KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK ESAT-SCD (SISTA)
Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee)

MODELLING TRAFFIC ON MOTORWAYS: STATE-OF-THE-ART, NUMERICAL DATA ANALYSIS, AND DYNAMIC TRAFFIC ASSIGNMENT

Promotoren:
prof. dr. ir. B. DE MOOR
prof. ir. L.H. IMMERS

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Sven MAERIVOET

Juni 2006

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK ESAT-SCD (SISTA)
Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee)

MODELLING TRAFFIC ON MOTORWAYS: STATE-OF-THE-ART, NUMERICAL DATA ANALYSIS, AND DYNAMIC TRAFFIC ASSIGNMENT

Examencommissie:
prof. dr. ir. P. Van Houtte, voorzitter
prof. dr. ir. B. De Moor, promotor
prof. ir. L.H. Immers, co-promotor
prof. dr. ir. J. Vandewalle
prof. dr. ir. A. Barbé
prof. dr. ir. G. Campion (UCL, Louvain-la-Neuve)
dr. ir. B. De Schutter (TU Delft)
dr. A. Schadschneider (Universität zu Köln)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Sven MAERIVOET

U.D.C. 656

Juni 2006

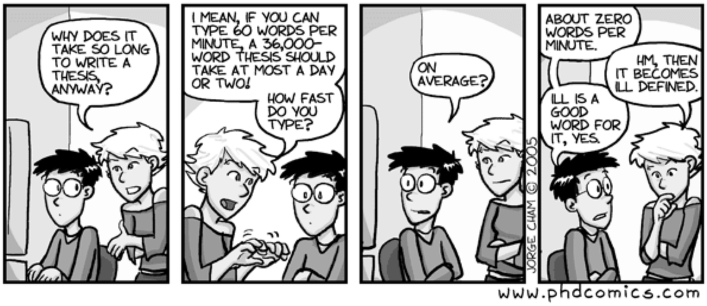
Wettelijk depot D/2006/7515/42

ISBN 90-568-2708-1

“Hofstadter’s Law: It always takes longer than you expect,
even when you take into account Hofstadter’s Law.”

— Douglas A. Hofstadter

Dankwoord



"Piled Higher and Deeper" by Jorge Cham[1]

Het is een vermaard gegeven, het dankwoord is absoluut en onafstrijdbaar het meest gelezen deel van een doctoraat. Naast de begeleidingscommissie die er haar oordeel over velt, en die vier mensen in de wereld die het ter hand nemen en ook daadwerkelijk gebruiken, bestaat het overgrote deel van het lezerspubliek uit zij die dit schrijfsel via het Internet downloaden of die het op de dag van de verdediging in fysieke vorm openslagen. Daarbij geldt trouwens ook de universeel academische waarheid, dat van alle stukken die er in een dergelijk boekje worden neergeschreven, het dankwoord ongetwijfeld tot het meest plezante wordt gerekend; hier vloeien de woorden pas echt (voor zij die het wensen te weten, de meest afgrijselijke delen om te schrijven zijn de Nederlandse samenvatting, de abstract, de vertaling van de abstract, de inleiding en de conclusies, in die volgorde).

Goed, laat ons even terugkeren naar de essentie van het alles, namelijk het tot stand komen van een dergelijk stukje tekst. Een typische platitude die menig auteur hier pleegt te verkondigen, is het feit dat je een doctoraat niet alleen schrijft … Maar als we eerlijk zijn, dan moeten we toch toegeven dat het, ondanks de netwerk-kwaliteiten van menig assistent, vaak vele eenzame uren achter de computer zijn. Voor sommigen is doctoreren een job die ze van negen tot vijf doen, echter voor mij is dat absoluut niet het geval. Denk ik maar even terug aan de meest bizarre uren waarop ik soms het toetsenbord beroerde, de rare manier van leven die zich uitte in een halve onregelmaat

[1] http://www.phdcomics.com

qua eettijden, het concept ‘ochtend’ dat plots al zijn betekenis verliest, et cetera. Eraan beginnen is niet het grote werk, maar het afwerken dat is vaak de karaktertest. Hierbij valt me plots op dat ik zelf de niet-zo-fijne dingen vaak tot op het laatste moment uitstel. Misschien had ik toch maar beter Joan Bolker’s boek *“Writing Your Dissertation in Fifteen Minutes a Day”* [Bol98] op voorhand eens gelezen. Nu ja, de afgelopen jaren waren best een levendige periode, met de gekende toppen en dalen (de eerste zitten vaak aan het begin en einde van een hoofdstuk, de laatste in de schrijfperiode daartussen).

