{down} (%)	RAF + WUF			WAF			
	50-pct	90-pct	99-pct	%	50-pct	90-pct	99-pct
10	0.9	8.2	41.5	0.03	0.1	1.6	15.7
2	0.8	7.6	31.2	0.13	0.2	3.6	18.6
0.5	0.7	6.2	25.7	0.65	0.1	1.7	11.5

Table 9-4: Sensitivity of computational gain to ε_{down}

ε_{down} (%)	% updates	% comp time
10	2.7	3.9
2	3.9	4.2
0.5	8.2	10.2

9.3.3 Turning fraction interval (*T*)

As expected, the accuracy of the results is not very sensitive to *T* (see Table 9-5). However, the turning fraction interval does influence the computation time (Table 9-6). On the one hand, less intersection updates are required for smaller *T* since the tracing of approximation errors is reduced, thus lowering the computation time. However, this is overcompensated by the fact that the computational overhead increases significantly for very small *T*, especially during pre-processing to update the turning fractions according to a demand variation.

Table 9-5: Sensitivity of $|\Delta q_{\exp} - \Delta q_{MaC}|$ to *T*

T (min)	RAF + WUF			WAF			
	50-pct	90-pct	99-pct	%	50-pct	90-pct	99-pct
10	1.1	8.6	30.9	0.29	0.4	3.2	12.5
3	0.8	7.6	31.2	0.13	0.2	3.6	18.6
1.5	0.8	7.6	31.2	0.08	0.1	5.6	22.3

Table 9-6: Sensitivity of computational gain to *T*

T (min)	% updates	% comp time
10	4.9	5.7
3	3.9	4.2
1.5	3.5	7.0

9.3.4 Size of variations

In the previous sections, the sensitivity to the parameters of MaC was discussed. This section inquires into the effect of the size of the variations that are imposed onto the base scenario. Therefore, the absolute value of the relative error is compared for different sizes of demand variations (Figure 9-6). Only RAF for which both $\Delta q_{\exp}$ and $\Delta q_{MaC} \geq 1$ veh/h are included in these results.

Not surprisingly, MaC approximates the explicit results better if the variations are small, since then the variations are closer to the base scenario. Approximation errors made in MaC – neglecting the change of turning fractions for secondary downstream effects and due to changing travel times (see Section 9.1.2) – obviously cause larger errors for more severe perturbations[46]. For extreme variations (e.g. +1000 veh/h for one route), the MaC results are no longer reliable. It should be noted, however, that – at least in this case study – adding 1000 veh/h to a route demand usually implies an increase to many times the base value. Even an increase with 100 veh/h can already be considered as a large perturbation for most routes.

[46] Note that supply variations (to link capacities) do not (directly) require changes to the turning fractions as is the case for demand variations. Since approximation errors to the turning fractions are the main source of error in MaC (see Section 9.1.2), and because a supply variation is confined to a single location, link capacities may be varied more strongly than the demands before large approximation errors are to be expected.

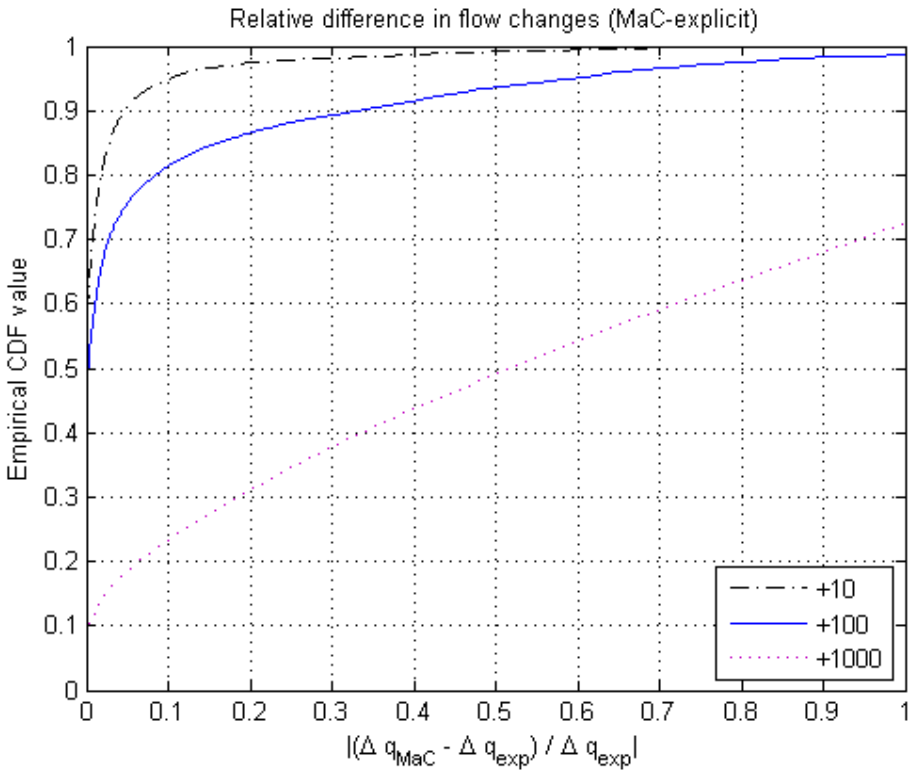

Figure 9-6: Sensitivity of results to the size of the demand variations

Naturally, also the required intersection updates and computation time increase with the size of the variations (Table 9-7).

Table 9-7: Sensitivity of computational gain to the size of the demand variations

variation	% updates	% comp time
+10	2.8	3.3
+100	3.9	4.2
+1000	15.6	21.9

These results show that the range in which variations can diverge from the base scenario is not unlimited. To improve the results for large perturbations, MaC could be run in an iterative loop to update the relation between the turning fractions and the demand according to the new travel times, thus improving the approximation (at the cost of higher computation time however). On the other hand, extreme perturbations such as adding 1000 veh/h can no longer be regarded as a local approximation of a derivative (in optimization procedures) or even DTD stochasticity in variability studies – except perhaps for a very select number of routes, e.g. on busy highways. For some applications, where large perturbations are expected (for instance event analysis), these perturbations will have to be included in the base scenario(s). MaC can then be used to quantify the stochasticity as variations to these expected perturbations in the base scenario.

9.4 Conclusion

In this chapter, the polyvalent marginal DNL simulation algorithm MaC is presented that can replace repeated explicit DNL simulations (with high overlap) if the computation times of the latter are impracticably high. This is because MaC is not only computationally efficient. Since it is based on first-order traffic flow theory, it is far more realistic than existing fast tools such as static models, which are often used as an alternative if the time budget does not allow the use of dynamic simulation models. Also, it exhibits smaller approximation errors than the previously developed MIC (Chapter 8). This is because it is algorithmically much closer to the maternal base model LTM. Contrary to MIC, intersection updates in MaC are performed entirely analogous as in LTM and also the link model has the same functionalities[47]. Checks are introduced to limit calculations to the active part of the network in which changes (are expected to) occur from the variation to the base scenario.
This improved approach has many implications, mostly advantageous. The fact that the approximation errors are reduced has already been mentioned. As a result, it enables evaluation on a finer level than MIC, e.g. of route or link travel times or flows instead of aggregated output such as VHL[48]. This, in combination with the ability to simulate both demand and supply variations, renders MaC generally applicable to simulation-based problems that are currently infeasible or at least highly computationally demanding (except on a very small scale). In TTV studies, MaC can be used to examine the impact of not only incidents, but also of DTD stochasticity of demand and supply. Other applications include numerically approximation of the derivatives of an objective function to input variables in optimization problems such as (robust) network design[49], DTM optimization (ramp metering, variable speed limits, route guidance, etc.) and dynamic OD estimation. Frederix et al. (2011) validates the use of MaC for the latter.
Finally, on the downside, the enhanced accuracy and extended applicability slightly reduce the computational efficiency of MaC compared to the previously developed MIC. Therefore, MIC can still be preferred for large-scale incident analysis, in particular for quick scan purposes. Still, the computational gain of MaC compared to explicit simulation with LTM is significant.

The case study in Section 9.2 (on the medium-scale network around Ghent) compares the performance of MaC with that of its maternal base model LTM. The comparison is made based on the link flow changes resulting from 2032 demand variations. This makes for a general evaluation of MaC. For some applications (e.g. when deriving the sensitivity of the link flows to the demand), the flows are directly of interest. Furthermore, since link flows are the most basic output of MaC (and LTM), they serve also as a good indicator for the accuracy of other, derived outputs such as travel times and VHL, which are of interest in other applications (e.g. TTV studies).

---

[47] The only remaining discrepancy between MaC and LTM is that MaC is SC, which means it regards traffic as one homogeneous flow on the link level rather than considering disaggregated flows by route like the MC LTM.
[48] Still, marginal simulation should be preserved for applications where the detailed outcome of one particular simulation is not directly of interest.
[49] In robust network design, which incorporates the effect of variability to create reliable traffic networks (see e.g. Snelder, 2010), a marginal DNL model could be used to simulate multiple scenarios with varying demand and supply (e.g. due to incidents) as well as to determine the sensitivity of the network's performance to changes in the network design (e.g. changes in link capacities).

The case study shows that MaC approximates the results of repeated explicit simulations well: 77.5 % of all explicit flow changes are captured with an approximation error of less than 10 % (see Figure 9-5). Since only 1.6 % of all link flows have changed, this corresponds to 99.6 % of all link flows. Furthermore, merely 0.13 % of all link flows have wrongfully changed in MaC. We note that in many applications (e.g. TTV studies) the errors in estimating the model inputs (traffic demand, network characteristics) and the deviations between traffic flow theory and reality overshadow the additional error introduced by replacing an explicit DNL simulation model with its marginal counterpart. If the marginal simulation is used to determine derivatives (e.g. in optimization problems), this may not always be the case, however.
A computational gain of a factor 25 is attained with MaC compared to LTM. This gain is largely due to limiting calculations to the active part of the network for each variation, so that MaC only requires 4 % of the intersection updates of the explicit model. Further computational improvement to MaC could be established by implementing the possibility to de-activate intersections once no further flow changes are detected and expected; however, this also requires introducing additional checks. The computational gain of MaC increases with the network size (number of intersections and links) and with the number of routes. Finally, we believe that a thorough code optimization could further increase the computational gain significantly.

The sensitivity analysis in Section 9.3 shows that in particular the downstream accuracy threshold ε_{down} - determining whether secondary downstream changes are tracked or neglected - strongly influences the results and computational efficiency of MaC. Furthermore, it is shown that if the marginally simulated variations deviate extremely from the base scenario, the results become unreliable. Firstly, MaC could be run in an iterative loop to update the relation between the turning fractions and the demand according to the new travel times (as explained in Appendix B). Another option could be to develop a MC marginal algorithm from LTM. As then there would be no discrepancy with LTM, this would result in fewer approximation errors. However, both approaches would come at the cost of an increased computation time. Moreover, the observed range in which MaC is sufficiently accurate should suffice for most applications. In some cases (if large perturbations are expected; e.g. in case of mass events or extreme weather forecasts), the best approach might be to include the most severe perturbations in several base scenarios. MaC could then be applied to simulate variations to these anticipated perturbations.

Finally, it should be noted that an important contribution is that, analogous to how MaC is derived from LTM, other marginal DNL algorithms can be drafted from other existing explicit DNL models by following the procedure presented in Section 9.1.1. Indeed, simulation-based DNL models inherently apply a discretization in space and time. Hence, marginal algorithms can be developed that confine the calculations to a (small) part of the network and time period. While in MaC (as in LTM) the calculations are steered by intersection updates, other DNL models (and their marginal derivations) may operate for instance by updating cells (CTM). However, this does not constitute a fundamental difference. Indeed, analogous to MaC, other marginal DNL simulation algorithms can first isolate the local variations (e.g. to some intersections and links as in MaC, or to some cells).

Within this affected area, calculations are to be carried out according to the base model's principles (or – if desirable for reasons of computational efficiency – a close approximation thereof). If changes are detected (in traffic flow or whatever indicator is used), the marginal algorithm should expand the affected area in correspondence with the propagation assumptions adopted from the base model.

A notable advance would be the development of a marginal DNL simulation tool that includes en-route rerouting, as this would greatly improve the realism in many potential applications. Following this observed opportunity, a proof-of-concept implementation of MaC extended with an en-route rerouting model is provided in Appendix G. Ideally, however, such a marginal algorithm should be developed from a different base model that already includes en-route rerouting instead of adding an en-route rerouting component in a later phase.

10
CONCLUSION

The general aim of this PhD research has been to further advance first-order macroscopic simulation-based DNL modelling. Furthermore, we have limited the scope to single-class and single-lane models. As motivated in Chapter 1, these are the simplest models that are sufficiently realistic for a wide range of applications in congested road networks. Hence, it makes sense for advances that are fundamental (intersection modelling) or entirely new to transportation modelling (marginal simulation) to be developed first for this type of models. An obvious and interesting future research step would be to let the presented developments form the basis for analogous advances to other types of DNL models (e.g. multi-class, multi-lane or second-order models). For some types of models, this may be quite straight-forward, for others additional research and modelling efforts may be required.

Two technically independent research directions have been pursued in this thesis, namely the (further) development of the macroscopic DNL intersection model and of marginal DNL simulation. Both constitute an important advance that generally serves a broad range of DNL applications. In Section 10.1, the main findings and contributions of the part of the thesis on intersection modelling are orderly presented, and necessary and interesting future research directions are elaborated on. Likewise, Section 10.2 treats marginal simulation.

10.1 Intersection modelling

10.1.1 Main findings and contributions

The initial objective of this research – as stated in Section 1.2.1 - has been to develop theoretically consistent intersection models that realistically represent traffic flows and congestion dynamics, particularly at oversaturated urban and regional intersections. This aim has been subdivided in two steps, corresponding to the two main functions of the DNL intersection model as explained in Section 2.1. The first function is to find a consistent solution for the flows over the intersection taking into account all demand and supply constraints. The second function is to consider not only the external constraints of the adjacent links but also internal supply constraints arising from internal conflicts within the intersection itself (e.g. conflicts between crossing flows). Chapters 3 and 4 have addressed the first function. The second function has been introduced in Chapters 5 and 6. Hereby, we have started from a general intersection model in the first chapters. In Chapter 6, specific intersection models are presented for each of the intersection types mentioned in Section 2.2.2 (PTTR, AWSC, priority-controlled, roundabout and signalized). Also, we have adopted the more traditional point-like approach to DNL intersection modelling. This means that the intersection model is a dimensionless model that solves a large set of strongly coupled equations, rather than a spatial model (see Section 2.2.1). Finally, while our developments have been made with implementation in DNL models based on first-order traffic flow theory in mind, the presented intersection models can also be combined with vertical or horizontal queuing link models. DNL models based on second-order traffic flow theory, however, require more complex intersection models that, in addition to transferring flow, also transfer speeds (or momentum).

The main findings that form the basis of (the contributions of) our research are the following:

- The vast majority of existing macroscopic DNL intersection models (limited to external constraints) either make unrealistic or artificial assumptions, or are macroscopically impractical, or may produce solutions that are inconsistent in some or several ways. The latter implies a violation of FIFO, of the invariance principle of Lebacque & Khoshyaran (2005) or a failure to (individually) maximize the flows with respect to all prevailing constraints. This finding comes forward from the review of existing macroscopic DNL intersection models presented in Section 3.3.

- Very few state-of-the-art intersection models impose internal supply constraints that account for the fact that supply of the inner infrastructure of an intersection – i.e. at the conflict points of crossing movements, or the merge points on the arcs of a roundabout - may limit the flows (see the literature overview in Section 5.1.2.1). This omission is acceptable for simple (highway) merges and diverges. Also under certain traffic conditions, e.g. in case of low traffic volumes or in situations in which an external supply is dominant (for instance if heavy congestion spills back onto the intersection), one may do without these internal supply constraints. In many other situations, however, such internal constraints are decisive at busy urban and regional intersections. Consequently, to provide a realistic representation of traffic flows and

congestion formation for such intersections, the development of the macroscopic DNL intersection models (as in this thesis) is highly necessary.

- It has been found that the uniqueness of the solution of the intersection model, in particular in the presence of internal supply constraints, is – contrary to the common assumption in the literature - not trivially guaranteed (see Section 5.2). This holds not only for the intersection models developed in this thesis, but also for macroscopic DNL intersection models in general. Moreover, some existing models (Gentile et al., 2007; Gentile, 2010; Yperman, 2007; Flötteröd & Rohde, 2011[50]) may exhibit multiple solutions. Moreover, this solution non-uniqueness is found to appear under realistic behavioural assumptions. This severely complicates the process of establishing a unique solution, which is necessary in traditional, deterministic DNL modelling. Finally, it is shown in Section 5.3 that the non-uniqueness problem is not solved by adopting a spatial modelling approach. This reinforces our preference for point-like intersection models in macroscopic DNL, also for complex intersections with internal conflicts.

Several theoretical and practical contributions are presented in this dissertation in response to the above described shortcomings in the state-of-the-art. Firstly, we list the theoretical contributions:

- A set of seven generic requirements for first-order macroscopic DNL intersection models is formulated that ensures the proper fulfilment of the first function (see Section 3.4.1):

    - General applicability to any number of incoming and outgoing links and any combination of boundary conditions
    - Non-negativity of flows
    - Conservation of vehicles
    - Satisfying the demand and supply constraints
    - Ensuring FIFO: Conservation of turning fractions (CTF)
    - Individual flow maximization (each flow is actively constrained by demand or supply)
    - Satisfying the invariance principle of Lebacque and Khoshyaran (2005)

    This set of requirement helps model developers to build realistic intersection models that are consistent with all demand and supply constraints and with sensible traffic flow dynamics; see for instance Gibb (2011) and Flötteröd & Rohde (2011) who continued on this contribution.

[50] Flötteröd & Rohde recognize and acknowledge this. They also present a heuristic algorithm with guaranteed convergence towards a compromise solution. Hence, Flötteröd & Rohde (2011) is an important seminal work regarding solution non-uniqueness of the intersection model.

- In addition to the above generic requirements, Section 3.5 explains how the Supply Constraint Interaction Rules (SCIR) govern the supply distribution, considering the interdependencies that exist between the various demand and supply constraints. To our knowledge, the need for and function of the SCIR have never before been properly elaborated on. It should be noted that – while the vast majority of existing (often incomplete) SCIR apply a distribution based on certain proportionalities – several plausible definitions for the SCIR are conceivable (as in Gibb, 2011).

- Internal supply constraints are introduced in Chapter 5. Analogous to how external supply constraints are universally treated in the state-of-the-art, a distribution of internal supply is proposed based on proportionality of priority parameters of an incoming link for a certain supply. This leads to a general intersection model framework that accounts for internal conflicts within the intersection itself. This is mainly a theoretical contribution, showing how practical models can be build from this base. Hereby, different assumptions regarding the driver behaviour with respect to the priority rules in a conflict – distinguishing absolute and limited compliance to the priority rules and turn-taking (see Section 5.1.1) – can be translated to a corresponding selection of the priority parameters for that conflict.

- A sufficient and necessary condition for solution uniqueness of the intersection model is presented in Section 5.2.2, expressing that the proportionality of the priority parameters or priority ratio between any two incoming links must be the same no matter which (internal or external) conflict is considered. This comes down to the requirement that the priority parameters must be single-valued for each incoming link. Importantly, this uniqueness condition appears behaviourally unrealistic in the presence of internal supply constraints. Indeed, it is (often) in contradiction with the priority rules, and thus with how one would naturally define the priority parameters. Therefore, it is necessary to empirically study whether or not the non-uniqueness that is observed in the model exists in reality as well. For now, to enable the further use of traditional deterministic DNL models, which require a unique solution of the intersection model, we see no other possibility as to solve this problem in a rather pragmatic way (see below). Future research should determine whether these pragmatic approaches will have lasting value (if an intersection model producing a unique solution is shown to be sufficiently realistic) or not.

The above theoretical contributions are our primary achievements regarding intersection modelling. They provide essential insight for model developers; namely how to ensure theoretical consistency - firstly for models limited to external constraints, secondly when increasing the model's realism (and complexity) by adding internal supply constraints. The following practical contributions of our research are important advances along these lines:

- Section 4.1 defines a general intersection model limited to external constraints[51], which is presented as a set of implicit equations. The SCIR of this model consistently distribute the supplies based on proportionality of general, single-valued priority

[51] Still, in some situations, e.g. in case of heavy downstream congestion, this simplified model can suffice.

parameters. This model complies with the seven requirements listed above and produces a unique solution. Hence, unlike the vast majority of existing intersection models, it properly fulfils the first function of the intersection model in first-order macroscopic DNL.
A specification of this model is put forward in Section 4.2, by detailing the priority parameters as oriented capacities. This means that the strength of an incoming link in the competition for the supply of an outgoing link is determined by the incoming link's capacity and its turning fraction towards this supply. An efficient solution algorithm is presented in Section 4.3. Due to their analogy with the model presented in Chapter 4, this algorithm can be readily extended to solve the subsequently presented intersection models.

- Chapter 6 presents specific intersection models for AWSC, PTTR, roundabout, priority-controlled and signalized intersections. For these models, internal supply constraint functions are defined based on conflict theory (Brilon & Wu, 2001) in Section 6.1.1. Also, the pragmatic approach of pre-processing the priority parameters (as first suggested in Section 5.2.4), i.e. the weighting of multiple-valued priority parameters per conflict into a single-valued priority per incoming link, is specified in Section 6.1.2. Although sensible and thus practically valuable, this approach requires validation in future research (in particular the weights that are given to each priority parameter in this pre-processing). To our knowledge, these models are the first to combine both functions of the intersection model into a consistent solution that is guaranteed to be unique.

In conclusion, this PhD research has aimed to enhance the theoretical soundness and practical accuracy of the intersection model regarding both congestion propagation – connecting the traffic flow and shockwave dynamics in adjacent links – and congestion formation due to limited supply of the intersection itself. The theoretical soundness is conclusively solved by the theoretical contributions as described above – although still, additional research into the properties of the solution non-uniqueness is highly meaningful (see the next section). Regarding the practical accuracy, clearly the elimination of theoretical inconsistencies (that result in errors in the solution) is also advantageous for the model's realism. Furthermore, the mere fact that internal conflicts are accounted for puts these models ahead of most existing models.
Considering the above simplifications, the practical models of Chapter 6 are to be considered as an important intermediate step in the process towards realistic complex intersection models, rather than the end of the line.

10.1.2 Future research

Future research on DNL intersection modelling is necessary to further enhance the realism of the models (primarily focused on the internal supply constraints) and to better understand and deal with the solution non-uniqueness.

- Solution non-uniqueness:
  The further unravelling of non-uniqueness in the intersection model requires both theoretical and empirical efforts:

  - Theoretical:
    Studying the parallels between the observed non-uniqueness in the intersection model and in the equilibrium DTA problem constitutes interesting theoretical work that may enhance the understanding of both problems. The similarity between the two is clear from the following statement – made with regard to equilibrium DTA: "The problem is inherently characterized by ill-behaved system properties that are imposed by the need to adequately represent traffic realism and human behaviour. This is further exacerbated by the time-dependency and randomness in system inputs. A fundamental consequence of this reality is that a theoretical guarantee of properties such as existence, uniqueness, and stability can be tenable only through compromises in depicting traffic theoretic phenomena and potentially restrictive assumptions on driver behaviour. Viewed from the complementary perspective, an ability to adequately capture traffic dynamics and driver behavioural tendencies precludes the guarantee of the standard mathematical properties." (Peeta & Ziliaskopoulos, 2001). Indeed, this applies very much to the non-uniqueness in the intersection model as well.

  - Empirical:
    It is envisaged that the solution non-uniqueness poses a more severe burden to the practical applicability of the DNL intersection model than to the equilibrium DTA assumption. To clarify the severity and implications of the problem, empirical research would be highly useful. This dissertation aids by helping to understand the phenomenon and when it occurs in the model, so that these specific circumstances can be sought in the field. If the existence of multiple flow patterns under identical boundary conditions is indeed observed, their characteristics – e.g. probability, frequency of switches, duration of stable periods - and the (external) factors that lead to these characteristics – e.g. history, neighbouring (signalized) intersections, intersection geometry – should be identified. As discussed in Section 5.2.3.2, such a finding may necessitate (for some applications) a paradigm shift towards stochastic or chaotic intersection models. For instance stochastic DNL models (e.g. Sumalee et al., 2011; Osorio et al., 2011) could treat the solution non-uniqueness of the intersection model probabilistically, by replacing the non-unique flows by a unique distribution. If this empirical work disproves the existence of non-uniqueness in reality (or shows it to be negligible for at least some applications), it should support the development of stronger modelling guidelines for intersection models producing a realistic, unique solution.

- Realism:
  Additional research aiming to further enhance the realism of the intersection model, in particular with regard to internal conflicts, is essential. Several ways for improvement can be identified:

  - Turning lanes:
    The separation (or grouping) of partial flows into turning lanes is not considered in the presented models. Keeping the models in their current form, this could be treated by introducing a diverge on an incoming link where the turning lanes start. Then, these lanes are physically separated and considered as different incoming links in the intersection model. However, this introduces short links into the DNL model, which, for LTM, has a negative impact on the computation time due to the CFL condition. Therefore, a better approach to handle this problem is desirable. Further research is needed to clarify how to include turning lanes properly, how this may affect the definition of the priority parameters in the model, and if (and which) changes to the current model framework are required.

  - Driver behaviour:
    The presented model assumptions are to be validated and (possibly) refined in order to improve the representation of driver behaviour, foremost in internal conflicts. Particularly empirical research is necessary for this. Moreover, the optimal model definitions are likely to be specific to a particular type of intersection and to the expected driver behaviour regarding priority compliance.
    Firstly, the SCIR (see Sections 3.5 and 4.1-4.2) that govern the distribution of supplies as defined in this thesis (which is closely related to other definitions in the state-of-the-art) are to be compared to possible alternatives (e.g. the definition of Gibb, 2011). Also the definition of the internal supply constraint functions based on conflict theory (Brilon & Wu, 2001) may have to be refined. In any case, stronger guidelines are needed for the configuration and composition of the conflict zones (see Section 6.1.1.2). Finally, the possibility of defining priority functions that depend on the flows, rather than constant priority parameters to determine the competitive strength of incoming links for the supplies is to be investigated (see Section 5.1.2.2). This may be necessary to realistically capture the transitions between different driver behaviour, e.g. based on the saturation level. Realistically defining priority functions while ensuring the uniqueness of the solution seems very difficult, however.

  - Influence of microscopic dependencies:
    Accounting for the influence of microscopic dependencies (see the discussion in Section 6.5) constitutes a challenge in future research. This includes foremost the interrelationship of (internal) supply constraints due to simultaneous decision making of drivers who have to traverse several conflict points. Another microscopic influence that is currently neglected is the order

in which a flow encounters conflicts. Future research is needed to determine how these dependencies influence the definition of the internal supply constraint functions. Also, this may again create a need for priority functions rather than constant parameters.

In conclusion, in this thesis we have developed intersection models that are foremost fully theoretically consistent. Achieving this - knowing that in a later phase perhaps some of this consistency may have to be sacrificed - is a logical first step towards the development of realistic, complex DNL intersection models (that are as theoretically consistent as possible). Indeed, it is envisaged that model extensions as listed above – particularly the last one - cannot be fully reconciled with all of the desirable model properties acquired in this thesis. More precisely, solution uniqueness and compliance with the requirements of CTF, individual flow maximization and the invariance principle of Lebacque & Khoshyaran (2005) limit the range of action of the modeller. Consequently, it may be necessary to (partially) release some of these (in itself desirable) model properties. It is important to realize the consequences of model adaptations (i.e. what properties the model possesses and which not), so that a full understanding of the model output is retained. The insight provided in this thesis will be vital for making a sensible compromise between theoretical consistency and realism in future intersection models.

Apart from a potential loss of consistency, an enhanced realism likely comes at the cost of increased complexity, computational cost and calibration efforts. Depending on the application, different trade-offs between realism, theoretical consistency, model complexity and data requirements of the intersection model will be desirable. Hopefully, this part of the research domain will evolve comparably to that of link models, where various theories – e.g. travel time functions, vertical and spatial queuing – provide different levels of complexity and realism.

To validate (potentially) satisfactory compromise(s), empirical research is indispensable. Data should be collected on different types of intersections and under varying saturation levels, as (compliance to) the priority rules may vary accordingly. Hereby, the effects of intersection geometry, priority rules, (shared) turning lanes, simultaneous driver decisions and the interaction between different conflicts need to be studied. Gathering and analyzing the necessary data will be a considerable challenge from a theoretical as well as a practical perspective. In our opinion, validation against microscopic simulation models – an approach that is often turned to in the literature in absence of empirical data – does not provide a suitable alternative. This is so because microscopic models are highly steerable through parameter calibration. Moreover, most existing microscopic simulation models do not realistically represent the vehicle interactions at intersections; especially not in congested conditions (see e.g. Chevallier & Leclercq, 2009).

10.2 Marginal simulation

10.2.1 Main findings and contributions

The initial objective of this second part of our research was to improve the computational efficiency of DNL simulation in the context of stochastic DNL (and DTA) modelling of

variable traffic conditions (see the discussion in Appendix F.1). Hereto, we have aimed to develop marginal derivations of the simulation-based DNL model LTM (Yperman, 2007) to retain the realistic representation of congestion dynamics while significantly reducing the required computation time. This is beneficial for a wide range of applications, from long-term planning for large-scale projects (including variability of traffic conditions) to real-time traffic prediction, management and control. Furthermore, marginal algorithms are expected to have a long lasting value, since reducing computation time always increases the scope that can be handled. Therefore, the increasing speed of computers will further enhance the possibilities of marginal simulation in the future, rather than rendering it obsolete.
Firstly, the Marginal Incident Computation (MIC) algorithm was developed (Chapter 8). This model is able to quantify the congestion effects resulting from incidents in a highly efficient way. To enable also the simulation of demand variations, a slightly different approach was necessary. This resulted in the more versatile Marginal Computation (MaC) algorithm (Chapter 9).
The development of MaC significantly increased our understanding of marginal DNL simulation. The main finding of this part of the dissertation is the insight in the proper development and characteristics of a marginal simulation algorithm. As is described in Section 7.2, a marginal algorithm should be directly derived from that of the maternal base model (i.e. a traditional, explicit simulation model). This way, the marginal simulation tool can perform marginal simulations as variations to a base scenario (run with the maternal base model) with minimal error. It became clear that the potential application of marginal DNL simulation far exceeds its original aim. Any application or study that requires a large number of successive simulations with large overlap (i.e. largely identical input and thus largely identical output) may benefit from a marginal simulation approach that limits the focus to only the differences among the successive simulations, saving computation time by (largely) avoiding identical calculations.

The main contributions of this part of the PhD thesis are the following:

- MIC, outlined in Section 8.1, is designed for fast Monte-Carlo simulation of incidents (i.e. reduced link capacities) on road networks. It superimposes the congestion effects due to an incident onto the CVN from a base simulation. It exhibits a realistic representation of queue propagation and congestion spillback, consistent with first-order traffic flow theory (as in Newell, 1993). In the vulnerability analysis case study of Section 8.2, MIC produces results (in terms of VHL) that are highly similar to those of explicit simulation with LTM. Furthermore, MIC reduces the computation time to less than 1 %. An example of the practical usefulness of MIC is given by Snelder (2010), who applies MIC to evaluate the network robustness in robust network design.

- The more versatile MaC (see Section 9.1) deviates much less than MIC from the maternal base model LTM. Consequently, it can simulate both demand and supply variations (to link capacities) – significantly broadening its applicability - and its results are more detailed and accurate. This improved and extended functionality is shown in Section 9.2, which presents a case study that compares the sensitivity of the link flows to route flow changes in MaC and LTM. While the results are close - 98.3

% of all link flows are rightfully unchanged in MaC; and 77.5 % of all explicit flow changes are captured with an approximation error of less than 10 % - the computation time is reduced to 4 %. Since the comparison is made on the basis of a basic model output (link flows), other, derived outputs (e.g. travel times, VHL) are accurate as well. Hence, this case study serves as a validation of MaC for a wide range of applications. Moreover, we refer to Frederix et al. (2011), in which the use of MaC in gradient-based OD estimation is validated.

- An important contribution is that, analogous to how MaC is derived from LTM (see Section 9.1), other marginal DNL algorithms can be drafted from other existing simulation-based DNL models. Hereby, it is important to follow the general philosophy and scope of marginal simulation as discussed in Section 7.2. Firstly, this implies that the marginal algorithm should be as closely related as possible to the maternal base model. Secondly, since marginal simulation is intended for fast iterative, finite difference or Monte-Carlo simulation, its application should be limited to problems where the outcome of each individual simulation is not directly of interest. Rather, the aim should be to quantify the aggregate properties of the set of simulations, for instance its probabilistic characteristics (in TTV studies) or the optimization direction it produces.

In conclusion, the marginal DNL simulation algorithms developed in this thesis approximate the results of explicit simulation (with LTM) well. A very significant computational gain is achieved that is proportional to the size of the network and route set. This enables the use of DNL simulation in a wide range of applications (see below) that are currently infeasible or at least highly computationally demanding. Finally, particularly for MaC, it should be possible to further improve the computational efficiency thorough code optimization – moreover, this is probably necessary to render it feasible for real-time applications.
While the computationally more efficient MIC may be preferred for some more coarse-grained incident-related applications (e.g. vulnerability analysis or incident-related DTM support – see further below), MaC (and marginal DNL simulation in general) serves a wide range of potential applications. In TTV studies, it can be used to examine the impact of not only incidents, but also of DTD stochasticity of demand and supply. In optimization problems such as dynamic OD estimation, optimal control (e.g. ramp metering) and evacuation planning, MaC could be used to numerically approximate the gradient of the objective function[52]. Robust network design combines the two aforementioned categories, since it incorporates the influence of variability to find optimal traffic network designs. A marginal DNL algorithm could be used both on the (stochastic) DNL (or DTA) level to simulate multiple scenarios with varying demand and supply as on the optimization level to determine the sensitivity of the network's performance to changes in the network design (e.g. changes in link capacities). Furthermore, marginal simulation may also support (real-time) DTM decisions. For instance, if an incident is detected, the marginal simulation algorithm could be used as a quick scan tool to predict a probabilistic range of the possible impact of

[52] Especially in the first iterations - when the distance to the optimum is still large – fast marginal simulation probably suffices to determine the search direction. In the last iterations, one could divert to explicit simulation if higher accuracy is required to reach the optimum.

the incident on the network. If the marginal simulation tool encompasses an en-route rerouting model, it could be a valuable aid in the drafting and evaluation of potential route guidance strategies.

10.2.2 Future research

Firstly, it appears interesting to seek for a way to mathematically analyze the complexity of marginal algorithms compared to explicit simulation. This would provide a theoretical comparing ground for the computational benefit of marginal simulation - in addition to the empirical analysis in this thesis. However, this seems very difficult for the marginal algorithms developed in this thesis, for which no analytical expression is readily available[53].
Secondly, additional (empirical) validation of MaC for various applications as mentioned in the previous section constitutes important future research. Furthermore, it is worthwhile to further inquire into the possibilities to deploy MaC (or newly developed marginal simulation algorithms) for other purposes. Foremost, it is interesting to explore the possibility of applying marginal simulation in DUE algorithms that are based on decomposition of the master DUE problem into restricted sub problems - see Nie (2010) for static assignment algorithms that exploit such a decomposition. The restricted sub problems may relate to a set of routes between a single OD pair or to bushes towards a single destination or from a single origin. By definition, finding the equilibrium for the sub problems involves only manipulations to restricted parts of the network. Marginal simulation may be used on these restricted parts to marginally evaluate the effect of route (or departure time) swaps - instead of performing explicit DNL simulations on the full network as in Lu et al. (2009) - and hence to approximate equilibrium in the restricted network. This marginally calculated equilibrium is expected to be a very efficient and close approximation of the optimization direction and step size towards equilibrium in the original model – which gives MaC a role comparable to the local linear approximation of the gradient used in Gentile & Noekel (2009) for the static assignment problem.

To extend and enhance the potential application of marginal simulation, developing new marginal simulation algorithms is of course to be encouraged. Since simulation-based DNL models inherently apply a discretization in space and time, marginal algorithms can be developed that confine the calculations to a (small) part of the network and time period. This will open up new and improved applications for marginal simulation. Indeed, specific marginal algorithms (e.g. multi-class, multi-modal, second-order or microscopic) may be tailored for specific purposes. Finally, a notable advance in DNL would be the development of a marginal algorithm that includes en-route rerouting, as this would greatly improve the realism in many potential applications. Following this observed opportunity, a proof-of-concept implementation of MaC extended with an en-route rerouting model is provided in Appendix G. Ideally, however, such a marginal algorithm should be developed from a different base model that already includes en-route rerouting (e.g. the EVAQ model; see Pel, 2011) instead of adding an en-route rerouting component to the marginal algorithm in a later phase.

[53] Also the works describing marginal simulation techniques for digital circuits (Hwang et al., 1988; Salz and Horowitz, 1989) do not provide such a theoretical analysis, so these cannot serve as a starting point.

Moreover, marginal simulation should not be limited to the field of transportation. It is expected to be beneficial for other domains with similar needs (for instance pedestrian modelling, factory design and supply chain management). In fact, it has turned out that marginal simulation techniques have already been used for the design of digital hardware circuits (Hwang et al., 1988; Salz and Horowitz, 1989). A notable difference with our approach is that these implementations compare the inputs of sub models instead of the model outputs to check for changes between the marginal and base simulations. If the inputs to a sub model are unchanged compared to the base scenario, it is not run, since its outputs would also be the same. It would be interesting to also develop marginal DNL algorithms that have this property. An alternative MaC algorithm in that sense would compare link demands and supplies (i.e. the inputs to the intersection model) and only calculate flows over an intersection in case of a change of these inputs. This would allow performing sensitivity analyses of link flow variations, analogous to route flow variations as is done in Chapter 9. This could further enhance the computational gain for some applications (e.g. optimal control and gradient-based OD estimation) significantly. In this way, marginal DNL simulation could benefit from the experiences in the digital circuit domain. Hopefully, in the future, such transfers of knowledge will rapidly advance marginal simulation applications across different research domains.

A THE LINK TRANSMISSION MODEL

The Link Transmission Model (LTM) is a macroscopic Dynamic Network Loading (DNL) model that combines high realism in the representation of congestion formation and spillback with computational efficiency. It is briefly introduced in this appendix; for more details see Yperman et al. (2006) and Yperman (2007).

Section A.1 familiarizes the reader with Cumulative Vehicle Numbers (CVN). The CVN from the basis of the calculations during an LTM simulation and are also the primary output of LTM. Section A.2 briefly explains the basics of the traffic propagation on links according to first-order traffic flow theory with a triangular fundamental diagram. This is the underlying theory of the link model in LTM. The intersection model is thoroughly discussed in the main text of this dissertation. The brief explanation in Section A.3 of this appendix is limited to the steering role of the intersection model in the LTM algorithm.

A.1 Cumulative Vehicle Numbers

N(x,t) expresses the CVN that have passed location *x* by time *t*. In LTM, CVN are calculated and stored at all upstream and downstream link boundaries. Since LTM adopts a Multi-Commodity (MC) approach where each commodity corresponds to a specific predefined route between an origin and a destination, CVN curves are composed for the total traffic streams on a link as well as for each route commodity. These CVN curves form the basis of the calculations (see Section A.2) and the main output of LTM.
When post-processing the CVN output, link travel times can be derived as the horizontal distance between the curves of $N(x^0,t)$ (at the upstream link boundary) and $N(x^L,t)$ (at the downstream link boundary). In the example in Figure A-1, vehicle *h* experiences travel time *tt(h)*. Analogously, route travel times can be derived from the route CVN in the origin and destination link. This determination of link travel times requires first-in-first-out (FIFO) behaviour on each network link, which is ensured by the LTM algorithm. The vertical distance between the CVN curves represents the number of vehicles $\Delta N(t)$ on the link at

time *t*. Additional information such as flow, density, Total Time Spent (TTS) and Vehicle Hours Lost (VHL) can be calculated from the CVN as well. The flow *q* (veh/h) is simply the slope of the CVN curve. The density *k* (veh/km) on a link can easily be calculated by dividing the number of vehicles ΔN by the length *L*. The TTS of all vehicles on a link is calculated as the integral between the upstream and downstream CVN curve. The VHL is the excess (or reduction in case of negative VHL) in TTS compared to a base reference, usually the free flow scenario.

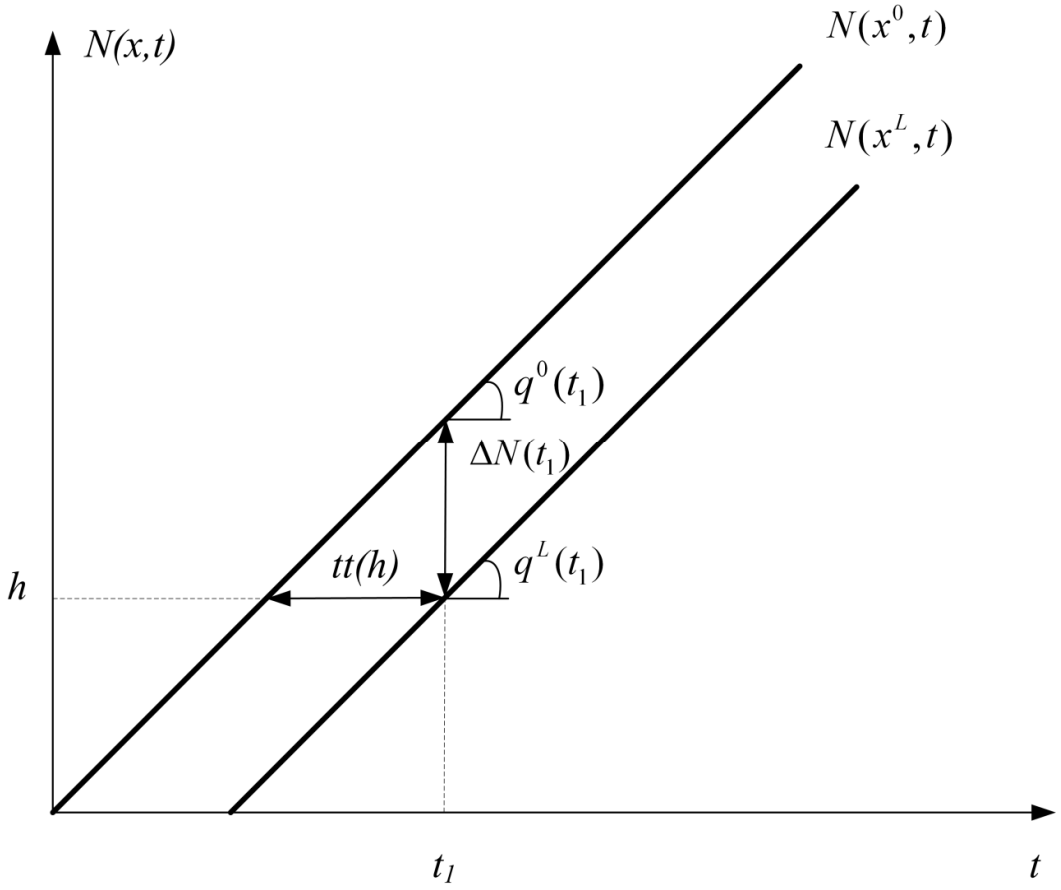

Figure A-1: Cumulative vehicle numbers

A.2 Traffic propagation on links

In LTM, traffic propagation on the unidirectional links is consistent with first-order traffic flow theory. The fundamental diagram is assumed to be triangular (Figure A-2). In free flowing conditions – i.e. densities below the critical density k_c that corresponds to the capacity flow *C* - vehicles are assumed to travel with a fixed free flow speed v_f (km/h). The speed *v* of congested traffic states ($k > k_c$) is given by *q/k*. The maximum or jam density k_j thus corresponds to vehicles standing still. The characteristic waves of traffic states move through the links of the network with a wave speed *dq/dk*. Different traffic states are separated by shock waves which may propagate up- or downstream with a speed w_s. This

shockwave speed w_s can be derived as the slope of the connecting line between the adjoining traffic states, as illustrated in Figure A-2. The maximum negative shockwave speed w is the fastest possible speed with which congestion may spill back in the upstream direction.

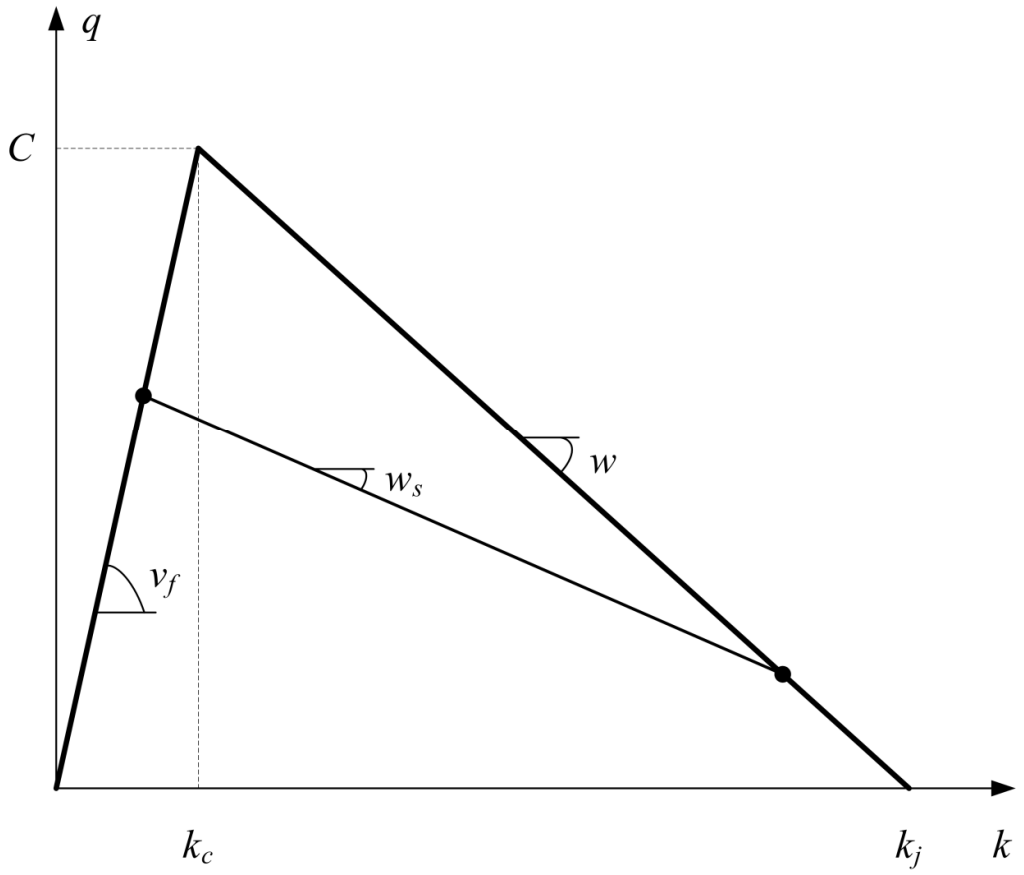

Figure A-2: Triangular fundamental diagram

Now, we will briefly explain how the above can be captured in CVN according to the simplified solution procedure for first-order traffic flow theory with a triangular fundamental diagram of Newell (1993). In this thesis, it suffices to understand that the constraints on the flows at link boundaries can be derived from the CVN calculated by the simulation at earlier time steps. Firstly, as vehicles and downstream moving traffic states cannot travel faster than v_f, each established point[54] $N(x^0,t)$ on the upstream CVN curve forms a constraint on the downstream CVN a free flow link travel time ($\frac{L}{v_f}$) later. Hence:

$$N(x^L,t) \le N(x^0,t-\frac{L}{v_f}) \qquad \text{(A.1)}$$

[54] By this, we mean calculated and stored earlier in the simulation.