Ok, het is zover ... tijd voor het echte werk.

El numero uno is de geprezen promotor Bart De Moor, of ‘den BDM’ zoals hij in de wandelgangen en op die gele Post-it notes wordt genoemd. Qua dynamiek, flamboyantie en stijl steekt hij met kop en schouders boven iedereen uit. Onder de vaak gehoorde kenmerken vallen zijn inspirerende kracht, zijn ambitieuze ideeën, zijn enthousiasme, ... Mijn persoonlijke inbreng in deze laudatio is dat ik hem dank voor de vrijheid van meningsuiting, wat soms resulteerde in nogal harde en luide discussies die we hielden over de zin en onzin van doctoreren, de doctoraatsopleiding, de administratie, ... Ik diende me hierbij nooit een blad voor de mond te nemen, en als ik al eens te ver ging met de scherpe speerpunt die ik vooral in het begin was, dan gaf hij wel gepaste repliek waardoor ik de volgende keer twee keer nadacht. Gezien zijn drukke agenda verliep onze communicatie bijna uitsluitend over e-mail, en nu ik er bij stilsta, geschiedde dit ook meestal ’s nachts. Dank ook voor de talrijke gastlezingen die ik voor je mocht geven, ze zorgden ervoor dat ik mijn boekenplank degelijk heb kunnen spijzen. Ook bedankt om me onder je vleugels in SISTA te nemen, en me de kans te geven mijn doctoraat verder af te kunnen werken in de periode dat het plots veel langer duurde dan ik aanvankelijk had gedacht (tiens, ik hoor hier Hofstadter klinken).

Speciale dank gaat uit naar mijn co-promotor professor Ben Immers, de man die me bruisend van energie vaak de zonnige zijde van een doctoraat liet zien. Discussies met hem waren steeds inspirerend, en vaak zat ik gretig te luisteren naar zijn relaas over de projecten die hij met TNO in Nederland uitvoerde. En of je nu bij hem in zijn bureau zit, met hem tafelt in een restaurant, of een glas nuttigt op een terras, zijn enthousiasme zal altijd even aanstekelijk werken.

Ik bedank ook de andere leden van de jury voor hun bereidheid om mijn proefschrift te lezen en beoordelen, meerbepaald professor Paul Van Houtte als voorzitter, professor Joos Vandewalle waarmee het altijd fijne momenten waren tijdens het mondelinge examen van HK05, professor Guy Campion om deel van de begeleidingscommissie en jury te willen uitmaken, professor André Barbé om op de valreep nog in mijn jury te willen zetelen en op heel korte tijd de grote brok tekst door te lezen, Bart De Schutter voor de vele waardevolle opmerkingen en suggesties (hij liet me de puntjes op de ‘i’ zetten); I also like to thank Andreas Schadschneider for agreeing to participate in my PhD jury.

SISTA is een dynamisch clubje mensen, individuen die niet schrikken van woorden zoals systeem, chaos, regeling, verkeer, quantum, genoom, tensor, classificatie, LS-SVMs, spraak, cryptografie, document, bier, warme hapjes en recepties. Onder de vele collega's vermeld ik in het bijzonder mijn bureaugenoten, namelijk Tom Bellemans waarmee ik gedurende twee jaar de onderafdeling 'Verkeer@SISTA' vormde; Steven Bex die ik vaak tegenkwam op recepties, ofwel tijdens een meestal-lang-uitgelopen babbel telkens ik naar het tweede afzakte; Maarten Van den Nest met zijn immer cryptische formules waar ik nog altijd totaal geen snars van begrijp (die QIT'ers zijn een raar ras); Dries Van Dromme die me een glimp van de evil-world der table-top gamers liet opvangen ("The invasion shall take place, burning the sky, cleansing the earth, no one can escape the Zerg"). Een speciaal woord van dank gaat ook uit naar zij die de administratieve ruggengraat van SISTA vormen, namelijk Ida Tassens, Ilse Pardon, Veerle Duchateau, Péla Noé en Bart Motmans.