In other words, a vehicle that enters the link at time *t* can leave the link no sooner than $t+\frac{L}{v_f}$. Secondly, the time it takes for congested traffic states and congestion spillback waves to spill back from the downstream to the upstream link boundary is at least $\frac{L}{-w}$. During this time, at most k_jL vehicles may have entered the link. A higher number could cause the maximum possible density k_j to be exceeded somewhere on the link. Hence, an established point on the downstream CVN curve places a constraint on the upstream CVN of the following form:

$$N(x^0, t+\frac{L}{-w}) \leq N(x^L, t) + k_j L \tag{A.2}$$

More details can be found in Yperman (2007) and Newell (1993).

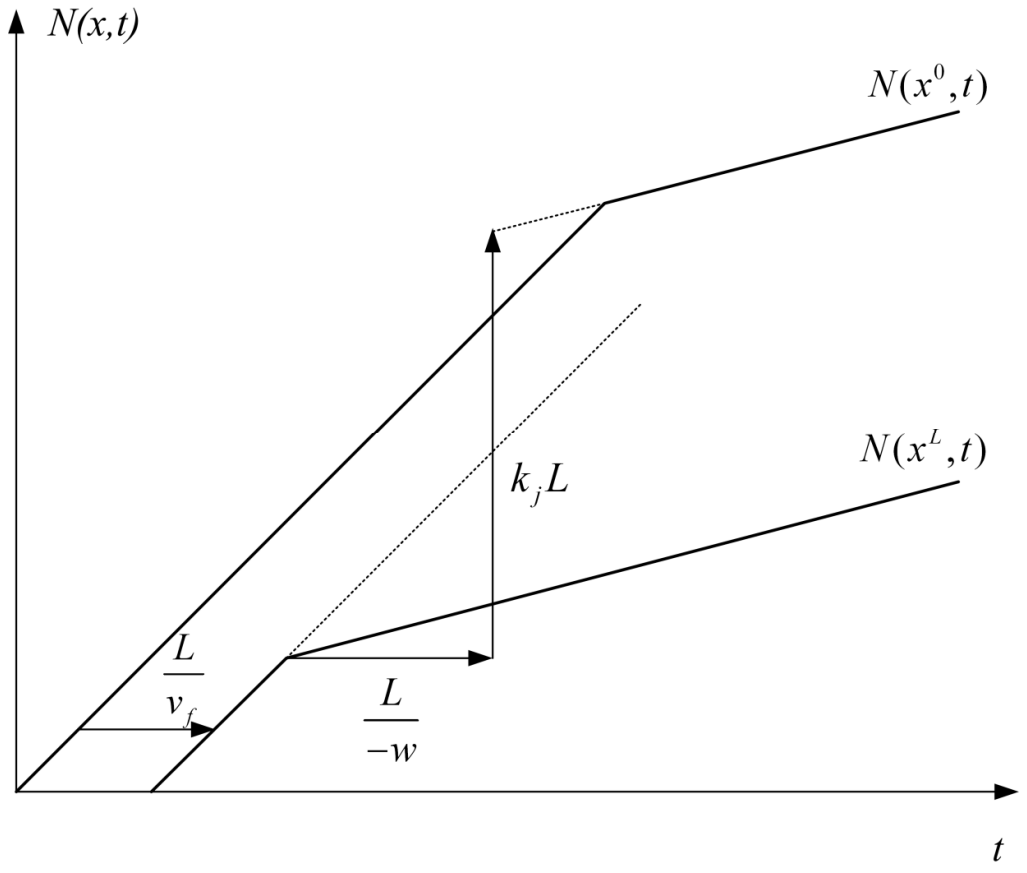

Figure A-3: Propagation of constraints in the CVN curves

According to what is described above, the link model passes constraints to the intersection model. Following (A.1), the maximum possible outflow of an incoming link (i.e the demand constraint *S*) between t and Δt can be computed as:

$$S(t) = \min((N(x^0, t + \Delta t - \frac{L}{v_f}) - N(x^L, t)), C\Delta t) \quad \text{(A.3)}$$

The second term in (A.3) stems from the fact that the outflow cannot exceed the link's capacity *C*. Based on (A.2), the maximum possible inflow into an outgoing link (the supply *R*) can be calculated as:

$$R(t) = \min((N(x^L, t + \Delta t - \frac{L}{-w}) + k_j L - N(x^0, t), C\Delta t) \quad \text{(A.4)}$$

Then, it is the task of the intersection model to determine the resulting flows based on these constraints.

A.3 The intersection model at the heart of the LTM algorithm

The LTM simulation algorithm progresses by calculating discretized flows *q* during update time steps Δt. This intersection-dependent Δt is fixed to the shortest free flow travel time of any of the adjacent links. This maximizes each Δt given the Courant-Friedrichs-Lewy (CFL) condition (Courant et al., 1928).
The flows *q* during Δt from each incoming link *i* to each outgoing link *j* are calculated by the intersection model. How this is done based on the demands S_i and supplies R_j - determined by the link model as described in the previous section - is extensively discussed in this thesis. The intersection model applied in all case studies involving LTM in this thesis is the oriented capacity proportional model presented in Chapter 4[55]. LTM also implements more sophisticated intersection models (see Yperman, 2007) that add internal supply constraints and delay in under-saturated conditions. These are briefly discussed in Section 5.1.2.1, but not used elsewhere in this thesis.

Once the flows *q* are computed, the CVN are updated according to (A.5). This is done after each intersection update for all the adjacent link boundaries, i.e. the downstream boundaries of incoming links and the upstream boundaries of the outgoing links. Equation (A.5) is only written for the CVN of the total flows; the CVN of each route commodity are updated analogously.

$$N(x, t + \Delta t) = N(x, t) + q\Delta t \quad \text{(A.5)}$$

This procedure is carried out for each intersection and for each update step until *t* reaches the end of the simulation period.

[55] Originally, LTM implemented a demand proportional intersection model. This has been replaced by the oriented-capacity proportional model of Chapter 4.

B

SINGLE-COMMODITY DYNAMIC NETWORK LOADING

As explained in Appendix A, LTM (Yperman, 2007) is a Multi-Commodity (MC) DNL model, with each commodity representing a (predefined) route. This means that LTM keeps track of separate route flows on the links. Single-Commodity (SC) DNL models have been proposed by Gentile et al. (2007), Taale (2008) and Blumberg & Bar-Gera (2009). Also both marginal DNL models introduced in this thesis adopt are SC. The advantage of the SC approach is that it is much more efficient for large-scale networks carrying a high number of routes; in terms of computation time as well as memory usage. On the other hand, obtaining a fully consistent DNL simulation requires some iteration, as explained in this appendix.

The difference between the MC and SC approach to DNL can be understood by means of the illustrations in Figure B-1. In a MC DNL, the flow composition in terms of route flows is known on the links. From this, the turning fractions f_{ij} at link ends can be calculated at the link level by summing up the route flows per outgoing link.

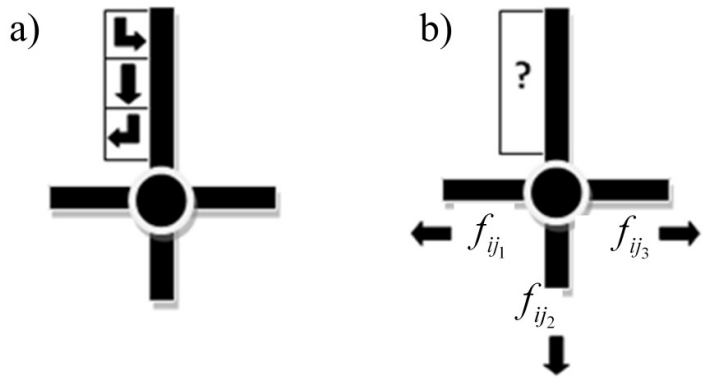

Figure B-1: Schematic illustration of MC (a) and SC (b) approach

In a SC DNL, instead of considering each route flow individually, traffic is propagated on the links as one homogenous flow. Exogenously defined turning fractions f_{ij} guide the propagation of traffic through the network by dividing the total flow on each incoming link over the various outgoing links after the intersection.

Unlike in a MC DNL model, route choice made at departure is not explicitly accounted for during a SC DNL simulation[56]. Therefore, consistency needs to be achieved through an iterative procedure that adjusts the turning fractions until they, together with the travel times, correctly correspond to the route choice at departure. This route choice may be updated via the choice component in a DTA model (see Section 1.1.1.1) or assumed to be fixed. Figure B-2 depicts this iterative process for the latter case.
First, the link travel times are initialized, e.g. assuming free flow. Also, the predefined OD matrix and route choice – yielding predefined route demands – are loaded as input. Then, the turning fractions at link ends are derived from the propagation of the dynamic route demands according to the dynamic link travel times. From this, the route flows at each downstream link boundary can be calculated as a function of time. By adding up the route flows that are headed towards the same outgoing link, discretized dynamic turning fractions are obtained at each downstream link boundary. Hereby, a turning fraction interval T is used during which the turning fractions are constant. This time interval T could be intersection-, link- or even time-dependent. Throughout this thesis, however, T is assumed fixed for the whole network and simulation period.
The SC DNL model then uses the turning fractions to propagate the traffic flows through the network. The SC DNL simulation yields new link travel times influenced by congestion (and intersection delays, if implemented). The turning fractions are then recalculated according to these new link travel times and the route flows, following the same procedure as described above. When the link travel times have converged, a consistent SC DNL according to the predefined route choice is obtained. For more details on this iterative procedure, we refer to Blumberg & Bar-Gera (2009)[57].
Nastase & Harehdasht (2010) report a fast convergence for these SC DNL iterations on small dummy networks (typically 2-3 iterations), using a SC variant of LTM. Blumberg & Bar-Gera (2009) report a similar speed of convergence for their SC model. Furthermore, Nastase & Harehdasht (2010) show that the number of required iterations rises with increasing volatility of the turning fractions. Also, they demonstrate that an iterated SC DNL approaches a MC DNL[58] as the turning fraction interval T is decreased. Further research into the properties and speed of the convergence is desirable, however.

[56] In case of en-route route choice - as in reactive assignment (see Section 1.1.1.2) – the turning fractions could be changed during the SC DNL simulation.
[57] Blumberg & Bar-Gera (2009) use arrival orders instead of turning fractions, but otherwise the procedure is identical.
[58] Of course, the comparison has to be made between models with exactly the same modeling principles. In Nastase & Harehdasht (2010), the traditional (MC) LTM is compared to a SC LTM variant.

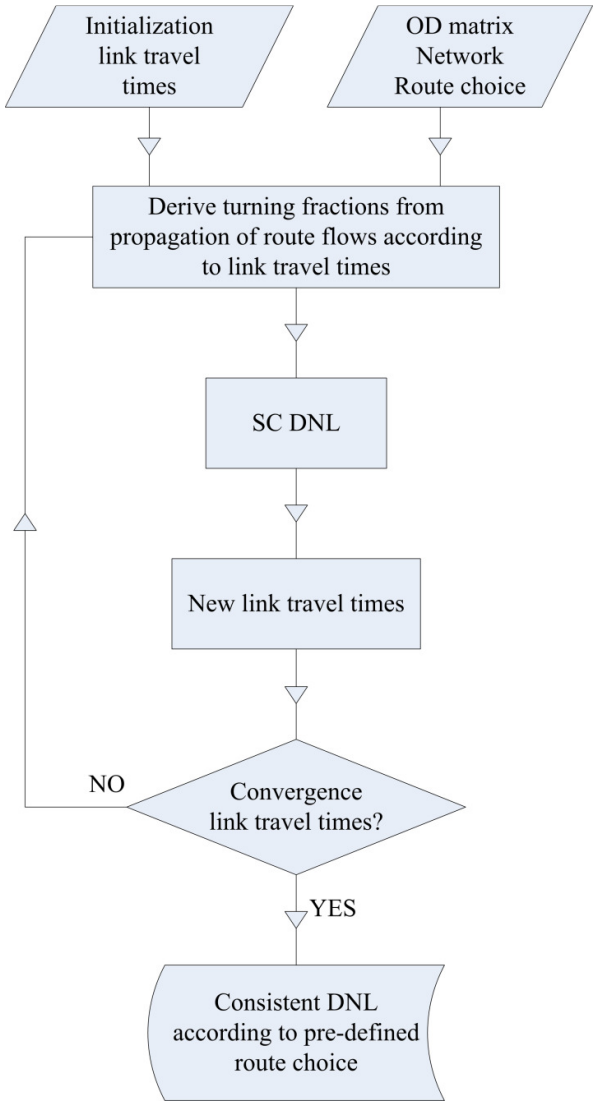

Figure B-2: Flowchart SC DNL iteration loop

Finally, different model frameworks are possible in an equilibrium DTA model. The SC iterations that update the turning fractions may form an inner loop, or may be integrated in the DTA iterations. It should be noted that the results in Nastase & Harehdasht (2010) suggest that this, in particular the number of SC iterations that are performed before each DTA iteration, may influence both the speed of the DTA convergence as the equilibrium route choice that is obtained.

To save computation time, however, the iterative procedure to update the turning fractions is omitted in the marginal algorithms developed in this thesis. Instead, the turning fractions that result from the base simulation are used for the marginal simulations as well (see e.g. Section 9.1).

C

COMPLIANCE OF THE MODEL OF CHAPTER 4 WITH THE REQUIREMENTS OF SECTION 3.4.1

In this appendix, it is shown that the intersection model presented in Chapter 4 (using the formulation of Section 4.1) satisfies all requirements for first-order macroscopic DNL intersection models postulated in Section 3.4.1.

1. General applicability to any number of incoming and outgoing links and any combination of boundary conditions

As all definitions are formulated generally without any restrictions of the number of incoming or outgoing links, or of the values of the boundary conditions, this requirement is trivially guaranteed.

2. Non-negativity of flows

Let us assume a valid link model that passes boundary conditions $S_i \geq 0$, $f_{ij} \geq 0$, $R_j \geq 0$ and $C_i \geq 0$. As all $\alpha_i > 0$ by definition, (4.9) only contains positive variables so that $q_i \geq 0 \quad \forall i$.

3. Conservation of vehicles

This is guaranteed by unambiguously defining the solution of the intersection model in (4.9) by single variables q_i for each i. All partial and outgoing flows are then derived via given f_{ij} (see (3.16)).

4. Compliance with the demand and supply constraints

From (4.9) results immediately that $q_i \le S_i \quad \forall i$.
Compliance with the supply constraints is demanded by definition of the sets (4.4).

5. Ensuring FIFO: Conservation of turning fractions (CTF)

As the solution is defined in terms of the q_i and $q_{ij} = f_{ij}q_i$ with fixed, given f_{ij} derived by the link model or exogenously, this requirement is always guaranteed.

6. Flow maximization from the users' perspective

Individual flow maximization requires that each q_i is either actively limited by demand or by a supply that is fully consumed (i.e. $\hat{R}_j = 0$). Definition (4.5) states that either $q_i = S_i$ or $i \in U_j \Leftrightarrow q_i = \frac{R_j^i}{f_{ij}}$. If $q_i = S_i$ then q_i is obviously actively limited. In the latter case, it follows from (4.9) that $f_{ij}q_i \le R_j^i \quad \forall i, j \,|\, U_j \ne \varnothing \Leftrightarrow \sum_i f_{ij}q_i \le \sum_i R_j^i \quad \forall j \,|\, U_j \ne \varnothing$. Then, it follows from (4.4) and (4.8) that $\sum_{i \in U_j} f_{ij}q_i = \sum_{i \in U_j} R_j^i = \widetilde{R}_j$ so that (with (4.7)) $\sum_{i \in U_j} f_{ij}q_i + \sum_{i \notin U_j} f_{ij}q_i = R_j$.
Hence, the active supply constraints are fully consumed so that q_i is also actively limited if it is supply constrained.

7. Compatibility with link traffic flow dynamics: compliance with the invariance principle

From (4.4) and (4.5) follows that $q_i < S_i \Leftrightarrow \exists j \,|\, q_i = \frac{R_j^i}{f_{ij}}$. Furthermore, R_j^i is independent of S_i ((4.7)-(4.8)) and $q_i < S_i \le C_i$. Hence, q_i is invariant to replacing S_i by C_i if $q_i < S_i$.
If $q_j < R_j \Leftrightarrow U_j = \varnothing$ (4.6), then R_j does not appear in (4.9). Hence, q_j is invariant to replacing R_j by C_j if $q_j < R_j$.

D

EXACTNESS AND CONVERGENCE OF THE SOLUTION ALGORITHM OF SECTION 4.3

First, it is proven that the algorithm presented in Section 4.3 finds the exact solution of the oriented capacity proportional intersection model of Section 4.2 (which is an instance of the general intersection model of Section 4.1). Then, it is shown that this solution is found in at most *I* iterations; *I* being the number of incoming links.

D.1 Exactness of the solution

It can easily be derived that the algorithm in Section 4.3 yields the exact same solution as the model presented in Section 4.2. The question that needs to be answered is "does the algorithm produce a_j as defined by (4.15)?". If it does, the exact solution follows immediately from (4.16) and (4.17).

It can be shown that the a_j can only increase (or be set to 1 if $U_j = \varnothing$) as they are recomputed by the algorithm (i.e. from $k \to k+1$). Therefore, we consider steps 4(a) **-** at least one *i* is demand constrained **-** and 4(b) **-** all $i \in U_{\hat{j}}^{(k)}$ are constrained according to $a_{\hat{j}}^{(k)}$ **-** of the algorithm in Section 4.3 separately. We will demonstrate this for the case $k = 0 \to 1$; the general case $k \to k+1$ is entirely analogous.

In step 4(a), some link(s) i are demand constrained and the following procedures are carried out:

- $\widetilde{R}_j^{(1)} = \widetilde{R}_j^{(0)} - \sum_{i|q_i=S_i} S_{ij} \quad \forall j \in J^{(0)}$ (reducing the numerator of the a_j)
- $U_j^{(1)} = U_j^{(0)} \setminus \{i \mid q_i = S_i\} \quad \forall j \in J^{(0)}$ (reducing the denominator of the a_j with $\sum_{i|q_i=S_i} C_{ij}$)

The algorithm then returns to step 3 (unless all $U_j = \varnothing$) and the a_j are recomputed. The a_j for the next iteration become:

$$\begin{aligned} a_j^{(1)} &= \frac{\widetilde{R}_j^{(1)}}{\sum_{i \in U_j^{(1)}} C_{ij}} = \frac{\widetilde{R}_j^{(0)} - \sum_{i|q_i=S_i} S_{ij}}{\sum_{i \in U_j^{(0)}} C_{ij} - \sum_{i|q_i=S_i} C_{ij}} && \text{if } U_j^{(1)} \neq \varnothing \\ a_j^{(1)} &= 1 && \text{if } U_j^{(1)} = \varnothing \end{aligned} \tag{D.1}$$

In (D.1), the ratio of what is deducted from the numerator over what is deducted from the denominator is smaller than $a_j^{(0)}$ for all j that are still under consideration, as is shown by:

$$\forall i \mid q_i = S_i, \forall j \in J^{(0)} : S_{ij} \leq a_{\hat{j}}^{(0)} C_{ij} \;\&\; a_{\hat{j}}^{(0)} \leq a_j^{(0)} \Rightarrow \frac{\sum_{i|q_i=S_i} S_{ij}}{\sum_{i|q_i=S_i} C_{ij}} \leq a_{\hat{j}}^{(0)} \leq a_j^{(0)} \qquad \forall j \in J^{(0)} \tag{D.2}$$

Since the denominator is relatively stronger reduced than the numerator, each a_j (that is not set to 1) can only increase, i.e. $a_j^{(1)} \geq a_j^{(0)} \quad \forall j \in J^{(0)}$.

In step 4(b), all links $i \in U_{\hat{j}}^{(0)}$ are able to consume their oriented capacity proportional share and the following procedures are carried out:

- $\widetilde{R}_j^{(1)} = \widetilde{R}_j^{(0)} - \sum_{i \in U_{\hat{j}}^{(0)}} a_{\hat{j}}^{(0)} C_{ij} \quad \forall j \in J^{(0)} \setminus \{\hat{j}^{(0)}\}$ (reducing the numerator of the a_j)
- $U_j^{(1)} = U_j^{(0)} \setminus \{i \in U_{\hat{j}}^{(0)}\} \quad \forall j \in J^{(0)} \setminus \{\hat{j}^{(0)}\}$ (reducing the denominator of the a_j with $\sum_{i \in U_j^{(0)}} C_{ij}$)

Only the links $i \notin U_{\hat{j}}^{(0)}$ (i.e. $i \mid f_{i\hat{j}} = 0$) remain in other sets $U_j^{(1)}$. The following a_j for the next iteration result:

$$a_j^{(1)} = \frac{\widetilde{R}_j^{(1)}}{\sum_{i \in U_j^{(1)}} C_{ij}} = \frac{\widetilde{R}_j^{(0)} - \sum_{i \in U_{\hat{j}}^{(0)}} a_{\hat{j}}^{(0)} C_{ij}}{\sum_{i \in U_j^{(0)}} C_{ij} - \sum_{i \in U_{\hat{j}}^{(0)}} C_{ij}} \quad if\ U_j^{(1)} \neq \varnothing$$
$$a_j^{(1)} = 1 \qquad if\ U_j^{(1)} = \varnothing \tag{D.3}$$

Analogous to the previous argumentation we find that all a_j (that are not set to 1) can only increase:

$$\forall j \in J^{(0)} \setminus \{\hat{j}^{(0)}\} : \frac{\sum_{i \in U_{\hat{j}}^{(0)}} a_{\hat{j}}^{(0)} C_{ij}}{\sum_{i \in U_{\hat{j}}^{(0)}} C_{ij}} = a_{\hat{j}}^{(0)} \leq a_j^{(0)} \Rightarrow a_j^{(1)} \geq a_j^{(0)} \quad \forall j \in J^{(0)} \setminus \{\hat{j}^{(0)}\} \tag{D.4}$$

Therefore, the currently smallest $a_{\hat{j}}^{(0)}$ can be fixed and $j = \hat{j}^{(0)}$ is not considered in further iterations. Analogously, it can be shown that the above is true for each step $k \rightarrow k+1$. Hence, the a_j can only increase (or be set to 1) as the algorithm proceeds. Therefore, it is possible to fix $a_{\hat{j}}^{(k)}$ and $U_{\hat{j}}^{(k)}$ when step 4(b) of the algorithm is reached in iteration *k*. We will now show that each resulting a_j fulfils (4.15).

Since the smallest $\alpha_{\hat{j}}$ over all *j* can only be influenced by the demand constraints – as all other supply constraints are less restrictive -, this $\alpha_{\hat{j}}$ and the corresponding $U_{\hat{j}}$ of the model solution can be written as:

$$a_{\hat{j}} = \frac{R_{\hat{j}} - \sum_{i | S_i \leq a_{\hat{j}} C_i} S_{i\hat{j}}}{\sum_{i \in U_{\hat{j}}} C_{i\hat{j}}} \quad with\ U_{\hat{j}} = \{i \mid f_{i\hat{j}} > 0\ \&\ S_i > a_{\hat{j}} C_i\} \tag{D.5}$$

Since the a_j's can only increase as the algorithm proceeds, the smallest $\alpha_{\hat{j}}$ (and the corresponding $U_{\hat{j}}$) will be the first to be fixed in the algorithm. This means that step 4(b) is entered for the first time with this $\hat{j}$ as the restrictive outgoing link.

From the initialization, $\widetilde{R}_{\hat{j}}^{(0)} = R_{\hat{j}}$ and $U_j^{(0)} = \{i \mid f_{i\hat{j}} > 0\}$. If necessary, demand constrained links ($S_i < a_{\hat{j}} C_i$) are removed from $U_{\hat{j}}$ and the $\alpha_{\hat{j}}$ are updated according to (D.1) in step 4(a) in the first iteration(s). This implies that, when step 4(b) is entered for $\hat{j}$, $\alpha_{\hat{j}}$ and $U_{\hat{j}}$ are fixed to the exact value given by (D.5), since (D.3) cannot have been carried out for $\hat{j}$.

For subsequent j (denoted as j_x), the algorithm fixes α_{j_x} and U_{j_x} if step 4(b) is entered with $\hat{j}^{(k)} = j_x$. It follows from (D.1) and (D.3) that these α_{j_x} and U_{j_x} can be written as:

$$a_{j_x} = \frac{\widetilde{R}_{j_x}^{(0)} - \sum_{i|q_i=S_i} S_{ij_x} - \sum_{z=1}^{x-1} \sum_{i \in U_{j_z}} a_{j_z} C_{ij_x}}{\sum_{i \in U_{j_x}^{(0)}} C_{ij_x} - \sum_{i|q_i=S_i} C_{ij_x} - \sum_{z=1}^{x-1} \sum_{i \in U_{j_z}} C_{ij_x}} = \frac{R_{j_x} - \sum_{i|q_i=S_i} S_{ij_x} - \sum_{z=1}^{x-1} \sum_{i \in U_{j_z}} a_{j_z} C_{ij_x}}{\sum_{i \in U_{j_x}} C_{ij_x}} \qquad \text{(D.6)}$$

$$with: \quad U_{j_x} = U_{j_x}^{(0)} \setminus \{ i \mid q_i = S_i \ \bigcup \ i \in U_{j_z} \} \quad \forall z = 1..x-1$$

Herein, a_{j_z} ($z = 1..x-1$) are the smaller a_j that have been fixed in previous iterations; U_{j_z} are the corresponding sets. Therefore, if these a_{j_z} and U_{j_z} have been calculated exactly, (D.6) matches (4.15) exactly. Since it was shown that the solution is exact for the smallest $a_{\hat{j}}$ that was calculated first, all subsequent a_j are also exact and the algorithm finds the exact model solution.

D.2 Maximum number of iterations

The algorithm is finished as soon as the set of j still to be considered in the next iteration is empty, i.e. $J^{(k+1)} = \varnothing$. A j is removed from $J^{(k)}$ as soon as U_j is definitively determined, i.e. when there is no i left that needs to be removed from it. The stop criterion is therefore equivalent to stating that all i need to be definitively appointed to one U_j or be removed from all U_j. In each iteration of the algorithm, at least one i is removed from all U_j in step 4(a) or definitively appointed to one in step 4(b). Therefore, the maximum number of iterations is equal to I.

E

PROOF OF UNIQUENESS CONDITION (5.13)

This appendix first proves that (5.13) is a sufficient condition for solution uniqueness of the intersection model. Then, it is demonstrated that it is also a necessary condition for the vast majority of real-world intersection topologies, in which the flows of at least two incoming links are mutually dependent in at least two (internal) supply constraints.

E.1 Proof of sufficiency

Only the ratio's of the priority parameters α_{ij} and α_{ik} are used to find the intersection model's solution. Hence, it is trivial to see that the solution is independent of β_j and β_k in (5.13) and that all ratio's are identical:

$$\frac{\alpha_{ij}}{\alpha_{i'j}} = \frac{\alpha_i \cancel{\beta_j}}{\alpha_{i'} \cancel{\beta_j}} = \frac{\alpha_i \cancel{\beta_k}}{\alpha_{i'} \cancel{\beta_k}} = \frac{\alpha_{ik}}{\alpha_{i'k}} \quad \forall i, i' \forall j, k \in F_i \bigcap F_{i'} \tag{E.1}$$

As such, the solution is the same for any β_j and β_k. Hence, it suffices to prove that having single-valued priority parameters α_i is a sufficient condition for solution uniqueness of the intersection model.

For this proof, let us assume that we have found a solution that fits the model definitions (6.1) -(6.4) (with $\alpha_{ij} = \alpha_{ik} = \alpha_i \quad \forall i, j, k$). We indicate this "starting solution" by adding the superscript 'old' to the variables. We will show that a transition from this starting solution to another solution (indicated by 'new') in which $\exists i \mid q_i^{old} \neq q_i^{new}$ is impossible. For this, we start from the assumption that the transition is initiated by some flow increasing (i.e. $\exists i \mid q_i^{old} < q_i^{new}$). The proof assuming a decreasing flow is entirely analogous. Also, only

internal conflicts *k* are included in the proof. Any *k* could be replaced by a *j*, yielding an equivalent proof.

Firstly, let us denote $i \mid q_i^{old} < q_i^{new}$ as i_n. Now, $q_{i_n}^{old} < q_{i_n}^{new} \Leftrightarrow q_{i_n}^{old} < S_{i_n} \Rightarrow \exists k_n \mid i_n \in U_{k_n}^{old}$, otherwise the new solution would violate the demand constraint. Since $i_n \in U_{k_n}^{old} \Leftrightarrow \widehat{N}_{k_n}^{old} = 0$, some other flow[59] competing for N_{k_n} must decrease to prevent the internal supply constraint from being exceeded as q_{i_n} increases, i.e.:

$$q_{i_n}^{old} < q_{i_n}^{new} \Leftrightarrow \exists i_{n+1} \mid f_{i_{n+1}k_n} > 0 : q_{i_{n+1}}^{old} > q_{i_{n+1}}^{new} \tag{E.2}$$

From (E.2), it follows that:

$$i_n \in U_{k_n}^{old} \Leftrightarrow \frac{q_{i_n}^{old}}{q_{i_{n+1}}^{old}} \geq \frac{\alpha_{i_n}}{\alpha_{i_{n+1}}} \overset{\frac{q_{i_n}^{old}}{q_{i_{n+1}}^{old}} < \frac{q_{i_n}^{new}}{q_{i_{n+1}}^{new}}}{\Rightarrow} \frac{q_{i_n}^{new}}{q_{i_{n+1}}^{new}} > \frac{\alpha_{i_n}}{\alpha_{i_{n+1}}} \tag{E.3}$$

Furthermore, it is clear that i_{n+1} must be supply constrained in the new solution:

$$q_{i_{n+1}}^{old} > q_{i_{n+1}}^{new} \Leftrightarrow q_{i_{n+1}}^{new} < S_{i_{n+1}} \Rightarrow \exists k_{n+1} \mid i_{n+1} \in U_{k_{n+1}}^{new} \overset{\frac{q_{i_n}^{new}}{q_{i_{n+1}}^{new}} > \frac{\alpha_{i_n}}{\alpha_{i_{n+1}}}}{\Rightarrow} f_{i_n k_{n+1}} = 0 \tag{E.4}$$

This means that – because of (E.3) – i_n cannot be a competitor for $N_{k_{n+1}}$. Now, since $i_{n+1} \in U_{k_{n+1}}^{new} \Leftrightarrow \widehat{N}_{k_{n+1}}^{new} = 0$, some other flow competing for $N_{k_{n+1}}$ must increase so that the internal supply constraint in k_{n+1} remains active despite the decreasing $q_{i_{n+1}}$, i.e.:

$$q_{i_{n+1}}^{old} > q_{i_{n+1}}^{new} \Leftrightarrow \exists i_{n+2} \mid f_{i_{n+2}k_{n+1}} > 0 : q_{i_{n+2}}^{old} < q_{i_{n+2}}^{new} \tag{E.5}$$

From (E.5), it follows that:

$$i_{n+1} \in U_{k_{n+1}}^{new} \Leftrightarrow \frac{q_{i_{n+1}}^{new}}{q_{i_{n+2}}^{new}} \geq \frac{\alpha_{i_{n+1}}}{\alpha_{i_{n+2}}} \tag{E.6}$$

[59] This can be any competing flow. The subscript '*n*' defines the order in which the incoming links appear in the derivation and is entirely unrelated to any predefined numbering or priority order of the incoming links.

Similarly to before:

$$q_{i_{n+2}}^{old} < q_{i_{n+2}}^{new} \Leftrightarrow q_{i_{n+2}}^{old} < S_{i_{n+2}} \Rightarrow \exists k_{n+2} \mid i_{n+2} \in U_{k_{n+2}}^{old} \overset{\frac{q_{i_{n+1}}^{old}}{q_{i_{n+2}}^{old}} > \frac{q_{i_{n+1}}^{new}}{q_{i_{n+2}}^{new}} \geq \frac{\alpha_{i_{n+1}}}{\alpha_{i_{n+2}}}}{\Rightarrow} f_{i_{n+1}k_{n+2}} = 0 \tag{E.7}$$

Now, it is clear that i_{n+2} is of the same 'type' as i_n, by which we mean that both of these links are supply constrained in the starting solution and increase from the old solution to the new one. This implies that taking the derivation any number of steps further leads to, by combination of (E.3) and (E.6):

$$\frac{q_{i_n}^{new}}{q_{i_{n+1}}^{new}} \geq \frac{\alpha_{i_n}}{\alpha_{i_{n+1}}} \quad \forall n \tag{E.8}$$

And:

$$\begin{aligned} &\frac{q_{i_n}^{new}}{q_{i_m}^{new}} > \frac{\alpha_{i_n}}{\alpha_{i_m}} \quad \forall n, m \\ &with: \quad m > n+1 \end{aligned} \tag{E.9}$$

Now, given a finite number of incoming links, at some point one of the following options occurs, disproving the transition from the old solution to the new one:

- There does not exist an i_{n+1} (or i_{n+2}) as in (E.2) (or (E.5)), which means that the flow q_{i_n} (or $q_{i_{n+1}}$) cannot be changed, meaning the previous one cannot be changed, etc.
- Some i exists twice in the derivation, such that $i_m = i_n$. Obviously: With (E.8), this implies that:

$$i_m = i_n \Rightarrow \frac{q_{i_n}^{new}}{q_{i_m}^{new}} = \frac{\alpha_{i_n}}{\alpha_{i_m}} = 1 \tag{E.10}$$

  Now, it is clear that $m > n + 1$, since subsequent q_i in the above derivation always change reversely (one increases and the other decreases). With $m > n + 1$, however, (E.10) contradicts (E.9).

In summary, assuming single-valued α_i, the above derivation leading from the starting solution to the new one is impossible. Hence, (5.13) is a sufficient condition for solution uniqueness of the intersection model.

E.2 Proof of necessity for mutually dependent flows

In the following, we will show that for the case of mutually dependent flows, specific circumstances exist that render the solution non-unique for any set of α's that does not meet (5.13). Note hereby that the supply constraint functions $\widehat{R}_j$ and $\widehat{N}_k$ only depend on boundary conditions such as the given (internal) supply and the turning fractions. Therefore, they are independent of the model parameters α_{ij} and α_{ik}.

Assume any arbitrary, strictly positive α_{ik}, $\alpha_{i'k}$, $\alpha_{ik'}$ and $\alpha_{i'k'}$ for two incoming links *i* and *i'* that are mutually dependent in two internal supply constraints $\widehat{N}_k$ and $\widehat{N}_{k'}$. Cases including external supply constraints are entirely analogous. The demand constraints are assumed arbitrarily high so that they never bind. Also other supply constraints are disregarded. As supply constraint functions are always continuous and decreasing, a setting is always conceivable so that (5.14) holds for these functions. Furthermore, the boundary conditions can be chosen such that the intersection point lies between the two priority ratios as in Figure E-1.

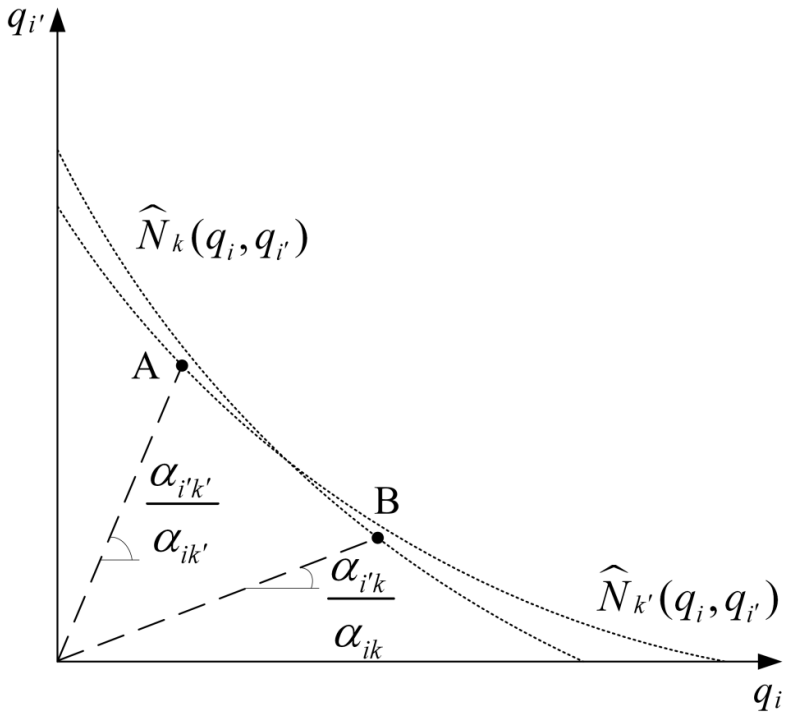

Figure E-1: Construction of non-unique solutions

Now, two solutions A and B exist that are both described by (5.5)-(5.8). Enforcing condition (5.13) leads to $\frac{\alpha_{ik}}{\alpha_{i'k}} = \frac{\alpha_{ik'}}{\alpha_{i'k'}} \Rightarrow \frac{q_i^A}{q_{i'}^A} = \frac{q_i^B}{q_{i'}^B}$ and, given that one of the internal supply constraints must bind, a unique solution. If (5.13) does not hold, an ambiguous setting as above can

always be constructed. Therefore, it is a necessary condition for solution uniqueness for any two mutually dependent flows.

Moreover, it appears that the model can become insolvable if $\widehat{N}_k$ and $\widehat{N}_{k'}$ were to coincide. In this case, (5.8) should hold for both *k* and *k'* in both A and B, which is impossible unless the priority ratios coincide. So again, this issue is only resolved if condition (5.13) holds.

Finally, we have limited the statement of (5.13) as a necessary condition to the simple case of two directly mutually dependent flows. We note that there exist cases of indirectly or circularly mutually dependent flows that may also cause solution non-uniqueness. This circular mutual dependency can be explained as follows: multiple flows $q_{i_1}, \ldots, q_{i_m}$ are tied in (internal) supply constraints that each bind (at least) two different flows in a circular sequence. In Figure E-2, an example of such a circular dependency is given. Four internal conflicts are present, governed by Priority-To-The-Right (PTTR). Consequently, each flow has priority in one of its conflicts, and has to give way in the other. The solution space is not depicted, but one can intuitively understand that two different solutions exist. In the first solution, q_1 and q_3 take their maximal share of N_5 and N_7 so that q_2 and q_4 are restricted. In the second solution, the distribution of N_6 and N_8 is followed, so that q_2 and q_4 flow maximally and q_1 and q_3 are restricted.

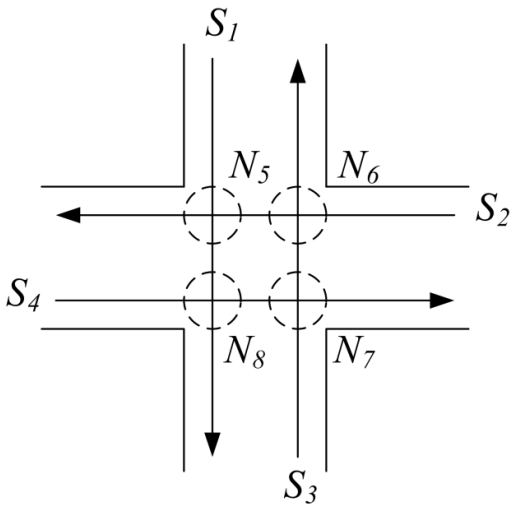

Figure E-2: Circular mutual dependency causing non-uniqueness (regardless of CTF)

It is clear that in this case the CTF requirement is not part of the problem of solution non-uniqueness, as all q_i are unidirectional. Also in this example, solution uniqueness would be obtained by imposing (5.13).

F

TRAVEL TIME VARIABILITY: CURRENT CHALLENGES AND THE ROLE OF MARGINAL SIMULATION

Traditionally, DTA models (see Chapter 1) are applied deterministically, i.e. producing an average or 'typical day' evaluation or prediction. However, numerous studies show that the variability of traffic conditions strongly influences the decisions of road users (see e.g. Lam & Small, 2001; De Palma & Picard, 2005; Hamer et al., 2005). Hence, there is an increasing interest to include (travel time) variability into the planning of network changes and the evaluation of Dynamic Traffic Management (DTM) measures. For this, stochastic evaluations of the network's performance are needed, accounting for the impact of variations in traffic demand and network supply. With traditional DNL simulation models, this can only be achieved by performing a large number of Monte-Carlo simulations with varying input (as in, e.g., Miete, 2011). The resulting high computation times limit the applicability of such an approach. Therefore, modellers have been forced to restrict their studies either in scope (small networks, few scenarios) or in realism by using faster, but less realistic (static or analytical) models.

In this appendix, first the current challenges in Travel Time Variability (TTV) studies are extensively discussed (Section F.1). Then, Section F.2 presents a proof-of-concept case study that exemplifies the potential role of marginal simulation in TTV studies. A stochastic DTA is considered on a dummy network, in which the TTV due to incidents is quantified with MIC (presented in Chapter 8). The influence of this TTV on travellers' route choice is examined in a Probabilistic User Equilibrium (PUE).

This appendix is an edited version of Corthout R., Tampère, C.M.J. & Immers, L.H. (2010). Stochastic Dynamic Network Loading for Travel Time Variability due to Incidents. *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment, pp. 179-198.*

F.1 Travel time variability

From time immemorial, reliability has been a vital issue in the planning, operation and evaluation of economically and socially important systems such as water distribution, health care, communication networks and so on. Reliability in transportation does not go back quite that far. However, this important topic is receiving increasing attention in the past three decades. For road networks, an extensive discussion of previous research and recommended future efforts can be found in Berdica (2002). Also, this work attempts to explain the concepts of reliability and vulnerability (and its causes). Before elaborating on the exact interpretation that will be given to these and related concepts in this dissertation, we further narrow our focus. Three different types of reliability studies are recognized in the literature. The first is connectivity (or terminal) reliability, which is concerned with the probability of whether or not the origins and destinations of the network remain connected and hence whether or not trips can be undertaken or not (see e.g. Wakabayashi & Iida, 1992). In dense road networks like the Western European network, however, connectivity is not an issue. The second type of studies inquires into the reliability of travel times, which is what we will be concerned with in the following. Thirdly, Chen et al. (1999) introduce capacity reliability as a new network performance index. It is defined as the probability that a given demand level can be accommodated by the network (maintaining a desired level of service), which is subject to capacity variations. This includes connectivity reliability as a special case and provides travel time reliability as a side product. However, this third type of reliability is not widely adopted in the state-of-the-art.

The term travel time reliability is often used interchangeably with travel time variability; the latter indicating the reverse of the former. However, Batley et al. (2008) state that this leads to confusion and misinterpretation. Travel time variability in a strict sense implies simply that travel times are not constant, but vary both within-day and from day-to-day (DTD). Reliability intuitively relates to variability on the one hand and information and expectation on the other. In the following, two examples are given to provide more clarity.
The first is adopted from Batley et al. (2008). Say a bus service consistently arrives 10 minutes later than its scheduled arrival time. Hence, there is no travel time variability. However, the reliability depends on the information that is available to users. For regular users, whose main information source is their own experience, the travel time is as expected, and thus reliable. Irregular users, who are informed by the bus timetable, did not expect this longer travel time and will perceive it as unreliable. However, it is the information provided to the irregular users (the bus timetable) which is unreliable in this case, not the travel times.
For the second example, suppose a trip by car that always takes 15 minutes on a Sunday morning, and 90 minutes on a Monday morning. For well-informed, experienced drivers, these travel times, although variable in the strict sense, are perfectly reliable. Again, for less experienced drivers it is not so much the travel times, but rather the available information (for instance based on knowledge of the road type or speed limit, or perhaps completely absent) and, with that, their expectations that are unreliable.
Although neither of the above examples constitute cases that are (nor should be) considered in reliability or variability studies, we conclude from them that reliability is actually (more) closely related to information and expectation (which are user-specific) than to travel time variability – although obviously high variability complicates providing reliable information.

The fact that information may be gained during trips, so that expectations are adjusted and may become more reliable, adds further confusion to the term reliability. Perhaps reliability can best be understood as the extent to which the travel time at any given time and for any given trip is as expected based on the information an individual has. From this point of view, reliability has a subjective character.

We therefore plead in favour of using the term travel time variability, except maybe for studies that specifically examine the effect of information provision under variable conditions (e.g. Arnott et al., 1996; De Palma & Picard, 2006). Still, variability in the strict sense does also not correctly cover the collection of studies that are typically classified as variability (or reliability) studies. Indeed, the second example above – with travel times being variable in the strict sense - is not a situation that should be included in such studies. Rather, variability should be considered specific to a well-defined trip, i.e. on a specific day and departure time, following a specific route and using a specific mode. Travel time variability (TTV) in the remainder thus describes variability to the expected value for the travel time of such a trip, arising from unpredictable, stochastic variations. The expected value is to be interpreted as expected by experienced users or by the modeller – usually the mean travel time is assumed. Unpredictable should be interpreted as describing situations for which it is impossible to provide reliable information prior to their occurrence and are therefore unexpected for all road users and the modeller[60]. These should include incidents and random DTD variations in demand and supply. Variations that are, in principle, predictable such as for instance within-day (peak vs. off-peak), day-of-week and seasonal effects are to be excluded (see Bates et al., 2001 and Batley et al., 2008). Although this may still leave room for debate on what constitutes an unpredictable, random variation and what not, this hopefully delineates the scope of the remaining discussion sufficiently.

Finally, we briefly discuss two related concepts, namely vulnerability and robustness. We regard vulnerability as the extent to which travel times increase (or more generally: performance degrades) as a consequence of unpredictable variations. Usually this is defined on a link level by considering capacity disruptions (incidents, link failures). This can be considered separately from the probability of these disruptions to occur (D'Este and Taylor, 2001). Robustness, which is typically used in reference to whole networks, is then the reverse, namely the ability to cope with disruptions without severe consequences on the network's performance.

To allow the adequate modelling of TTV and the response of drivers with DTA models, advances along the following lines are required:

- Uncovering the probabilistic characteristics of the various sources of TTV (i.e. the unpredictable variations as described above)
- Modelling of the impact on (the variability of) travel times (and other performance measures) of these variations
- Modelling the effect of TTV on drivers' decision making regarding mode, departure time and route

[60] Obviously, it is impossible to provide reliable information in advance of unpredictable variations. Therefore, one might address this again as 'unreliability'. However, the given explanation provides a sharper definition.