Ook de collega's Isaak Yperman, Chris Tampère en Jim Stada van de vakgroep Verkeer en Infrastructuur wens ik te bedanken voor de verkeerskundige babbels. In het bijzonder vermeld ik hier Steven Logghe, zonder wie ik waarschijnlijk nooit in Leuven was terecht gekomen, ware het niet dat hij op een dag tijdens het Googlen mijn digitale persoonlijkheid tegenkwam. Ik dank hem ook voor de vele fijne en puntige wetenschappelijke debatten die soms zo passioneel gevoerd werden dat een luide stemverheffing geen uitzondering was, alsmede voor het nalezen van mijn tekst en het geven van constructieve opmerkingen.

Ik dank het Federaal Wetenschapsbeleid (wat ooit de afkorting DWTC droeg) voor de financiering van mijn onderzoek. Ook Stefaan Hoornaert van het departement Mobiliteit en Openbare Werken van het Verkeerscentrum Vlaanderen wens ik te bedanken voor de massa's verkeersmetingen (tellingen) die ik voor mijn onderzoek toegestuurd kreeg.

Tot slot dank ik mijn vrienden van de jeugdbeweging K.S.A. 'Vlaamse Kerels' Zwijndrecht; op de weinige momenten die me nog resteerden kon ik me bij hen uitleven zonder me zorgen te maken. Ook mijn familie en in het bijzonder mijn ouders verdienen een woord van dank voor de mentale steun en blijvende motivatie; het zijn twee zeer fijne mensen die altijd in mij geloofden.

En dan is er Sanne; jij bracht de andere helft van regelmaat in mijn leven. Zonder jou zat ik nu waarschijnlijk aan hoofdstuk 327 te schrijven en raakte dit doctoraat nooit af. Ik dank jou voor je lieve woorden, je warme glimlach, de momenten dat je er gewoon was en geen vragen stelde, de andere momenten waarop je de ene na de andere vraag afvuurde en terecht achter mijn veren zat, de spiegel die je vaak voor me was telkens je me confronteerde met mijn eigen manier van doen, en hoe je me recht trok als ik te zeer van het pad afweek en dreigde verloren te lopen. Sanne, een kus want je bent de ene uit de 6.6 miljard !

Sven Maerivoet
Leuven, 27 juni 2006

Abstract

With the levels of congestion in cities and countries showing an ever-increasing trend, the modelling of road traffic continues to be a highly active field. Whereas numerous efforts have been undertaken towards the local and global control of traffic flows, our research is aimed at the modelling part of road traffic, more specifically traffic on motorways.

The goal of this dissertation is three-fold; for starters, we provide a complete nomenclature convention within traffic flow theory, built upon a consistent set of notations. In continuation, we give an in-depth literature survey on the mathematical models used for describing road traffic flows, both from a transportation planning and a flow propagation point of view. Special attention is given to the class of cellular automata models of road traffic. Secondly, we perform an exploratory data analysis of raw traffic flow measurements, discussing the operational characteristics of single-loop detectors. This analysis also provides researchers with tools to track statistical outliers, to quickly assess structural and incidental detector failures, to estimate travel times in an off-line fashion based on raw cumulative counts, and to obtain a visual representation of traffic flow dynamics in time and space. Finally, we provide, within the context of simulation-based dynamic traffic assignment, a straightforward method to tackle both departure time choice and dynamic route choice problems in a sequential manner, built around a traffic flow model that is represented as a computationally efficient cellular automaton.

Our contributions to the field of literature are distinct, in that such comprehensive overviews hitherto only existed in scattered form, whereas we provide a synthesis of the approaches concerning the description of road traffic flows. Furthermore, in contrast to most research on the numerical analysis of traffic flow measurements, we offer methods that are capable of dealing with large-scale data sets in order to get a global picture regarding the quality of the measurements. Finally, as opposed to many approaches towards the paradigm of simulation-based dynamic traffic assignment, we propose a methodology that sequentially integrates departure time choice with route choice within a simulation framework.

Korte samenvatting

Terwijl de filevorming in steden en landen een immer-toenemende trend vertoont, wordt het modelleren van wegverkeer een steeds maar actiever vakgebied. Daar waar reeds vele inspanningen werden gedaan met betrekking tot de lokale en globale regeling van verkeersstromen, is ons onderzoek gericht op het modelleren van wegverkeer op autosnelwegen.