In truly comprehensive TTV studies, all of the above elements should be included. However, as each of these topics is already very extensive and challenging on its own, the majority of research in the state-of-the-art is devoted to one specific topic. In the following, these three topics are briefly discussed.

Firstly, uncovering the probabilistic characteristics of the sources of variability is not a trivial task. The probability of an event (e.g. an incident or natural disaster) occurring at a certain time and a certain place needs to be known, as well as the probability distributions of the various characteristics of the event. For an incident for example, information on duration and severity is needed, which may depend on time and location of the incident. This complicates matters even further. To gain such insight requires gathering and studying a lot of data. For other, less prominent sources of variability such as DTD random demand and capacity variations, this is even much more difficult. The fact that the probability of occurrence and the probabilistic characteristics of sources of variability not only depend on the location, but also evolve in time ensures a continuous need for this type of empirical research.
Quite a few empirical studies exist on the probability of incidents and their characteristics, e.g. Smith et al. (2003), Knoop et al. (2008) and Meeuwissen et al. (2004). Research on DTD variations is somewhat scarcer. For example Dervisoglu et al. (2009) study daily variations of highway capacity. Brilon et al. (2005) use a year of empirical data to estimate the capacity distribution functions for two highway sections. Despite this large dataset, however, they could not estimate complete distribution functions. Meeuwissen et al. (2004) study demand variations, namely weather conditions and mass events; but also seasonal effects, which have been argued above to be not unpredictable and therefore to be excluded from TTV studies. Therefore, it is clear that more discussion on the delimitation of what is predictable and what not would be useful. This is indeed debatable; for instance weather conditions, natural disasters and (mass) events[61] are only predictable to some extent. Moreover, this delimitation may obviously change in time, as predictions (for instance weather forecasts) become more accurate and detailed.

Secondly, if the probabilities of the sources of variability are known, the next challenge is to model[62] their effect on the traffic flows and travel times. Typically, the effect of only one source at a time is examined, which is already challenging enough. TTV arising from daily demand fluctuations has been studied by e.g. Bell et al. (1999) and Clark & Watling (2005). Again, however, most studies focus on TTV arising from incidents, e.g. Chen et al. (2002), Inouye (2003) and Ritsema van Eck et al. (2004). A specific type of study is vulnerability analysis, which specifically aims to identify vulnerable links for which a blocking or capacity reduction (e.g. due to an incident) causes a significant increase in travel times; see for instance Murray-Tuite & Mahmassani (2004), Tamminga et al. (2005), Scott et al. (2006) and

[61] Of course, the effects of annually returning mass events are to a large extent predictable. For smaller-scale and/or new events, however, the modeller and users are probably not able to predict the effects. They may even not be aware of the occurrence of the event. Actually, in the latter case, one could argue that the event is essentially a random DTD demand and/or capacity fluctuation.
[62] For some applications, this may be derived empirically instead of with a model (as in Tu et al., 2005). However, many applications involve planning or evaluation of proposed or future measures or infrastructural changes. This of course excludes the possibility of empirical derivation of TTV.

Tampère et al. (2007). Depending on the objective of the study, different quantifications of TTV, e.g. in terms of route TTV, link TTV, or TTS or VHL in the network may be preferred.
To quantify TTV arising from a large set of demand or capacity variations, a traffic assignment model is often used; typically evaluating a large number of Monte-Carlo draws from a stochastic distribution of demand and/or supply (i.e. link capacities). Knoop et al. (2007) show that for this type of study, proper modelling of congestion spillback is vital. This means that macroscopic simulation-based DNL models consistent with first- or second-order traffic flow theory are well-suited for TTV studies. Congestion formation and spillback is far less realistically (or not at all) modelled in static models, analytical models, or simulation-based models with link exit or performance functions (see Section 1.1.2). Microscopic models on the other hand have the disadvantage of requiring a higher computation time; but especially the fact that capacity is much more difficult to control renders microscopic models less suited for (large-scale) TTV purposes.
Still, even with macroscopic DNL models, evaluating a large set of Monte-Carlo simulations in real-size networks is computationally troublesome or even practically infeasible. Many approaches can be found in the literature to circumvent these computational limitations. A general discussion, not limited to TTV studies, is provided in Section 7.1. In this thesis, marginal DNL simulation is put forward as a general approach that allows fast, repeated dynamic simulations featuring a realistic representation of congestion spillback.

Thirdly, modelling the effect of TTV on the travel choice process requires answering two questions:

- How should TTV be included in the cost that drivers perceive for a trip (more specifically for the trip's mode, departure time, and route)?

- How should the response of drivers to TTV be incorporated into DTA? What kind of changes does this require to the traditional equilibrium DTA framework?

Regarding the first question, an extensive survey can be found in Noland & Polak (2002). Two approaches exist in the state-of-the-art to include TTV in the cost function considered in the modelling of the choice behaviour, for which generally discrete choice theory is used (e.g. a logit model). These two are the mean-variance approach and the scheduling approach. Mean-variance, as the name suggests, includes the variance – or another measure expressing the variability (usually the standard deviation) – of the travel time in the cost function of the trip, together with the mean (or sometimes median) travel time. This approach is used for instance in Jackson & Jucker (1981), De Palma & Picard (2005) and Small et al. (2005). In the scheduling approach, the cost of TTV is obtained by penalizing deviations between a user's actual arrival time and his preferred arrival time – early and late deviations are separately weighted; sometimes an additional penalty is included for being late. The scheduling approach is applied in e.g. Small (1982), Noland & Small (1995) and Noland et al. (1998).
Which of the two approaches is best suited depends on the application. The general advantage of the mean-variance approach is that it is simpler to use and to interpret. In particular, it avoids estimating (distributions of) preferred arrival times, which are not directly observable. The scheduling approach on the other hand may provide a better insight into the

actual behavioural response of drivers to TTV. Batley et al. (2008) state that a stronger theoretical foundation and more comparison with empirical data are required for both approaches. Finally, we note that Fosgerau & Karlström (2010) identify the equivalence between the two approaches under some simplifying assumptions.
An important aspect – regardless of which of the approaches is used – is the valuation of TTV in the cost function, for instance in monetary terms or relative to the expected travel time. All studies indicate that it is indeed an important factor. Both stated preference and revealed preference surveys (see De Palma & Picard, 2005; Lam & Small, 2001 and several other of the references above) have shown that route choice is determined to a large extent by TTV. Liu et al. (2004) reported similar findings from empirical traffic counts on a congested corridor with a tolled, congestion-free bypass, indicating that the valuation of TTV even exceeds the value of expected travel time savings. Noland & Small (1995) and Noland et al. (1998) showed that also departure time choice is largely influenced by TTV. Despite such studies, a universal agreement on the exact valuation of TTV is still lacking[63]. Moreover, even which measure is best used to represent TTV in the cost function – this mainly applies to the mean-variance approach – is not indisputably determined. Lam & Small (2001) state that TTV is best expressed by the difference between the 90^{th} and the 50^{th} percentiles of travel times. But perhaps even higher-order moments of the distribution of travel times should be included, since travel times are typically not normally distributed. Rather, they follow a skewed distribution with a long tail on the high end (see Herman & Lam, 1974). At least, the fact that drivers perceive 'higher than expected' travel times much more negatively than 'lower than expected' travel times should be accounted for. Finally, like the valuation of time itself, the valuation of TTV is obviously an individual matter. Therefore, studies that explicitly investigate the impact of (the distribution of) risk aversion among drivers (e.g. Lo et al., 2006; De Palma & Picard, 2006) are useful.

Regarding the second question on page 178, the driver behaviour and system response in DTA are traditionally assumed to be in a (user) equilibrium (see Section 1.1.1.2). It is a difficult task to transform this traditional framework for TTV purposes. As discussed above, a large number of scenarios, each representing a specific, random situation with unpredictable demand and/or supply fluctuations, are to be evaluated to quantify TTV. Clearly, assuming that each of these specific situations is in equilibrium is unrealistic. Indeed, this would correspond to drivers having perfect knowledge of (the consequences of) the random events; for instance drivers would anticipate the congestion due to an incident even before the incident has occurred. This goes directly against the essence of TTV studies, namely to investigate the effect of unpredictable variations. Conceptually somewhat more defendable, is a Probabilistic User Equilibrium (PUE). In a PUE, equilibrium is not assumed in each single situation. Rather, drivers are assumed to assimilate their past experiences (regarding average travel times and TTV) from a large set of DNL runs (simulating the scenarios, which themselves are non-equilibrium) and settle into a long-term equilibrium. This approach is used by Lo & Tung (2003) to determine the route choice resulting from stochastic link capacity degradations. It is also adopted by Li et al. (2010), who assume a simultaneous route

[63] For instance Hamer et al. (2005) describes the outcome of an expert workshop in which the value of TTV is discussed for different transport modes and trip purposes. Considerable more debate among researchers will be required, however, before a universal agreement can be achieved.

and departure time choice in PUE under stochastic link capacities. Our proof-of-concept case study in Section F.2 also assumes a PUE route choice.
However, a PUE by itself still does not represent drivers' decision making under unpredictable conditions very realistically. Drivers react to each specific situation - not only in the long run based on the assimilation of their experiences. Most drivers try to gain information prior to and during their trip from various information systems (radio, variable message signs, GPS or smart phone, etc.). Within the boundaries set by the flexibility and information they have, they will respond to each specific situation by shifting their departure time, avoiding their traditional route or rerouting away from it, or even by changing the destination or not making the trip at all. Tampère & Viti (2010) suggest the general concept of equilibrium user strategies to reflect this behaviour. They state that drivers will pursue an optimal strategy – much like an if-then algorithm, including parameters - of considering past experiences and new information and make their pre- and en-route travel choices accordingly. Hence, the actual choices may differ in each specific situation, while the strategy (i.e. the decision process, resulting in specific choices per trip) of drivers converges to an equilibrium in which each user maximizes his valuation of the resulting probabilistic distribution of travel costs. Tampère & Viti (2010) suggest a dynamic process modelling framework. Earlier, Gao (2006) – see also Gao & Chabini (2006) - developed a specific instance of this general concept, namely equilibrium user routing strategies under stochastic demand and supply (in a PUE). Unnikrishnan & Waller (2009) analyze and exactly solve the static version of this problem, which they call the User Equilibrium with Recourse (UER). More research is needed to further explore the possibilities of this new modelling framework for DTA.

Marginal simulation, introduced in this thesis, can be used to address the second of the aforementioned challenges; i.e. the efficiently modelling of the TTV resulting from demand and supply variations with a realistic representation of congestion dynamics. The other two challenges discussed above, which concern the necessary input for these marginal simulation tools and the DTA modelling framework in which they are to be employed in TTV studies, are not part of the scope of this dissertation. Research in each of these three topics is necessary to reduce TTV through robust network management, planning and design.
The second part of this appendix provides a proof-of-concept case study that exemplifies the potential role of marginal simulation in TTV studies.

F.2 Proof-of-concept case study

First, it should be highlighted that this case study is intended as a proof-of-concept. Several simplifying assumptions are made. Foremost, a PUE is adhered; see Figure F-1. From a large set of incident cases (evaluated with MIC) the route TTV is quantified. This TTV is then accounted for in the route choice model. This new route choice is then adopted in both the base simulation and all incident scenarios of the next DNL iteration, which produces a new probabilistic distribution of route travel times. This process continues until convergence to a PUE is reached. This means that drivers are assumed to react indirectly to the encountered variability by adopting a fixed long-term route choice, rather than trying to optimize every single situation. Therefore, this case study is to be regarded as a proof-of-concept, showing

that accounting for TTV indeed has a significant impact on driver behaviour and the resulting traffic conditions in the model[64].

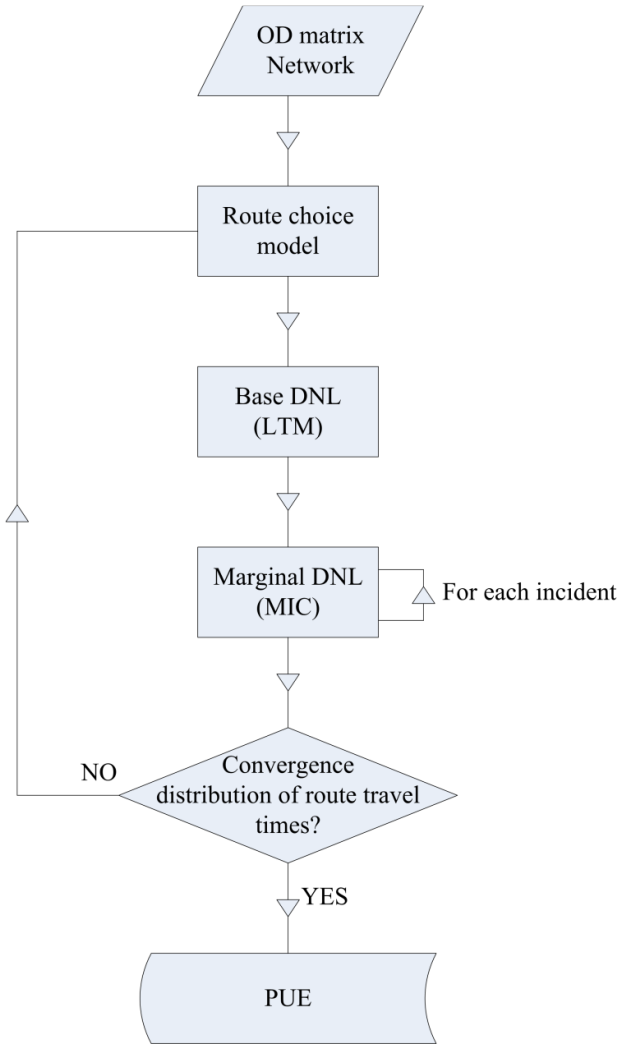

Figure F-1: MIC as a component in a PUE DTA

F.2.1 Network, demand and route choice initialization

A simple dummy network is chosen. It consists of six links and six nodes (Figure F-2). All traffic originates from A, heading towards destinations E or F.

[64] This has been repeatedly shown in the literature. Therefore, this appendix is to be seen as a demonstration of one of the possible applications of marginal simulation rather than as a stand-alone scientific contribution.

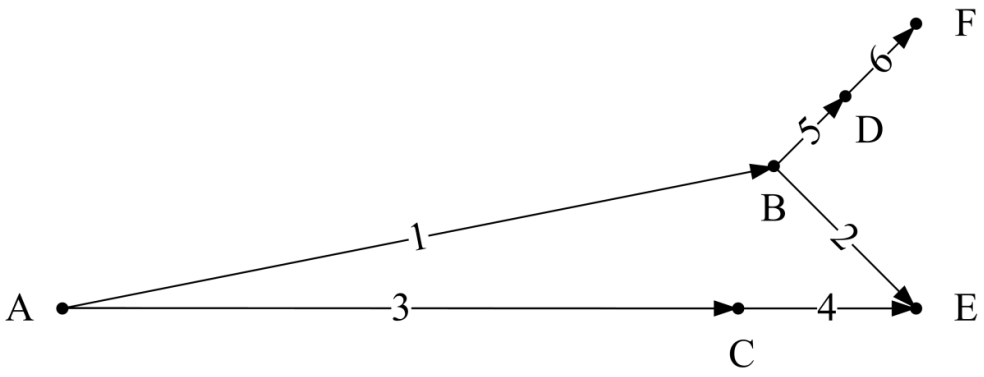

Figure F-2: Dummy network

The characteristics of the links (length L, capacity C, free flow speed v_f and jam density k_j) are given in Table F-1.

Table F-1: Link characteristics

Link	L (km)	C (veh/h)	v_f (km/h)	k_j (veh/km)
1	30	6000	100	375
2	5	6000	100	375
3	25	3000	70	250
4	5	1500	70	125
5	2.5	1500	70	125
6	2.5	1500	70	125

In a four hour simulation period, a constant travel demand of 4000 veh/h towards E and 1000 veh/h towards F is used. Only one route exists between A and F, while the demand towards E is divided over two alternative routes, r (links 1 and 2) and s (links 3 and 4). The free flow travel time of routes r and s are 21 and 25.7 min respectively. Based on this, the route choice (represented by route fractions p_r and p_s) is initialized for the first DTA iteration. The parameter of the logit route choice model (see Section F.2.3) is calibrated after Wohl & Martin (1967). They relate route fractions to the ratio of travel times of two alternative routes. This leads to initial route fractions p_r and p_s of 0.75 and 0.25 (for the entire simulation period). This corresponds to route demands of 3000 veh/h for route r and 1000 veh/h for route s. With this route choice, no congestion occurs on the network in the initial base scenario. Hence, the base route travel times in the first DTA iteration are equal to the free flow travel times. Therefore, this would be the user equilibrium if drivers would be insensitive to TTV.

F.2.2 Incident scenarios

TTV due to demand variations or small DTD capacity fluctuations is not considered; this case study focuses solely on incidents. Moreover, the same TTV due to incidents is assumed on the two routes themselves; and can therefore be excluded from the evaluation. However,

additional variability arises on route r due to incidents on links 5 and 6, causing congestion to spill back onto link 1 of route r. Only these incidents are simulated, adding TTV to route r. Consequently, even though the expected travel time remains lower, more drivers will opt for the reliable route s.

Thus, a set of incidents has to be generated between nodes B and F. Since the exact location of an incident has no decisive influence in this case study, all incidents are located in D (the middle between B and F). The characteristics of the incidents (duration, severity, probability) are based on Meeuwissen et al. (2004), while ensuring that queues never grow outside of the dummy network – i.e., spill back further than A. Only severe incidents are considered in this case study (reducing the link capacity to 5 % of its original value). Firstly, a set of 20 incidents is collected with varying duration, taken from a lognormal distribution with an average of 30 min and a standard deviation of 4 min. This set of incidents is repeated every 2.4 minutes to include the effect of varying starting time of the incidents. This leads to a total of 2000 incidents which are simulated – one by one – with MIC. Based on Meeuwissen et al. (2004), the probability of a severe incident occurring during the four hour simulation period is set to 2.5 % (very high risk). For this probability, a discrete distribution in time is assumed (Figure F-3). The possibility of more than one incident taking place during the simulation period is not considered. The incident probability is not directly accounted for in the sampling of incidents. Rather, the results are weighted with the incident probability before being considered in the route choice model.

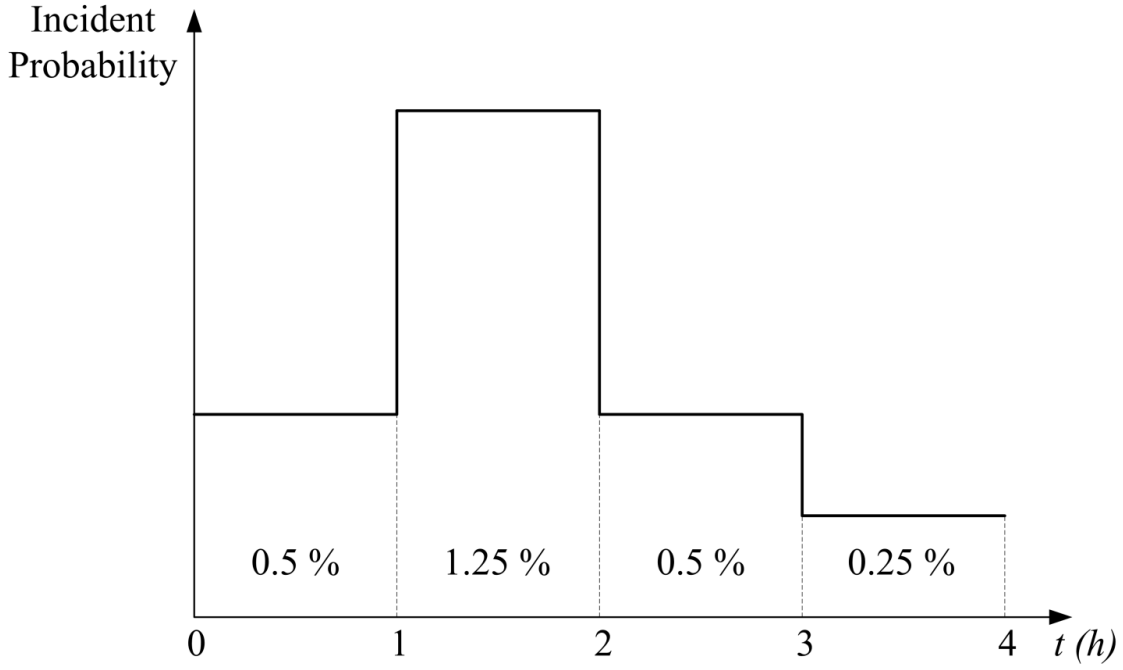

Figure F-3: Incident probability during the simulation period

F.2.3 Route choice model

In the following, we consider departure periods of 15 minutes and a route travel time for each of these periods for each (marginal) DNL simulation. A logit route choice model determines new route fractions for the next DTA iteration. The route fractions for route r are calculated for every departure period *d* according to (F.1). Since there are only two routes, $p_s(d) = 1 - p_r(d)$.

$$p_r(d) = \frac{e^{-\mu c_r(d)}}{e^{-\mu c_r(d)} + e^{-\mu c_s(d)}} \quad \forall d \tag{F.1}$$

The parameter μ in (F.1) is calibrated to obtain the initial route fractions 0.75 and 0.25, resulting in μ = 14. The cost functions *c(d)* are defined as a combination of the median route travel time $\tilde{tt}$ and a measure for the TTV for a departure period *d* (F.2). The median travel time was chosen instead of the mean because it is (arguably) a better representation of drivers' expected travel time than the mean; especially in this case study without regular congestion. The TTV is then expressed by a formula similar to the standard deviation; however, the mean travel time is replaced by the median. The TTV measure is weighted with a coefficient γ, which represents drivers' valuation of TTV with respect to the median travel time. For this proof-of-concept case study, the estimation γ = 2 is applied, which is on the high side.

$$c(d) = \tilde{tt}(d) + \gamma \sqrt{\frac{\sum_{x=1..2000} P_x (tt_x(d) - \tilde{tt}(d))^2}{N}} \quad \forall d \tag{F.2}$$

In (F.2), *x* denotes a specific incident scenario and $tt_x(d)$ the travel time for the currently considered route in this marginal simulation for departure period *d*. Note that the numerator in the TTV term in (F.2) is always zero if no incident occurs, i.e., for the base scenario. Since the total incident probability is 2.5 %, the base scenario holds in 97.5 % of all cases. Thus, the base travel time is equal to the median travel time. To give the correct weight to the base scenario, the total number of evaluations *N* = 2000/0.025 = 80000; since the 2000 incidents represent 2.5 % of all cases. Also, the varying incident probability (see Figure F-3) has not yet been taken into account. Therefore, the result of each marginal simulation *x* has to be multiplied with a factor P_x, given by:

$$P_x = \frac{P_{real}}{P_{sim}} \quad \forall x \tag{F.3}$$

P_{real} is the real probability of a specific incident sample, which varies per hour according to Figure F-3. For example, for each incident that occurs in the second hour, P_{real} = 0.0125/500; since the probability of any incident occurring during the second hour is 1.25 % and 500 is

the actual number of incidents that is simulated during that hour. Following the same reasoning, the simulated probability P_{sim} = 0.025/2000 for all incident scenarios.

F.2.4 Results

The route fractions for route r (after the first two iterations and in PUE) are depicted in Figure F-4. (The result for route s obviously equals $p_s = 1 - p_r$.)