Het doel van ons onderzoek is drievoudig; eerst geven we een volledige standaard omtrent nomenclatuur binnen het gebied van de verkeerskunde, gebaseerd op een consistente verzameling notaties. Dit wordt gevolgd door een gedetailleerd literatuuroverzicht omtrent de wiskundige modellen die gebruikt worden om het verkeer op wegen te beschrijven, dit vanuit zowel het standpunt van transportplanning als stromingsmodellen. Speciale aandacht gaat uit naar de klasse van cellulaire automaatmodellen van wegverkeer. Ten tweede voeren we een verkennende data analyse van ruwe verkeersmetingen uit, waarbij we de operationele karakteristieken van enkelvoudige lusdetectoren bespreken. Verder reiken we onderzoekers middelen aan om statistische uitschieters op te sporen, om op een snelle manier structurele en incidentele storingen van detectors te beoordelen, om reistijden te schatten op een off-line manier, gebaseerd op ruwe cumulatieve tellingen, en om een visuele voorstelling van de dynamica van verkeersstromen in tijd en ruimte te verkrijgen. Tot slot, voorzien we, binnen de context van simulatie-gebaseerde dynamische verkeerstoedeling, een duidelijke methode om zowel de problemen van de keuzes van vertrektijdstip en route op sequentiële wijze te combineren, dit gebouwd rond een verkeersstroommodel dat uitgewerkt wordt als een computationeel efficiënte cellulaire automaat.

Met betrekking tot de literatuur onderscheiden onze bijdragen zich doordat ze een synthese vormen van de benaderingen voor het beschrijven van wegverkeer, terwijl dergelijke samenvattingen tot op heden enkel verspreid bestonden. Om een globaal beeld te krijgen met betrekking tot de kwaliteit van verkeersmetingen, bieden wij daarnaast methodes aan die kunnen omgaan met grootschalige data, dit in tegenstelling tot het meeste onderzoek naar de numerieke analyse van verkeersmetingen wat vaak slechts op beperkte data wordt uitgevoerd. Tenslotte met betrekking tot de vele benaderingen van het paradigma van simulatie-gebaseerde dynamische verkeerstoedeling, stellen wij een methodologie voor die de keuze van het vertrektijdstip sequentieel met de routekeuze integreert.

Table of contents

Chapter 1

Introduction

Considering the current road traffic problems in cities and countries, it is becoming more apparent each day that we can not completely solve congestion. But all is not lost, as we can try to alleviate it in some way, by making the journeys as comfortable as possible, perhaps even diminishing the delays (which is an entirely different thing than eliminating them). It still remains a conundrum to tackle road traffic congestion on a global scale, requiring an integrated approach that combines several control techniques, e.g., the advanced traffic management systems (ATMS) such as dynamic route guidance, ramp metering, speed harmonisation, tidal flows, . . . , and policy measures decided upon by (local) governments. These latter are finding root in methods such as congestion pricing which is gaining appreciation, better and cheaper public transportation, . . . to even some of the most bizarre proposals encountered, e.g., our own liberal senator Jean-Marie De Decker who boldly put forward the concept of 'double-deck motorways' as a method to expand capacity, thereby reducing congestion. In contrast to some of the jump-the-gun measures, smoother traffic operations should be accomplished by using the existing road network, without the need for new infrastructure (note however that local adaptations of the current infrastructure are still allowed). The aim of our dissertation is to provide road traffic engineers with a solid background in road transportation modelling, whereby we spend attention on the literature part, as well as the analysis of numerical data, and the development of a framework for performing integrated dynamic traffic assignment.

In the introduction of this dissertation, we briefly depict the background on which our research was conducted, as well as the goals that were set. The subsequent section then provides a road map of the structure of the dissertation, after which the final section gives a chapter-by-chapter overview and highlights our own contributions.

1.1 Background and research goals

Over the course of the last ten to fifteen years, we have noticed how other scientific fields besides the traditional mathematical, physics, and engineering disciplines have entered the field of transportation research. Recognising the fact that individual drivers are human beings that perform simultaneous and complex operations, increases the push towards more psychologically-oriented domains (e.g., to investigate the reasons why collisions occur, why a driver's attention fails, . . .). At an even higher level, we see how collective dynamics, i.e., the socio-economic behaviour of large groups of travellers, has entered the field.

We have also seen an evolution towards more autonomy in vehicles, with older technological examples such as anti-locking brake system (ABS), electronic power steering (EPS), electronic stability programme (ESP), traction control system (TCS), . . . More recently, we notice an increased degree of advanced technologies such as lane guidance systems, e.g., Volvo's emergency lane assist (ELA), adaptive cruise control (ACC), collision avoidance systems, . . . One of the most culminating highlights in this area, is undoubtedly the *2005 Grand Challenge*[1] of the USA Department of Defense's (DoD) Defense Advanced Research Projects Agency (DARPA); unmanned vehicles were required to drive autonomously over a course of some 200 kilometres in the Mojave Desert.

As technological progress is more than ever present, this leads to the application of theories to the real world, e.g., the implementation of traffic control measures). And even though the modelling aspect remains important, the time has come to look at what is practically possible with respect to concrete applications and implementations. Today, powerful mathematical models can be put in computer, allowing them to be used, e.g., for predictions in an on-line control setting. Whether or not simple or complex models are used, it is the application that has become important, i.e., it is actually time to do something with all our knowledge. Note that although most of the discussed methods are also applicable to city traffic, the work in this dissertation is primarily aimed towards motorways.

From this perspective, the first goal of this dissertation is to provide practitioners in the field with a solid background in the modelling of road transportation. We still encounter a frequent confusion among traffic engineers and policy makers when it comes to transportation planning models and the role that traffic flow models play therein. The literature survey given in this dissertation, is unique in that it provides a rather complete overview, thereby eliminating the need to look for answers in the zoo of papers and notations that currently exists.

A second goal of this dissertation is aimed at the numerical data analysis of raw traffic flow measurements. Due to the advent of powerful, yet affordable, desktop computers, it has now become possible to perform large-scale data analyses. We therefore provide researchers with tools to track outliers, quickly assess structural and incidental

[1] http://www.darpa.mil/grandchallenge05/index.html

detector failures, a method for off-line travel time estimation, and reliability indicators based on tempo-spatial maps.

The third and final goal of this dissertation, finds its roots in the concept of dynamic traffic assignment. The current evolution in the scientific field is to endogeneously include both departure time choice (i.e., when will a commuter depart for his/her journey ?) and dynamic route choice (i.e., which route will the commuter be taking ?). We provide a straightforward method to tackle both problems in a sequential manner, built around a traffic flow model that is represented as a computationally efficient cellular automaton. The efficiency is furthermore enhanced with the concept of parallellisation through distributed computing.

1.2 Structure of the dissertation

The dissertation is divided into four large parts, centred around (i) *the physics of road traffic and transportation* (Chapters 2 and 3), (ii) *cellular automata models of road traffic* (Chapters 4 and 5), (iii) *numerical analysis of traffic data* (Chapter 6), and (iv) *integrated dynamic traffic assignment* (Chapter 7). The dissertation also contains a part with conclusions and perspectives (Chapter 8), and four appendices.

In Figure 1.1 we provide a road map that depicts the logical structure and coherence between the chapters: starting from Chapter 2 *"Traffic flow theory"*, the reader can move on to Chapter 3 *"Transportation planning and traffic flow models"*. From then on, the trajectory splits: on the one hand we have Chapter 6 *"Data quality, travel time estimation, and reliability"* which is rather self-contained, on the other hand we have Chapter 4 *"Traffic cellular automata"*. This latter Chapter finds continuation in Chapter 5 *"Relating the dynamics of the STCA to the LWR model"*, and draws upon the (didactical) software described in Appendix B *"TCA+ Java™ software"*. Finally, Chapter 7 deals with *"Dynamic traffic assignment based on cellular automata"*.

At the end of the dissertation, Appendix A provides the reader with a comprehensive glossary of terms, divided into a list with acronyms and abbreviations, and a list of symbols for each chapter separately. Appendix C talks about some thoughts related to the steps towards obtaining a PhD degree, and the final Appendix D provides a Dutch summary. The last three parts of the dissertation give an extensive list of literature references, and lists of the author's publications and presentations.