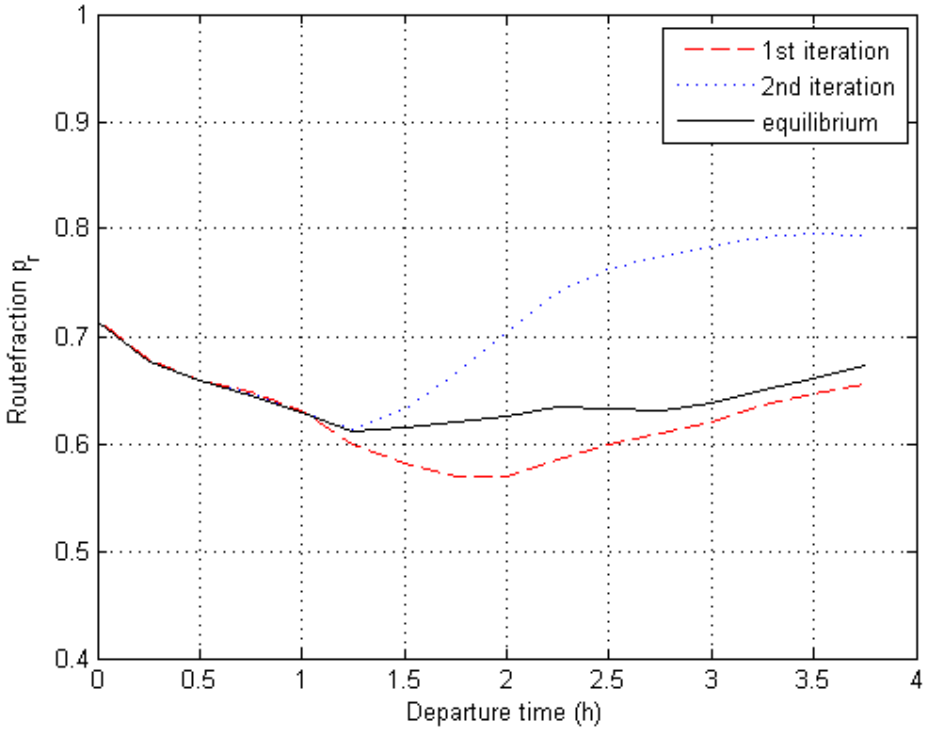

Figure F-4: Route fraction p_r

The dashed red line represents route fraction p_r after the first DTA iteration. The TTV motivates risk averse drivers to switch to the reliable route s. In the second iteration, this leads to congestion on route s in the base scenario, since the increase route demand activates the bottleneck at link 4. Meanwhile, the decreased route flow on r causes the TTV to decrease. Consequently, some drivers switch back to route s after the second iteration (dotted blue line in Figure F-4). The PUE (full black line) is a compromise between these extremes, where the cost of TTV on route r is balanced by the predictable delay at the bottleneck on route s.

The impact of TTV is clearly present, in that p_r is substantially lower than what it would be without variability – namely, equal to the initial p_r = 0.75. This difference is more pronounced during the period with higher incident probability. This is also visible in Figure F-5, which depicts the distribution of travel times for route r as a function of the departure time. (Note that the probability for free flow conditions, which is obviously > 97.5 %, is not depicted since this would render the graph impossible to interpret.) From the graph, the

higher incident probability during the second hour obviously results in more frequent high travel times.

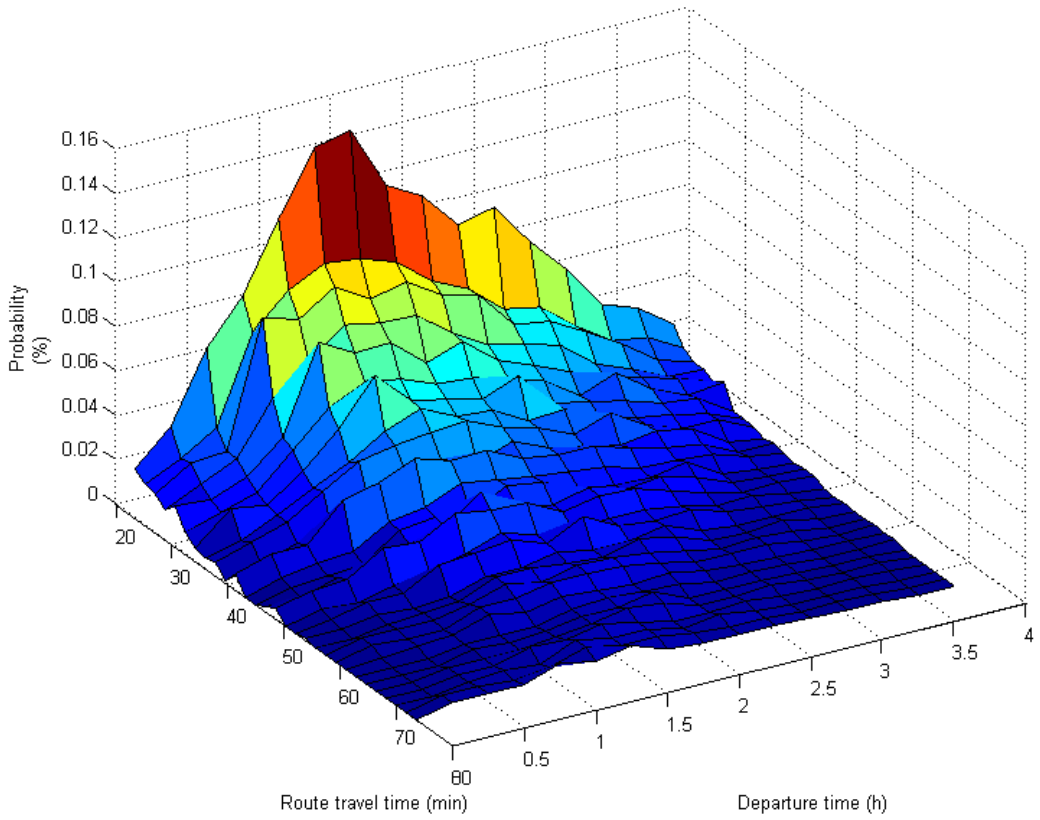

Figure F-5: Route travel time distribution for the incident scenarios

F.3 Conclusion

This appendix provides the reader with, firstly, an extensive discussion on TTV in general and an insight into the current challenges (Section F.1). Secondly, an interesting introduction of the role of marginal simulation in a TTV study is given in Section F.2, including many of the facets discussed in Section F.1.
In this case study, a base DNL is used with MIC simulations to form a stochastic set of DNL simulations, from which route travel time distributions are quantified. The observed TTV influences the route choice behaviour of drivers. A pair of routes is considered in the case study, one of which is prone to congestion spillback from incidents on links external to this route. Convergence to a PUE is reached.

Stochasticity is indeed an indispensable characteristic of DTA-based TTV studies, in which the sensitivity to TTV of the route and/or departure time choice of risk averse drivers is considered. However, only considering the long-term response of travellers to a large set of

experienced travel times as in the PUE model framework used here is probably insufficiently realistic. This approach implies that drivers only react indirectly to the encountered TTV, by adopting a fixed long-term route choice. Real drivers, however, are more likely to try to optimize every single situation. Therefore, TTV studies should ideally be performed using a more realistic modelling framework that allows for drivers to directly respond to unexpected conditions, as well as adapt their behaviour in the long term. For this, more research is necessary to further develop the concept of (equilibrium) user strategies into a well-established and accepted new modelling framework for DTA. Obviously, in such detailed studies, more driver decisions are to be included, foremost departure time choice and en-route rerouting. Regarding the latter, we refer to the proof-of-concept in Appendix G.

G

MaC EXTENDED WITH EN-ROUTE REROUTING

In this appendix, an en-route rerouting model is added to MaC (Chapter 9), namely the hybrid route choice model of Pel et al. (2009). As a result, drivers are no longer bound to the same route as in the base situation (no matter what the current conditions are).

Section G.1 briefly discusses the need for the computationally efficient modelling of drivers' response to unexpected traffic conditions. Then, Section G.2 explains how the hybrid route choice model of Pel et al. (2009) is incorporated in MaC. A proof-of-concept case study is presented in Section G.3. Finally, Section G.4 explains that implementing the hybrid route choice model into MaC poses several limitations that could be avoided if a new marginal DNL algorithm would be drafted from a more conveniently chosen maternal base model.

G.1 Introduction

Travel Time Variability (TTV) resulting from unpredictable variations such as Day-To-Day (DTD) demand and supply fluctuations and incidents may be quantified through Monte-Carlo simulation of various possible scenarios using macroscopic DNL models based on traffic flow theory. The computational limitations that exist in such applications can be (partially) relieved with marginal DNL simulation (see Chapters 7-9).
However, unpredictable traffic conditions also greatly influence the choice behaviour of travellers regarding mode, departure time and route choice (see Section F.1). Usually, the drivers' choices (whether or not the influence of TTV is accounted for) are only determined a priori, i.e. as a fixed input to the DNL model (e.g. as in Section F.2). Especially (but not only) route choice has a much more dynamic character. Drivers do not just determine their route prior to departure, but may deviate from their initial route if unexpected traffic conditions are met. This strategy is aided by various information systems available to drivers en-route. In turn, en-route rerouting behaviour influences the traffic conditions. This means that – at least for some applications – the effect of this behaviour should be accounted for in the DNL

simulation. By doing so, the mutual dependency between traffic conditions and drivers' rerouting can be modelled. Accounting for this is clearly beneficial for TTV studies. Also, it could enhance real-time control applications that invoke considerable rerouting such as incident-related traffic management by allowing the optimization of rerouting guidance around an incident. Similarly, other DTM measures such as ramp metering could be optimized.

En-route rerouting was not considered in the marginal simulation tools presented in earlier chapters. Since rerouting behaviour is of vital importance for some of the potential applications of marginal DNL simulation (notably TTV studies), marginal DNL algorithms that include en-route rerouting are an important development.

G.2 Incorporating en-route rerouting into MaC

In this section, the hybrid route choice model of Pel et al. (2009) is added to MaC. Consequently, MaC becomes a reactive assignment model (see Section 1.1.1.2) in which drivers may respond to the current traffic conditions by changing route at certain decision points. It is discussed in Section G.4 that stronger reactive marginal simulation algorithms can be developed by departing from an explicit simulation model that already encompasses an en-route rerouting model (such as EVAQ; see Pel, 2011) rather than LTM. Hence, this appendix is to be regarded as a proof-of-concept and an informative prospect into future research directions rather than as finished work.

In the hybrid route choice model, drivers re-evaluate their initial route (determined pre-trip) compared to attractive alternatives. A single parameter is introduced to express drivers' reluctance to move away from their initial route. This model is chosen because it assumes drivers to make a trade-off between their initially chosen, pre-trip route and en-route alternatives. This seems a more realistic representation of real driver behaviour – conceptually related to user strategies (see Tampère & Viti, 2010; Gao, 2006) as discussed in Section F.1 – than more traditional en-route models (see e.g. Kuwahara & Akamatsu, 2001) that assume drivers to simply choose the fastest route from their current location to their destination.

Note that this route choice is not equilibrated, as this does not realistically represent the driver behaviour resulting from unpredictable traffic conditions.

Firstly, let us define a route r as the collection of links towards destination d after the current link a. The route fraction $p^a_{r,ad}$ is then the fraction of drivers on link a that follow route r over all drivers travelling from a to d (G.1). These fractions are considered for time intervals with pre-specified length T_r. The time dimension is omitted for notational convenience.

$$p^a_{r,ad} = \frac{q^a_{r,ad}}{q^a_{ad}} \qquad with: q^a_{ad} = \sum_{r \in R_{ad}} q^a_{r,ad} \tag{G.1}$$

In the base simulation, drivers follow their initial routes *r*, which are chosen pre-trip. There may be a number of initial routes *r* from *a* to *d*, collected in set R_{ad}. From the base simulation, the route fractions $p^a_{r,ad}$ are known. The idea is then to adapt these fractions according to the en-route rerouting behaviour in each variation simulated with MaC. The change in $p^a_{r,ad}$ stems from drivers rerouting towards route alternative(s) *s* after this link *a* (see Figure G-1). (Of course, the route fractions $p^a_{s,ad}$ change accordingly.)

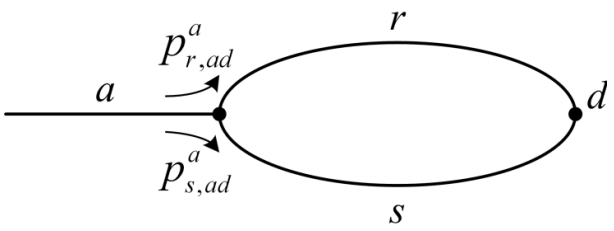

Figure G-1: Rerouting illustration

Alternatives *s* to an initial route *r* may include other initial routes $r' \in R_{ad}$, but also additional route alternatives that are not used in the base scenario. The route alternatives *s* to *r* at link *a* are collected in a set S^a_r. A shortest path search algorithm could be run to identify the route alternatives *s*. More easily implementable and more computationally efficient, a predefined selection of routes can be considered. This might be an interesting approach if mapped options for rerouting guidance (e.g. around an incident or road works) are to be compared. Only the latter approach with a predefined selection of route alternatives is implemented here.

The changes in route fractions should not only be imposed locally at the link where rerouting occurs. The altered route fractions should be generally accessible (e.g. from the origin) for subsequent rerouting decisions at downstream links. However, the structure of MaC (adopted from LTM) renders this difficult. Although tracing back the route fraction changes from a link to the origins of the routes is not impossible, this is not implemented at the moment. This means that subsequent rerouting would be prone to errors since the changes to the route fractions that have occurred earlier are neglected. Therefore, for this proof-of-concept, drivers are assumed to reroute only once.

To simulate the process of rerouting, the following questions need answering:

1. Where in the network and when will rerouting occur?
2. Which initial pre-trip routes *r* may be swapped for an alternative *s*?
3. Which route alternatives *s* should be considered?
4. How should the various route alternatives *s* be evaluated by drivers when reconsidering their route choice?
5. How are the flow changes that result from the changed route fractions simulated in MaC?

The first three questions are jointly answered in the following. Rerouting is considered at the end of a link *a* if two conditions are met:

- Unexpected delay is detected somewhere downstream of *a* on initial route(s) *r*. This delay is unexpected in the sense that it is not present in the base scenario - drivers did not account for it in their pre-trip route choice -, i.e. it has emerged in the current marginal simulation (e.g. due to an incident). As such, drivers can anticipate (increased) downstream congestion, as if information on the current traffic conditions is provided (e.g. on the radio or via VMS signs). A threshold *Q* is introduced so that rerouting is only considered for initial routes that carry a non-negligible flow ($q^a_{r,ad} > Q$).

- One or more attractive route alternatives *s* exist. An alternative *s* is considered attractive if its instantaneous travel time tt_s does not exceed the initial route's instantaneous travel time tt_r by more than a chosen percentage threshold tt_{add}, i.e. $tt_s < tt_r(1+\frac{tt_{add}}{100})$. In reality, drivers will have incomplete – and possibly incorrect - information about current and future traffic conditions and travel times. Perhaps, a prediction of travel times could be used instead of the instantaneous travel time. However, this is more difficult to determine and it may not even be the best representation of drivers' expectations. Also, at least under unpredictable and (highly) variable conditions, the instantaneous travel time arguably represents drivers' expectations better than past experiences; for an alternative route, they may not even have any past experience. In any case, it is preferable over using a fixed travel time (e.g. free flow travel time). Finally, we note that only the non-overlapping part of the routes is considered, i.e. from the next downstream link until the node where *s* and *r* join again.

Both of the above conditions are not static, but may or may not be met at different times during the simulation. If the conditions are met, the hybrid route choice model is run at that link with a pre-specified time interval T_r.

The fourth question is answered by explaining the hybrid logit route choice model of Pel et al. (2009) that is adopted in MaC. This model re-computes the route choice for each flow $q^a_{r,ad}$ separately. A fraction $p^a_{s,r}$ of $q^a_{r,ad}$ is derived that diverts from the initial route *r* to each route alternative *s*:

$$p^a_{s,r} = \frac{e^{-\mu c_s}}{e^{-\mu c_r} + \sum_{s' \in S^a_r} e^{-\mu c_{s'}}} \tag{G.2}$$

The fraction $p^a_{r,r}$ that remains on the initial route *r* is computed analogously. In (G.2), μ is the well-known parameter of the logit model. The instantaneous route travel time *tt* is

included in the generalized costs *c* (see (G.3)-(G.4)). Furthermore, a term η is added to account for route overlap. Ways to specify this overlap term can be found in e.g. Cascetta et al. (1996). For the alternative routes *s*, an additional term to express drivers' reluctance to move away from their initial route is introduced. This term consists of a difference factor $\lambda_{s,r}$, which is the length of the part of *s* that differs from *r* relative to the total length of *s*. Indeed, with dissimilarity between the initial and alternative route, also the reluctance to reroute can be expected to increase. The weighting parameter ω can be varied between 0 and 1, expressing absence of reluctance to reroute and absolute determination to sticking to the initial route respectively.

$$c_r = tt_r + \eta_r \tag{G.3}$$

$$\begin{aligned} c_s &= tt_s + \eta_s + \left(\frac{\omega}{1-\omega}\right)\lambda_{s,r} \\ with: \lambda_{s,r} &= \frac{\sum_{a \in s \setminus r} L_a}{\sum_{a \in s} L_a} \end{aligned} \tag{G.4}$$

Finally, for each route *u* (whether initial and/or alternative) the following total flow results from the rerouting process at the end of link *a*:

$$\begin{aligned} q^a_{u,ad} &= \delta_{u,R}\, p^a_{u,u} q^a_{u,ad} + \sum_{r' | u \in S^a_{r'}} p^a_{u,r'} q^a_{r',ad} \\ with: \quad \delta_{u,R} &= 0 \quad if \;\; u \notin R_{ad} \\ \delta_{u,R} &= 1 \quad if \;\; u \in R_{ad} \end{aligned} \tag{G.5}$$

Since the total destination-based flow q^a_{ad} is assumed not to change, the new route fractions $p^a_{r,ad}$ and $p^a_{s,ad}$ immediately result from the route flows of (G.5).

Finally, the route flow changes have to be propagated through the network. This can be done by updating the turning fractions at link ends along the path of each route for which the flow has changed. Hereby, the exact same procedure as described in Section 9.1 to propagate demand variations can be followed. An accuracy threshold $\varepsilon_{reroute}$ is used to determine whether or not a change in the route choice is significant enough for the resulting flow changes to be tracked or not. More precisely, if the route fraction of an initial route *r* has significantly changed compared to the base simulation ($\frac{| p^a_{r,ad,MaC} - p^a_{r,ad,base} |}{p^a_{r,ad,base}} > \varepsilon_{reroute}$), the resulting flow change is tracked on both the initial route *r* and the alternative routes *s*[65]. Otherwise, the change is neglected on the initial route *r* as well as on all route alternatives *s*.

[65] Additional checks could be added for each route alternative *s* separately so that changes on a practically unused route alternative are not tracked. Currently, however, such additional checks are not implemented.

G.3 Case study

In this section, a test case of the rerouting feature added to MaC is presented. A vulnerability analysis, with incidents that reduce the link capacities, is performed on a small network. Two scenarios – with and without rerouting – are compared.

G.3.1 Set-up

In the considered network (Figure G-2), traffic flows from two origins (1, 2) to one destination (7).

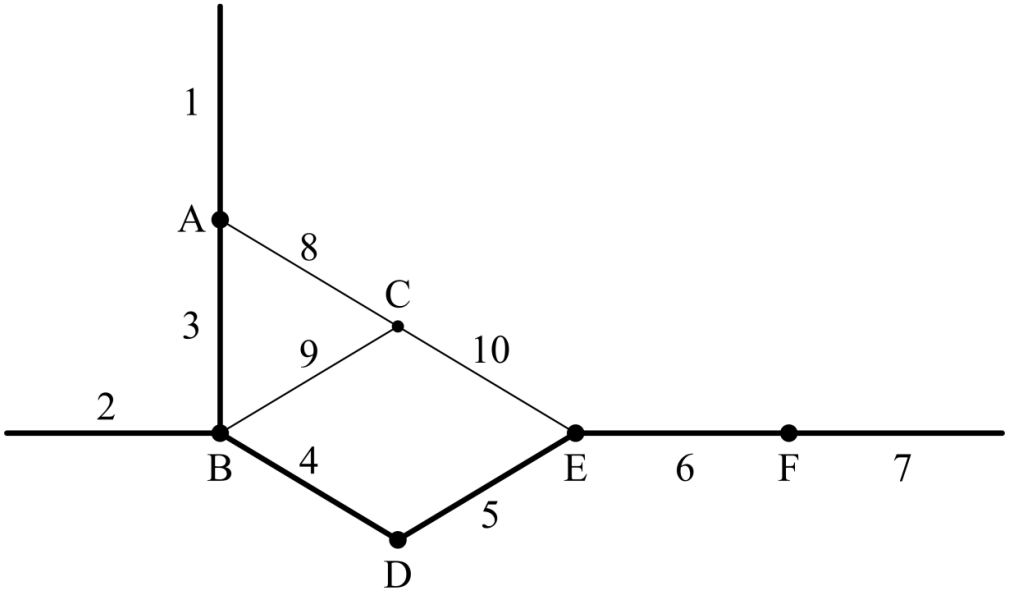

Figure G-2: Case study network

A five hour simulation period is considered, with a constant demand in the base situation (Table G-1). The link characteristics[66] are given in Table G-2. Links 1-7 is the highway part of the network; Links 8-10 represent urban roads.

Table G-1: Base demand

demand (veh/h)	7
1	1500
2	1500

[66] The update time step in MaC (as in LTM; see Yperman, 2007) is determined by the length and free flow speed of links. To reduce the numerical approximation errors due to large time steps resulting from long links, the links of the network in Figure G-2 are subdivided in 10 smaller segments in the simulations.

Table G-2: Link characteristics

	C (veh/h)	v_f (km/h)	k_j (veh/km)	L (km)
highway [1-7]	4000	100	280	12
urban [8-10]	2000	50	280	12

There are two decision points (A and B) where drivers may choose to reroute. In the base scenario, all drivers are assumed to use only the highway part of the network. This means that both at node A and B, there is only one initial route towards 7. In the scenario with rerouting, both at A and B one alternative route is present (which is not used in the base situation). Because at both A and B only one initial and one alternative route exists[67], the rerouting calculations can be simplified by only considering the non-overlapping part of the routes, so that route overlap does not have to be accounted for in the route choice model. The route sets then become:

- $R_{17} = \{r_1\}$ with route $r_1 = \{3, 4, 5\}$
- $R_{27} = \{r_2\}$ with route $r_2 = \{4, 5\}$
- $S^1_{r_1} = \{s_1\}$ with alternative route $s_1 = \{8, 10\}$
- $S^2_{r_2} = \{s_2\}$ with alternative route $s_2 = \{9, 10\}$

As such, (G.3) and (G.4) simplify to:

$$c_r = tt_r \tag{G.6}$$

$$c_s = tt_s + \left(\frac{\omega}{1-\omega}\right) \tag{G.7}$$

As in Section 9.2, the standard parameters in MaC are set to:

$$\begin{aligned} \varepsilon_{up} &= 0.5\% \\ \varepsilon_{down} &= 2\% \\ T &= 3\,\text{min} \end{aligned} \tag{G.8}$$

The additional rerouting parameters are set to:

[67] The possibility of drivers originating from 1 rerouting at B is neglected, since subsequent rerouting is not modeled as explained in Section G.2.

$$
\begin{aligned}
&Q: not\ applicable \\
&tt_{add} = 33.3\% \\
&\varepsilon_{reroute} = 2\% \\
&\mu = 14 \\
&\omega = 0.3 \\
&T_r = 5\,\text{min}
\end{aligned}
\tag{G.9}
$$

In a real-world case study, of course the above parameters need to be calibrated to the observed driver behaviour. For this demonstrative case study, they are chosen such that a credible rerouting response to the imposed incidental congestion is obtained.

Finally, the incidents that are analyzed to assess the vulnerability of different links need to be defined. For this, a set of incidents is chosen that will be inflicted (after one hour of simulation[68]) at the upstream boundary of each considered link. Hereby, the incident durations are varied from 30 to 60 min, in steps of 3 min; and the incident severity (i.e. the percentage reduction of a link's capacity) is varied from 30 to 80 %, in steps of 5 %. Thus, a set of 121 incidents is obtained, to be simulated on each considered link. Only the four links common to both initial routes are analyzed (links 4, 5, 6, 7). Each of these 484 incident variations is simulated twice with MaC; once for the scenario with and once for the scenario without rerouting.

G.3.2 Results

Firstly, a quick notion is provided on the computational implications of adding a rerouting model to MaC. In Table G-3, the computational effort for both scenarios is compared to the computation time that would be required with MaC's maternal base model LTM. The LTM case is of course without rerouting, as this possibility is not present in LTM. One can see that while adding rerouting absorbs some of the computational benefit provided by the marginal simulations, the increase in computation time is reasonable. It should be noted that in large-scale networks, much larger computational gains are to be expected from marginal compared to explicit simulation.

Table G-3: Computation time

LTM (no rerouting)	100%
MaC (no rerouting)	19%
MaC (rerouting)	34%

The vulnerability of each link is evaluated by the average and standard deviation of the VHL due the incidents on that link. In Table G-4, the results are given for the scenario without rerouting. Not surprisingly, in this case each link is classified as equally vulnerable, since the

[68] Since the demand is constant, the exact starting time of the incident is not important, as long as sufficient time is considered prior to and after the incident.

links have the same characteristics, the same incidents are imposed and the same amount of flow passes.

Table G-4: Vulnerability analysis without rerouting

VHL	4	5	6	7
avg	894	894	894	894
st dev	780	780	780	780

In the scenario with rerouting, links 4 and 5 are identified as substantially less vulnerable (Table G-5).

Table G-5: Vulnerability analysis with rerouting

VHL	4	5	6	7
avg	739	756	901	901
st dev	516	523	783	794

Incidents on these links can be bypassed by adaptive drivers via the alternative routes. This partially relieves the bottleneck and thus reduces congestion. For incidents on link 4, rerouting occurs only at diverge A, from r_1 to s_1. (The head of the queue is located at B, so at that point rerouting is no longer attractive.) Thus, only drivers originating from 1 are presented with an attractive alternative. For incidents on link 5, rerouting may occur at A and B. Still, also in this case, s_1 is a much more attractive alternative to r_1 than s_2 to r_2 (see Figure G-2). Incidents on links 6 and 7 on the other hand cannot be bypassed by rerouting. Still, drivers will divert to the urban alternative routes in an attempt to escape the queue on the highway, hereby dividing the congestion over the highway and urban routes. The results in Table G-5 show, however, that this rerouting behaviour even slightly increases the total number of VHL compared to the scenario without rerouting. Indeed, since the drivers' decisions are based on instantaneous travel times, they may be erroneous and rerouting may turn out to be disadvantageous.

In Figure G-3, the decreased vulnerability of links 4 and 5 is clear. The impact of severe incidents is significantly reduced compared to links 6 and 7, thanks to the existing route alternatives. Also, it can be seen that for mild incidents, these alternatives are not yet attractive. Furthermore, small differences can be observed in the vulnerability of links 4 and 5. Firstly, link 4 is somewhat less vulnerable to medium-impact incidents. A probable explanation for this is the following. For incidents on link 4, the incident location is closer to the decision point where drivers can reroute. Hence, fewer drivers will find themselves between the decision point and the incident location – and thus unable to reroute - when the incident occurs. Moreover, drivers can respond better, since the instantaneous travel times on which they base their decision deviate less from the real travel times than for incidents on the further away link 5. On the other hand, link 5 is less vulnerable to severe incidents. This is because now also drivers coming from origin 2 have an attractive alternative route; indeed, s_2 is only attractive in case of a severe incident due to the relatively large detour. For incidents

on link 4, drivers originated from 2 have no viable alternative, since the head of the queue is located at B.

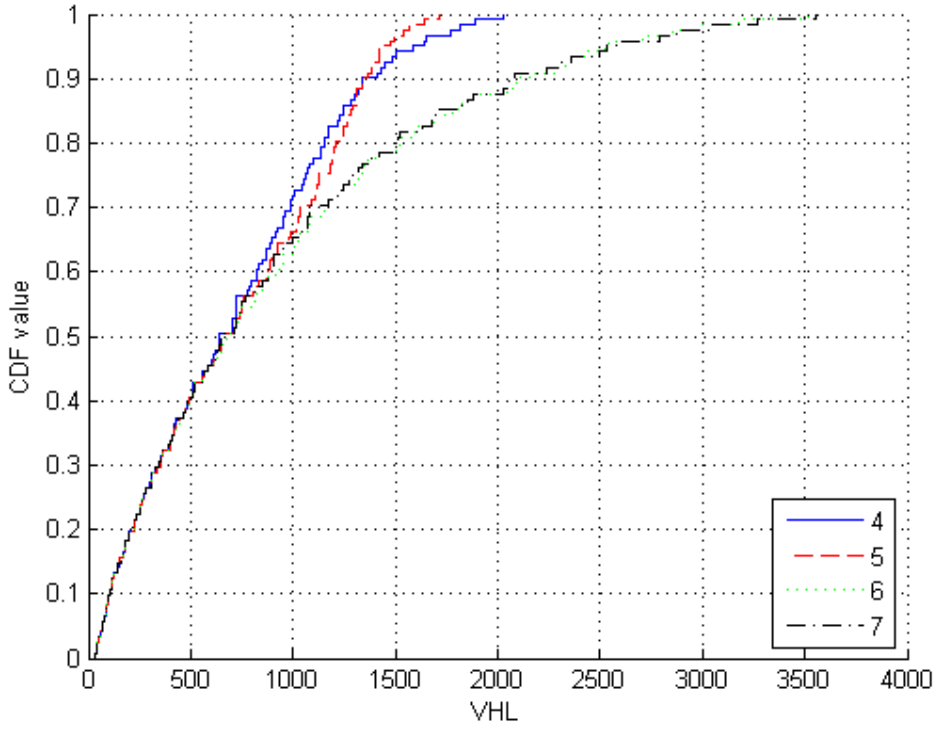

Figure G-3: Cumulative distribution of the impact of incidents in terms of VHL

In conclusion, this case study demonstrates that accounting for the effect of rerouting is indeed important in assessing vulnerability to unpredictable perturbations. Rerouting behaviour does not only influence the average system performance, but also the variability. In fact, the decrease of the standard deviation of VHL thanks to the presence of rerouting options is even larger than that of the average in the presented case study (compare Table G-4 and Table G-5).

Finally, of course the above presented results depend on the setting of the parameters, in particular the route choice parameters. In a real-world case study, these should be estimated from empirical data or surveys. Moreover, insight into the sensitivities of the drivers' rerouting behaviour enables additional research applications. For example, the benefit of various driver information systems could be compared, mainly by their influence on drivers' reluctance to move away from their initial route. Therefore, the impact of this reluctance parameter ω is indeed quite interesting. The following section provides a short analysis.

G.3.3 Impact of reluctance to reroute

The sensitivity of the results to the value of the reluctance parameter ω is shown in Figure G-4 (average VHL) and Figure G-5 (standard deviation of VHL).

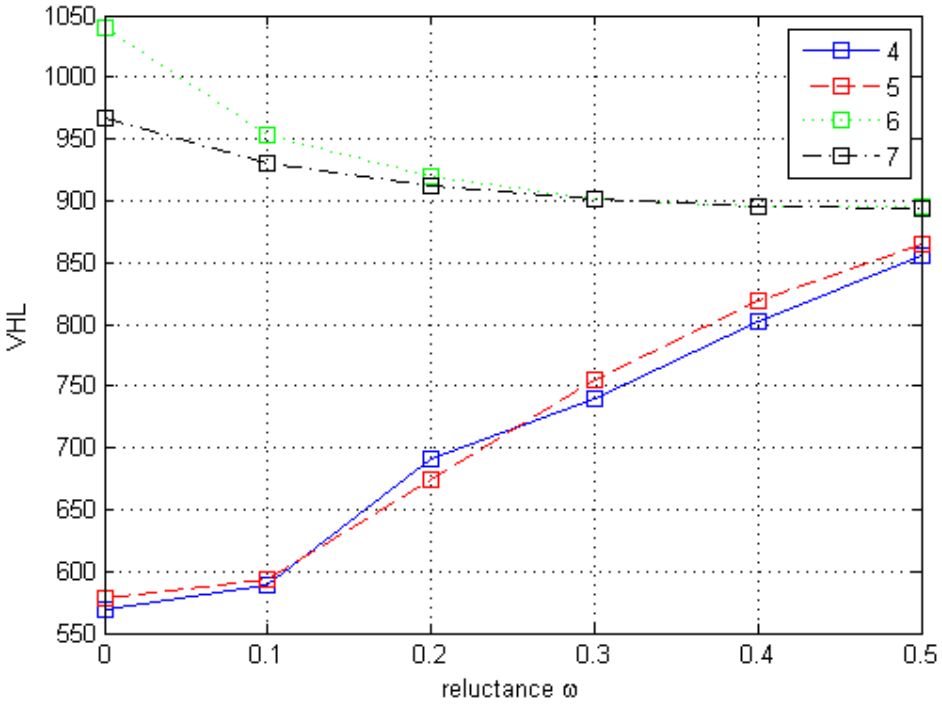

Figure G-4: Average VHL for different levels of reluctance ω

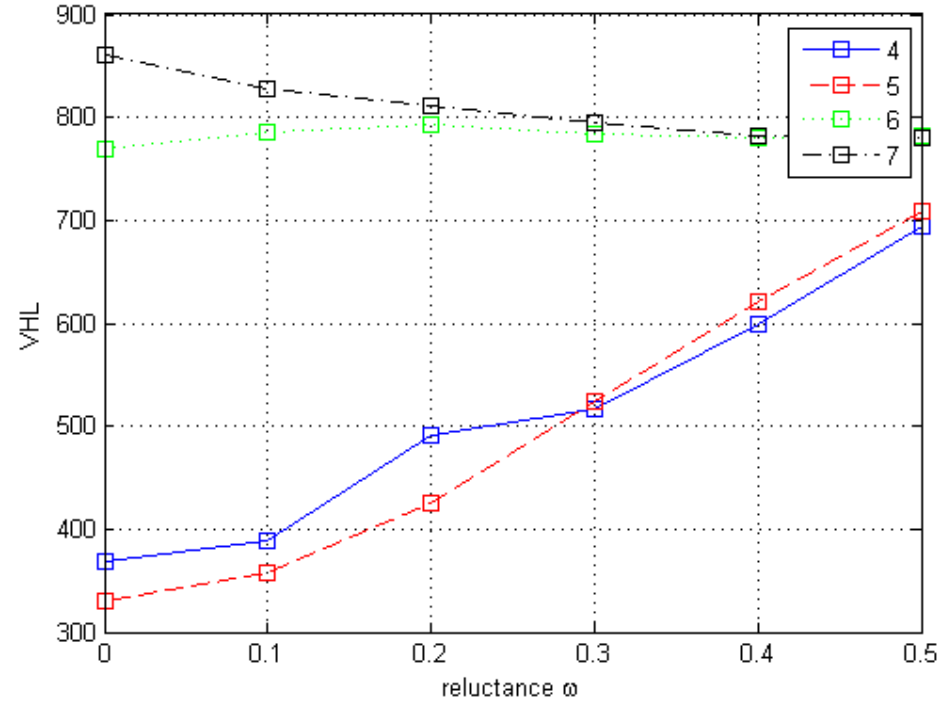

Figure G-5: Standard deviation of VHL for different levels of reluctance ω

For incidents on links 4 and 5, the negative effects of the incidents increase with increasing reluctance to rerouting. In this case study, the optimal behaviour would be that drivers evaluate the route alternative the same as their initial route, without an additional reluctance cost. For incidents on links 6 and 7 on the other hand, less vehicle hours are lost as reluctance increases. This again confirms that the rerouting decisions based on instantaneous travel times are often erroneous in these cases; and that dividing the congestion over highway and urban network does not compensate for the detour that rerouting drivers make. For $\omega > 0.3$, hardly any rerouting occurs for incidents on links 6 and 7.

In conclusion, rerouting should be carefully and precisely controlled so that drivers are preserved from potentially disadvantageous rerouting decisions based on incomplete and possibly incorrect information such as instantaneous travel times.

G.4 Conclusion

In this appendix, the extension of MaC (see Chapter 9) with the hybrid route choice model of Pel et al. (2009) is discussed. Indeed, the inclusion of drivers' rerouting response to unexpected conditions due to demand and/or capacity variations is valuable in numerous applications such as TTV studies (notably vulnerability analysis) and DTM that might invoke considerable rerouting such as incident management, ramp metering and evacuation planning. In such applications, a marginal simulation tool could be used to estimate the impact of incidents, road works, events or disasters on the network conditions in absence of management interventions. Secondly, it could assist in comparing and choosing between several possible strategies for route guidance.
However, the current implementation has a limited functionality. In retrospect, adding a rerouting model to MaC is not a convenient methodology. This is because MaC adopts its structure and functions from LTM, a DNL model that does not include en-route rerouting. Rather, a marginal DNL algorithm with en-route rerouting behaviour should be extracted from an explicit DNL model that already has the necessary properties. More specifically, the methodology in Chapter 9, explaining how MaC is derived from LTM, could be applied to the EVAQ model used in Pel et al. (2009) – see Pel (2011) for more details - which includes the hybrid route choice model. This way, the current limitations that inhibit real-world applications, such as the lack of a global rerouting memory and the need to manually input route alternatives, could be relieved more easily. Considering the above, the study presented in this appendix mainly serves as a proof-of-concept, exemplifying the potential benefits of developing marginal DNL simulation tools that not only efficiently simulate variable traffic conditions, but also the resulting response of drivers in a reactive assignment.

REFERENCES

AbdelFatah, A.S. & Mahmassani, H.S. (1998). System Optimal Time-Dependent Path Assignment and Signal Timing in Traffic Network. *Transportation Research Record 1645, pp. 185-193.*

Adamo, V., Astarita, V., Florian, M., Mahut, M. & Wu, J.H. (1999). Analytical modeling of intersections in traffic flow models with queue spill-back. *IFORS' 99 15Th Triennial Conference (ORSC), Beijing, P.R. China* (also published in *CRT di Montreal. Publication CRT 1999 nr. 52. National Libraries of Quebec and Canada*).

Akcelik, R. & Troutbeck, R.J. (1991). Implementation of the Australian Roundabout Analysis Method in SIDRA. *Highway Capacity and Level of Service, Proceedings of the International Symposium on Highway Capacity, pp. 17-34*.

Ansorge, R. (1990). What does the entropy condition mean in traffic flow theory? *Transportation Research Part B 24, pp. 133-143.*

Arnott, R., De Palma, A. & Lindsay, R. (1996). Information and Usage of Free-Access Congestible Facilities with Stochastic Capacity and Demand. *International Economic Review 37 (1), pp. 181-203.*

Bar-Gera, H. & Ahn, S. (2010). Empirical macroscopic evaluation of freeway merge-ratios. *Transportation Research Part C 18 (4), pp. 457-470.*

Bates, J., Polak, J., Jones, P. & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E 37, pp. 191-229.*

Batley, R., Grant-Muller, S., Nellthorp, J., de Jong, G., Watling, D., Bates, J., Hess, S. & Polak, J. (2008). *Multimodal Travel Time Variability: Final Report, Institute for Transport Studies, University of Leeds, Imperial College and John Bates Services.*

Beckmann, M.J., McGuire, C.B. & Winston, C.B. (1956). *Studies in the Economics of Transportation*. Connecticut: Yale University Press.

Bell, M.G.M., Cassir, C., Iida, Y. & Lam, W.H.K. (1999). A sensitivity based approach to network reliability assessment. *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT), pp. 283-300.*

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H.N. & Mishalani, R. (1998). DynaMIT: A Simulation-based System for Traffic Prediction and Guidance Generation. Presented at *the 3rd Triennial Symposium on Transportation Systems*, San Juan.

Berdica, K. (2002). An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy 9, pp. 117-127.*

Bezembinder, E. & Brandt, F. (2007). Junction Modeling in OmniTRANS, Technical Note – October 2007.

Bierlaire, M. & Crittin, F. (2004). An efficient algorithm for real-time estimation and prediction of dynamic OD tables. *Operations Research 52 (1), pp. 116-127.*

Bliemer, M.C.J., Versteegt, H.H. & Castenmiller, R.J. (2004). INDY: A New Analytical Multiclass Dynamic Traffic Assignment Model. Presented at *the TRISTAN V conference,* Guadeloupe.

Bliemer, M.C.J. (2007). Dynamic Queuing and Spillback in an Analytical Multiclass Dynamic Network Loading Model. *Transportation Research Record 2029, pp. 14-21.*

Blumberg, M. & Bar-Gera, H. (2009). Consistent node arrival order in dynamic network loading models. *Transportation Research Part B 43, pp. 285-300.*

Bovy, P.H. (1991). Zusammenfassung des schweiterischen Kreiselhandbuchs, Strabe und Verkehr, nr. 3.

Brilon, W. & Wu, N. (2001). Capacity at unsignalized intersections derived by conflict technique. *Transportation Research Record 1776, pp. 82-90.*

Brilon, W. & Miltner, T. (2005). Capacity at intersections without traffic signals. *Transportation Research Record 1920, pp. 32-40.*

Brilon, W., Geistefeldt, J. & Regler, M. (2005). Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proceedings of the 16th International Symposium on Transportation and Traffic Theory, pp. 125-144.*

Buisson, C., Lebacque, J.P. & Lesort, J.B. (1995). Macroscopic modelling of traffic flow and assignment in mixed networks. *Proceedings of the Berlin ICCCBE Conference.*

Buisson, C., Lebacque, J.P. & Lesort, J.B. (1996). STRADA, a discretized macroscopic model of vehicular traffic flow in complex networks based on the godunov scheme. *Proceedings of the CESA '96 IEEE Conference.*

Cantarella, G.E. & Cascetta, E. (1995). Dynamic Processes and Equilibrium in Transportation Networks: Towards a Unifying Theory. *Transportation Science 29, pp. 305-329.*

Carey, M. (1992). Nonconvexity of the Dynamic Traffic Assignment Problem. *Transportation Research Part B 26 (2), pp. 127-133*.

Cascetta, E., Nuzzolo, A., Russo, F. & Vitetta, A. (1996). A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and some Calibration Results for Interurban Networks. In Lesort (Ed.), *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France.

Cascetta, E. & Postorino, M.N. (2001). Fixed Point Approaches to the Estimation of O/D Matrices Using Traffic Counts on Congested Networks. *Transportation Science 35 (1), pp. 134-147*.

Cascetta, E. (2001). *Transportation systems engineering: theory and methods.* Kluwer Academic Publishers, the Netherlands.

Cassidy, M.J. & Ahn, S. (2005). Driver Turn-Taking Behavior in Congested Freeway Merges. *Transportation Research Record 1934, pp. 140-147*.

Chen, A., Yang, H., Lo, H.K. & Tang, W. (1999). A capacity related reliability for transportation networks. *Journal of advanced transportation 33 (2), pp. 183-200.*

Chen, A., Yang, H., Lo, H.K. & Tang, W.H. (2002). Capacity reliability of a road network: an assessment methodology and numerical results. *Transportation Research Part B 36, pp. 225-252.*

Chen, L., Jin, W-L. & Zhang, J.H. (2008). An Urban Intersection Model Based on Multi-Commodity Kinematic Wave Theories. *Proceedings of the 11th International IEEE Conference on ITS, pp. 269-274*.

Chevallier, E. & Leclercq, L. (2007). A macroscopic theory for unsignalized intersections. *Transportation Research Part B 41 (10), pp. 1139-1150.*

Chevallier, E. & Leclercq, L. (2008). A Macroscopic Single-Lane Roundabout Model to Account for Insertion Delays and O-D Patterns. *Computer-Aided Civil and Infrastructure Engineering 23, pp. 104-115*.

Chevallier, E. & Leclercq, L. (2009). Do microscopic merging models reproduce the observed priority sharing ratio in congestion? *Transportation Research Part C 17 (3), pp. 328–336.*

Chiu, Y.C., Zheng, H., Villalobos, J. & Gautam, B. (2007). Modelling no-notice mass evacuation using a dynamic traffic flow optimization model. *IIE Transactions 39 (1), pp. 83-94*.

Clark, S. & Watling, D. (2005). Modelling network travel time reliability of a road network: an assessment methodology and numerical results. *Transportation Research Part B 39, pp. 119-140.*

Coclite, G. M., Garavello, M. & Piccoli, B. (2005). Traffic Flow on a Road Network, *SIAM Journal on Mathematical Analysis 36 (6), pp. 1862-1886.*

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2009). Marginal Incident Computation, an Efficient Algorithm to Determine Congestion Spillback due to Incidents. *Transportation Research Record 2099, pp. 22-29.*

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2010). Stochastic Dynamic Network Loading for Travel Time Variability due to Incidents. *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment, pp. 179-198.*

Corthout, R., Flötteröd, G., Viti, F. & Tampère, C.M.J. (2012). Non-unique flows in macroscopic first-order intersection models. *Transportation Research Part B 46 (3), pp. 343-359*.

Corthout, R., Tampère, C.M.J., Frederix, R. & Immers, L.H. (forthcoming). Improving the efficiency of dynamic network loading applications through marginal simulation. Submitted to *Transportation Research Part C*.

Corthout, R., Tampère, C.M.J. & Immers, L.H. (forthcoming). Introducing en-route rerouting into the Marginal Computation Model. Submitted to *IET Intelligent Transport Systems.*

Courant, R., Friedrichs, K. & Lewy, H. (1928). Über die partiellen Differenzengleichungen der mathematischen Physik, *Mathematische Annalen 100 (1), pp. 32-74*. (English translation: On the partial difference equations of mathematical physics, *IBM Journal, March 1967, pp. 215–234*).

Daganzo, C. & Sheffi, Y. (1977). On Stochastic Models of Traffic Assignment. *Transportation Science 11, pp. 253-274.*

Daganzo, C.F. (1994). The cell-transmission model. A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B 28 (4), pp. 269-288.*

Daganzo, C.F. (1995). The cell transmission model part II: Network traffic. *Transportation Research Part B 29, pp. 79-93*.

Daganzo, C.F. (1998). Queue spillovers in transportation networks with a route choice. *Transportation Science 32 (1), pp. 3-11*.

Daganzo, C.F. (2005a). A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B 39 (2), pp. 187-196*.

Daganzo, C.F. (2005b). A variational formulation of kinematic waves: Solution Methods. *Transportation Research Part B 39 (10), pp. 934-950*.

De Palma, A. & Picard, N. (2005). Route choice decision under travel time uncertainty. *Transportation Research Part A 39, pp. 295-324.*

De Palma, A. & Picard, N. (2006). Equilibria and Information Provision in Risky Networks with Risk-Averse Drivers. *Transportation Science 40 (4), pp. 393-408.*

Dervisoglu, G., Gomes, G., Kwon, J., Horowitz, R. & Varaiya, P. (2009). Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity. Presented at *the 88th Annual Meeting of the Transportation Research Board*, Washington D.C.

D'Este, G.M. & Taylor, M.A.P. (2001). Network vulnerability: an issue for regional, national and international strategic transport networks. *Proceedings of the First International Symposium on Transport Network Reliability,* Kyoto.

Durlin, T. & Henn, V. (2005). A Delayed Flow Intersection Model for Dynamic Traffic Assignment. *Advanced OR and AI Methods in Transportation, Proceedings of the 10th EWGT Meeting and 16th Mini-EURO Conference, pp. 441-449*.

Fellendorf, M. & Vortisch, P. (2010) Microscopic Traffic Flow Simulator VISSIM. *Fundamentals of Traffic Simulation: International Series in Operations Research & Management Science 145, pp. 63-93*.

Flötteröd, G. & Nagel, K. (2005). Some practical extensions to the cell transmission model. *Proceedings of the 8th IEEE Intelligent Transportation Systems Conference, pp. 510-515*.

Flötteröd, G., 2008. *Traffic State Estimation with Multi-Agent Simulations.* PhD dissertation, Berlin Institute of Technology, Germany.

Flötteröd, G., Bierlaire, M. & Nagel, K. (2011). Bayesian demand calibration for dynamic traffic simulations. *Transportation Science 45 (4), pp. 541-561*.

Flötteröd, G. & Rohde, J. (2011). Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B 45 (6), pp. 903-922*.

Fosgerau, M. & Karlström, A. (2010). The value of reliability. *Transportation Research Part B 44, pp. 39-49.*

Frederix, R., Viti, F., Corthout, R. & Tampère, C.M.J. (2011). A New Gradient Approximation Method For Dynamic Origin-Destination Matrix Estimation On Congested Networks. *Transportation Research Record 2263, pp. 19-25.*

Frederix R., Tampère, C.M.J. & Viti, F. (in press). Dynamic OD estimation in congested networks. *Transportmetrica, doi: 10.1080/18128602.2011.619587.*

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, pp. 179-191*.

Gao, S. (2006). Traffic Assignment with Adaptive Routing Choices in Stochastic Time-Dependent Networks. Presented at *the First International Symposium on Dynamic Traffic Assignment (DTA2006)*, Leeds.

Gao, S. & Chabini, I. (2006). Optimal routing policy problems in stochastic time-dependent networks. *Transportation Research Part B 40, pp. 93-122.*

Gartner, N.H., Messer, C.J. & Rathi, A.K. (2000). *Revised Monograph of Traffic Flow Theory. Online publication of the Transportation Research Board, FHWA (http://www.tfhrc.gov/its/tft/tft.htm).*

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 41, pp. 1114-1138.*

Gentile, G. & Noekel, K. (2009). Linear user cost equilibrium: the new algorithm for traffic assignment in VISUM. Presented at *the European Transport Forum*, Brussels.

Gentile, G. (2010). The General Link Transmission Model for Dynamic Network Loading and a Comparison with the DUE Algorithm. *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment, pp. 153-178.*

Gibb, J. (2011). A Model of Traffic Flow Constraint through Nodes for Dynamic Network Loading with Spillback. Presented at *the 90th Annual Meeting of the Transportation Research Board*, Washington, DC.

Gleue, A.W. (1972). Vereinfachtes Verfahren zur Berechnung signalgeregelter Knotenpunkte (Simplified method for the calculation of signalized intersections. *Forschung Strassenbau und Strassenverkehrstechnik 136.*

Hamer, R., De Jong, G. & Kroes, E. (2005). *The value of reliability in transport: Provisional values for the Netherlands based on expert opinion.* RAND Europe/AVV.

Heidemann, D. & Wegmann, H. (1997). Queuing at unsignalized intersections. *Transportation Research Part B 31 (3), pp. 239-263.*

Herman, R. & Lam, T. (1974). Trip Characteristics of Journeys To and From Work. *Proceedings of the Sixth International Symposium on Transportation and Traffic Theory, pp. 57-86.*

Herty, M. & Klar, A. (2003). Modeling, Simulation and Optimization of Traffic Flow Networks. *SIAM Journal on Scientific Computing 25 (3), pp. 1066-1087.*

Herty, M., Kirchener, C. & Moutari, S. (2006). Multi-class traffic models on road networks. *Communications in Mathematical Science 4 (3), pp. 591–608.*

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), pp. 999-1017.*

Hoogendoorn, S.P. & Bovy, P.H.L. (2001). State-of-the-art of vehicular traffic modeling. *Proceedings Institution of Mechanical Engineers 215 (1), pp. 283-303.*

Hwang, S.Y., Blank, T. & Coi, K. (1988). Fast Functional Simulation: an Incremental Approach. *IEEE Transactions on Computer-Aided Design 7 (7), pp. 765-774.*

Inouye, H. (2003). An evaluation of the reliability of travel time in road networks based on stochastic user equilibrium. *The Network Reliability of Transport: proceedings of the 1st International Symposium on Transportation Network Reliability (INSTR), pp. 79-91.*

Jackson, W. & Jucker, J. (1982). An empirical study of travel time variability and travel choice behavior, *Transportation Science 16 (4), pp. 460–475.*

Jenelius, E., Petersen, T. & Mattsson, L.-G. (2006). Importance and exposure in road network vulnerability analysis. *Transportation Research Part A 40, pp. 537–560.*

Jenelius, E. (2010). User inequity implications of road network vulnerability. *Journal of Transport and Land Use 2 (3/4), pp. 57-73.*

Jepsen, M. (1998). On the Speed-Flow Relationships in Road Traffic: A Model of Driver Behaviour. *Proceedings of the Third International Symposium on Highway Capacity, pp. 297-319.*

Jin, W.L. & Zhang, H.M. (2003). On the distribution schemes for determining flows through a merge. *Transportation Research Part B 37, pp. 521-540.*

Jin, W.L. & Zhang, H.M. (2004). A multicommodity kinematic wave simulation model of network traffic flow. *Transportation Research Record 1883, pp.59-67.*

Jin, W.L. (2010). Continuous kinematic wave models of merging traffic flow. *Transportation Research Part B 44 (8/9), pp. 1084-1103.*

Knoop, V., Hoogendoorn, S.P. & van Zuylen, H.J. (2007). Quantification of the impact of spillback modeling in assessing network reliability. Presented at *the 86th Annual Meeting of the Transportation Research Board*, Washington D.C.

Knoop, V., Hoogendoorn, S.P. & van Zuylen, H.J. (2008). Capacity Reduction at Incidents: Empirical Data Collected from a Helicopter. *Transportation Research Record 2071, pp. 19-25.*

Kurzhanskiy, A. & Varaiya, P. (2010). Using Aurora road network modeler for active traffic management. *Proceedings of the 2010 American Control Conference, pp. 2260-2265.*

Kuwahara, M. & Akamatsu, T. (2001). Dynamic user optimal assignment with physical queues for a many-to-many OD pattern. *Transportation Research Part B 35 (5), pp. 461-479.*

Lam, T.C. & Small, K.A. (2001). The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E 37, pp. 231-251.*

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), pp. 647-677.*

Lebacque, J.P. & Lesort, J.B. (1999). Macroscopic traffic flow models: a question of order. *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT), pp. 3-26.*

Lebacque, J.P. (2003) Intersection modeling, application to macroscopic network traffic flow modeling and traffic management. *Proceedings of the Traffic and Granular Flow '03, pp. 261-278.*

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), pp. 365-386.*

Li, H., Bliemer, M., Tu, H. (2010). Reliability-based Departure Time User Equilibrium. Presented at *the 3rd International Symposium on Dynamic Traffic Assignment (DTA2010),* Takayama.

Lighthill, M.J. & Whitham, G.B. (1955). On kinematic waves (II): A theory of traffic flow on long crowded roads. *Proceedings of Royal Society A229, pp. 281-345.*

Liu, H.X., Recker, W. & Chen, A. (2004). Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A 38, pp. 435-453.*

Liu, Y., Yu, J., Chang, G.-L. & Rahwanji, S. (2008). A Lane-group Based Macroscopic Model for Signalized Intersections Account for Shared Lanes and Blockages. *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, pp. 639-644.*

Lo, H.K. & Tung, Y.-K. (2003). Network with degradable links: capacity analysis and design. *Transportation Research Part B 37 (4), pp. 345-363.*

Lo, H.K., Luo, X.W. & Siu, B.W.Y. (2006). Degradable transport network: Travel time budget of travelers with heterogenuous risk aversion. *Transportation Research Part B 40, pp. 792-806.*

Lu, C.-C., Mahmassani, H.S. & Zhou, X. (2009). Equivalent gap-function-based reformulation and solution algorithm for the dynamic user equilibrium problem. *Transportation Research Part B 43, pp. 345-364.*

Meeuwissen, A.M.H., Snelder, M. & Schrijver, J.M. (2004). *Statische analyse variabiliteit reistijden voor SMARA.* TNO Inro rapport 2004-31. (in Dutch)

Merchant D.K. & Nemhauser G.L. (1978a). A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science 12 (3), pp. 183-199.*

Merchant D.K. & Nemhauser G.L. (1978b). Optimality conditions for a dynamic traffic assignment model. *Transportation Science 12 (3), pp. 200-207.*

Mercier, M. (2009). Traffic flow modelling with junctions. *Journal of Mathematical Analysis and Applications 350, pp. 369-383.*

Messmer, A. & Papageorgiou, M. (1990). METANET: A macroscopic. simulation program for motorway networks, *Traffic Engineering and Control 31, pp. 466-470.*

Miete, O. (2011). *Gaining new insights regarding traffic congestion, by explicitly considering the variability in traffic.* Master Thesis, Delft University of Technology.

Mun, J.-S. (2007). Traffic Performance Models for Dynamic Traffic Assignment: An Assessment of Existing models. *Transport Reviews 27 (2), pp. 231-249.*

Murray-Tuite, P.M. & Mahmassani, H.S. (2004). A Methodology for the determination of Vulnerable Links in a Transportation Network. *Transportation Research Record 1882, pp. 88-96.*

Nastase, D. & Harehdasht, S.A. (2010). *Comparison of Traditional and Novel Ways of Calculating Time-Dependent Equilibrium Patterns in Transportation Networks.* Master Thesis, Katholieke Universiteit Leuven.

Newell, G.F. (1993). A simplified theory of kinematic waves in highway traffic, Part I: General Theory, Part II: Queuing at freeway bottlenecks, Part III: Multi-destination flows. *Transportation Research B 27, pp. 281-313.*

Ng, M. & Waller, S.T. (2009). A Dynamic Route Choice Model in Face of Uncertain Capacities. Presented at *the 88th Annual Meeting of the Transportation Research Board*, Washington D.C.

Ngoduy, D., Hoogendoorn, S.P. & van Lint, J.W.C. (2005). Modeling Traffic Flow Operation in Multilane and Multiclass Urban Networks. *Transportation Research Record 1923, pp. 73-81*.

Ngoduy, D. (2006). *Macroscopic Discontinuity Modeling for Multiclass Multilane Traffic Flow Operations.* PhD Thesis TRAIL series, Delft University of Technology.

Ni, D. & Leonard, J.D. (2005). A simplified kinematic wave model at a merge bottleneck. *Applied Mathematical Modelling 29, pp. 1054-1072.*

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), pp. 1-14*.

Nie, X. & Zhang, H.M. (2005). A Comparative Study of Some Macroscopic Link Models used in Dynamic Traffic Assignment. *Networks and Spatial Economics 5 (1), pp. 89-115*.

Nie, Y., Ma, J. & Zhang, H.M. (2008). A Polymorphic Dynamic Network Loading Model. *Computer-Aided Civil and Infrastructure Engineering 23, pp. 86-103.*

Nie, Y.M. (2010). A Class of Bush-based Algorithms for the Traffic Assignment Problem. *Transportation Research Part B 44 (1), pp. 73-89*.

Noland, R.B. & Small, K.A. (1995). Travel time uncertainty, departure time choice, and the cost of morning commutes, *Transportation Research Record 1493, pp. 150–158.*

Noland, R.B., Small, K.A., Koskenoja, P.M. & Chu, X. (1998). Simulating travel reliability. *Regional Science and Urban Economics 28 (5), pp. 535-564.*

Noland, R.B. & Polak, J.W. (2002). Travel time variability: a review of theoretical and empirical issues. *Transport Reviews 22 (1), pp. 39-54.*

Osorio, C., Flötteröd, G. & Bierlaire, M. (2011). Dynamic network loading: a differentiable model that derives full link state distributions. *Transportation Research Part B 45 (9), pp. 1410-1423*.

Pavlis, Y. & Papageorgiou, M. (1999). Simple Decentralized Feedback Strategies for Route Guidance in Traffic Networks. *Transportation Science 33, pp. 264-278.*

Payne, H.J. (1971). Models of freeway traffic control. *Mathematical Models of Public Systems 1, pp. 51-61.*

Peeta, S. & Ziliaskopoulos, A.K. (2001). Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future. *Networks and Spatial Economics 1, pp. 233-265.*

Pel, A.J. (2011). *Transportation Modelling for Regional Evacuations*. PhD Thesis TRAIL series, Delft University of Technology.

Pel, A.J., Bliemer, M.C.J. & Hoogendoorn, S.P. (2009). Hybrid route choice modeling in dynamic traffic assignment. *Transportation Research Record 2091, pp. 100-107.*

Qian, Z. & Zhang, M. (2011). Computing individual path marginal cost in networks with queue spillbacks. Presented at *the 90th Annual Meeting of the Transportation Research Board*, Washington D.C.

Raadsen, M.P.H., Mein, E.H., Schilpzand, M.P. & Brandt, F. (2010). Implementation of a single dynamic traffic assignment model on mixed urban and highway transport networks including junction modelling. Presented at *the 3rd International Symposium on Dynamic Traffic Assignment (DTA 2010)*, Takayama.

Ran, B., David E. Boyce & Larry J. Leblanc (1993). A New Class of Instantaneous Dynamic User-Optimal Traffic Assignment Models. *Operations Research 41 (1), Special Issue on Stochastic and Dynamic Models in Transportation, pp. 192-202.*

Richards, P.I. (1956). Shockwaves on the highway. *Operations Research 4, pp. 42-51.*

Ritsema van Eck, J., Hilbers, H. & Schrijver, J.M. (2004). De reistijdbetrouwbaarheid op het Nederlandse wegennet gemodelleerd met SMARA. *Tijdschrift Vervoerswetenschap 40 (4).* (in Dutch)

Rubio-Ardanaz, J.M., Wu, J.H., Florian, M. (2001). A Numerical Analytical Model for the Continuous Dynamic Network Equilibrium Problem with Limited Capacity and Spillback. *Proceedings of the 2001 IEEE ITS Conference, pp. 263-267.*

Salz, A. & Horowitz, M. (1989). IRSIM: An Incremental MOS Switch-Level Simulator. *Proceedings of the 26th ACM/IEEE Design Automation Conference, pp. 173-178.*

Schilpzand, M. (2008). *X-Stream in MaDAM: New junction modelling in macroscopic dynamic traffic assignment models.* Graduation Thesis, University of Twente.

Schrijver, J. (2004). *Het voorspellen van betrouwbaarheid van reistijd met SMARA*. TNO Inro rapport (in Dutch)

Scott, D.M., Novak, D.C., Aultman-Hall, L. & Guo, F. (2006). Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks. *Journal of Transport Geography 14 (3), pp. 215–227.*

Small, K.A. (1982). The Scheduling of Consumer Activities: Work Trips. *The American Economic Review 72 (3), pp. 467-479.*

Small, K.A., Winston, C. & Yan, J. (2005). Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. *Econometrica 73 (4), pp. 1367-1382.*

Smith, M.J. (1979). The Existence, Uniqueness and Stability of Traffic Equilibrium. *Transportation Research Part B 13, pp. 295–304.*

Smith, B.L., Qin, L. & Vankatanarayana, R. (2003). Characterization of Freeway Capacity Reduction Resulting from Traffic Accidents. *Journal of Tranportation Engineerging 129 (4), pp. 362-368.*

Snelder, M., Wagelmans, A.P.M., Schrijver, J.M., van Zuylen, H.J. & Immers, L.H. (2007). Optimal Redesign of the Dutch Road Network. *Transportation Research Record 2029, pp. 72-79*.

Snelder, M. (2010). *Designing Robust Road Networks.* PhD Thesis TRAIL series, Delft University of Technology.

Sumalee, A., Zhong, R.X., Pan, T.L. & Szeto, W.Y. (2011). Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. *Transportation Research Part B 45 (3), pp. 507-533*.

Taale, H. (2008). *Integrated Anticipatory Control of Road Networks: A game theoretical approach.* PhD Thesis TRAIL series, Delft University of Technology.

Tamminga, G.F., Maton, J.C., Poorteman, R. & Zee, J. (2005). *De Robuustheidscanner - Robuustheid van netwerken: een modelmatige verkenning*. Publication I&M-99366053-GT/mk. Grontmij Nederland, i.o.v. Adviesdienst Verkeer & Vervoer. (in Dutch)

Tampère, C.M.J., Stada, J., Immers, L.H., Peetermans, E. & Organe, K. (2007). Methodology for Identifying Vulnerable Sections in a National Road Network *Transportation Research Record 2012, pp. 1-10.*

Tampère, C.M.J, Viti, F., Corthout, R. & Frederix, R. (2009). Predicting unreliability of travel times in congested regional and urban road networks. Presented at *the 4th Kuhmo-Nectar Conference,* Copenhagen.

Tampère, C.M.J. & Viti, F. (2010). Dynamic Traffic Assignment under Equilibrium and Non-Equilibrium: Do We Need a Paradigm Shift? Presented at *the 3rd International Symposium on Dynamic Traffic Assignment (DTA 2010)*, Takayama.

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 45 (1), pp. 289-309.*

Troutbeck, R.J. & Kako, S. (1999). Limited priority merge at unsignalized intersections. *Transportation Research Part A 33, pp. 291-304.*

Tu, H., van Lint, J.W.C. & van Zuylen, H.J. (2005). Real-time modelling travel time reliability on freeway. Proceedings of *the 10th Jubilee Meeting of the EURO Working Group of Transportation (EWGT)*, Poznan.

Ukkusuri, S.V. & Waller, S.T. (2006). Single-Point Approximations for Traffic Equilibrium Problem Under Uncertain Demand. *Transportation Research Record 1964, pp. 169-175.*

Ukkusuri, S.V., Ramadurai, G. & Patil, G. (2010). A robust transportation signal control problem accounting for traffic dynamics. *Computers & Operations Research 37 (5), pp. 869-879.*

Unnikrishnan, A. & Waller, S.T. (2009). User Equilibrium with Recourse. *Networks and Spatial Economics 9, pp. 575-593.*

van Hinsbergen, C., Tampère, C.M.J., van Lint, J. & van Zuylen, H. (2009). Urban intersection in first order models with the Godunov scheme. Presented at *mobil.TUM 2009 – International Scientific Conference on Mobility and Transport – ITS for large Cities*, Munich.

Viti, F. & Tampère, C.M.J. (2010). Dynamic Traffic Assignment: Recent Advances and New Theories towards Real-time Applications and Realistic Travel Behavior (Editorial). *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment, pp. 1-25.*

Wakabayashi, H. & Iida, Y. (1992). Upper and lower bounds of terminal reliability in road networks: an efficient method with Boolean algebra. *Journal of Natural Disaster Science 14 (1), pp. 29–44.*

Wardrop, J.G. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers Part II 1, pp. 325-378.*

Watling, D. & Hazelton, M.L. (2003). The Dynamics and Equilibria of Day-to-Day Assignment Models. *Networks and Spatial Economics 3, pp. 349–370.*

Webster, F.V. (1958). *Traffic Signal Settings*. Road Research Laboratory Technical Paper No. 39.

Wohl, M. & Martin, B.V. (1967). *Traffic System Analysis for engineers and planners*. McGraw-Hill Series in Transportation.

Wu, N. (2000). Determination of Capacity at All-Way Stop-Controlled Intersections. *Transportation Research Record 1710, pp. 205-214.*

Yperman, I., Logghe, S., Tampère, C.M.J. & Immers, L.H. (2006). The Multi-Commodity Link Transmission Model for Dynamic Network Loading. Presented at *the 85th Annual Meeting of the Transportation Research Board*, Washington D.C.

Yperman, I. & Tampère, C.M.J. (2006). Multi-commodity Dynamic Network Loading with Kinematic Waves and Intersection Delays. Presented at *the First International Symposium on Dynamic Traffic Assignment (DTA2006)*, Leeds.

Yperman, I. (2007). *The Link Transmission Model for Dynamic Network Loading*. Ph.D. Thesis, Katholieke Universiteit Leuven.

AUTHOR'S PUBLICATIONS

Journal articles and book chapters

Corthout, R., Tampère, C.M.J., Frederix, R. & Immers, L.H. (forthcoming). Improving the efficiency of dynamic network loading applications through marginal simulation. Submitted to *Transportation Research Part C*.

Corthout, R., Flötteröd, G., Viti, F. & Tampère, C.M.J. (2012). Non-unique flows in macroscopic first-order intersection models. *Transportation Research Part B 46 (3), pp. 343-359*.

Frederix, R., Viti, F., Corthout, R. & Tampère, C.M.J. (2011). A New Gradient Approximation Method For Dynamic Origin-Destination Matrix Estimation On Congested Networks. *Transportation Research Record 2263, pp. 19-25.*

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 45 (1), pp. 289-309.*

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2010). Stochastic Dynamic Network Loading for Travel Time Variability due to Incidents. *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment, pp. 179-198.*

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2009). Marginal Incident Computation, an Efficient Algorithm to Determine Congestion Spillback due to Incidents. *Transportation Research Record 2099, pp. 22-29.*

Peer-reviewed conference articles

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2011). Combining Drivers' Rerouting Response and Marginal Computation for Variability Studies and Traffic Management Applications. Presented at *the 2nd International Conference on Models and Technologies for Intelligent Transportation Systems*, Leuven.

Frederix, R., Viti, F., Corthout, R. & Tampère, C.M.J. (2011). A new gradient approximation method for dynamic origin-destination matrix estimation on congested networks. Presented at the 90th Annual Meeting of the Transportation Research Board, Washington DC.

Corthout, R., Tampère, C.M.J., Frederix, F. & Immers, L.H. (2011), Marginal Dynamic Network Loading for Large-scale Simulation-based Applications. Presented at *the 90th Annual Meeting of the Transportation Research Board*, Washington DC.

Corthout, R., Tampère, C.M.J., Viti, F. & Immers, L.H. (2010). Macroscopic intersection modelling in simulation-based dynamic network loading. Presented at *the 11th TRAIL Congress*, Rotterdam.

Corthout, R., Tampère, C.M.J. & Deknudt, P. (2010). Assessment of Variable Speed Limits from the Drivers' Perspective. *Proceedings of the 13th International IEEE Annual Conference on Intelligent Transportation Systems, pp 499-506*.

Tampère, C.M.J., Corthout, R., Viti, F. & Cattrysse, D. (2010). Link Transmission Model: an efficient dynamic network loading algorithm with realistic node and link behavior. Presented at *the Tristan VII conference*, Tromsö.

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2010). Quantifying variability due to incidents including en-route rerouting. Presented at *the Tristan VII conference*, Tromsö.

Tampère, C.M.J., Viti, F., Corthout, R. & Frederix, R. (2009). Predicting unreliability of travel times in congested regional and urban road networks. Presented at *the 4th Kuhmo-Nectar Conference,* Copenhagen.

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2009). Marginal Incident Computation: Efficient Algorithm to Determine Congestion Spillback due to Incidents. Presented at *the 88th Annual Meeting of the Transportation Research Board*, Washington DC.

Corthout, R., Tampère, C.M.J., Van Oudheusden, D., Cattrysse, D. & Immers, L.H. (2008). A generic node model for unsignalized intersections in dynamic macroscopic simulation. Presented at *the 12th Meeting of the Euro Working Group on Transportation*, Ischia.

Corthout, R., Tampère, C.M.J. & Immers, L.H. (2008). Stochastic dynamic network loading for travel time variability due to incidents. Presented at *the 2nd International Symposium on Dynamic Traffic Assignment (DTA2008)*, Leuven.