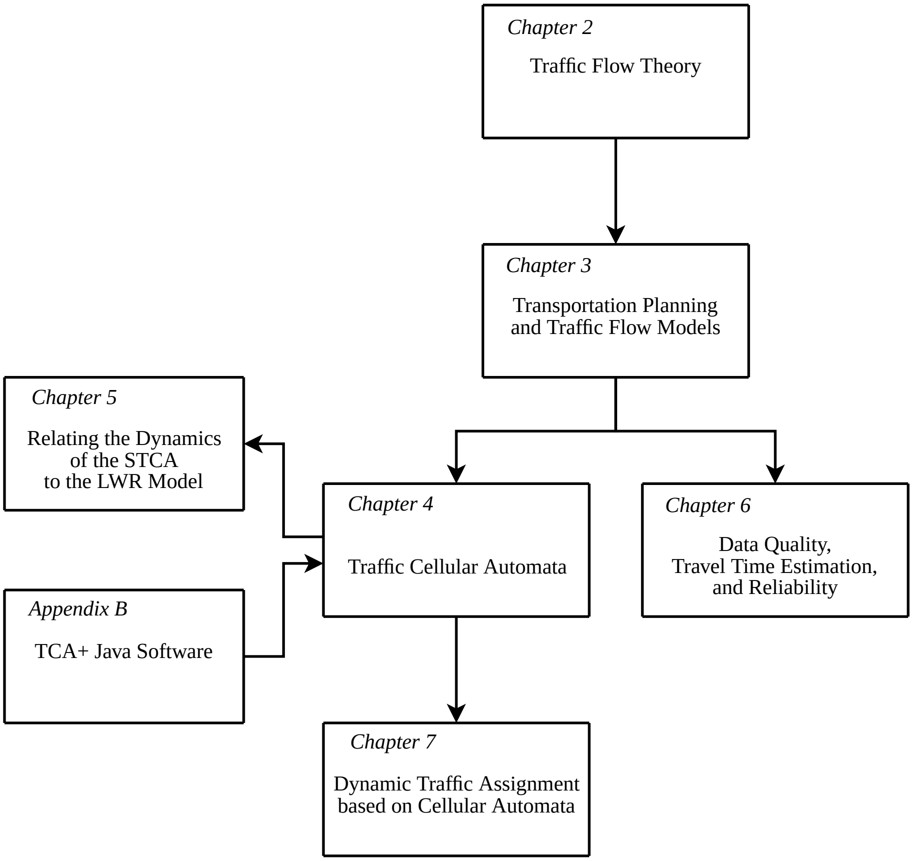


Figure 1.1: A road map depicting the logical structure and coherence between the chapters in this dissertation.

1.3 Overview and contributions to the state-of-the-art

Chapter 2 – "*Traffic flow theory*", provides a complete nomenclature convention, built upon a consistent set of notations. These encompass the classical traffic flow variables, some performance indicators, and a description of the different traffic flow regimes and the correlations between the traffic flow characteristics. Finally, we also discuss some of the different points of view with respect to the causes of congestion, as adopted by the traffic engineering community.

Chapter 3 – "*Transportation planning and traffic flow models*", gives a comprehensive overview of transportation planning models, operating on a high level, and traffic flow models that explicitly describe the physical propagation of traffic flows, typically on a lower level. We first focus on land-use models, both for the classical and modern approach, after which we highlight the traditional trip-based transportation models, followed by an elaboration of the activity-based approach. Before moving on to traffic flow propagation models, we give a brief account of the field of transportation economics, concluding with a view on road pricing policies. From then on, the chapter gives detailed information on the macroscopic, mesoscopic, and microscopic traffic flow models. Our contribution to the state-of-the-art in the literature, is an integrated survey that is able to help any researcher wishing to partake in the field, thereby alleviating the need to dive into tons of course texts, papers, ... most of the time spread over different scientific areas. Note that our work excludes fields such as traffic control theory and practice, as this is not the focus of our research.

Chapter 4 – *"Traffic cellular automata"*, details the field of traffic cellular automata (TCA) models; they allow for computationally efficient, yet still detailed enough, calculations of the propagation of traffic flows. Already, several reviews of TCA models exist, but none of them considers all the models exclusively from the behavioural point of view, as we do. As this kind of survey did not hitherto exist in the current scientific field, our overview fills this void, caused by the need for researchers to have such a comprehensive insight. In the chapter, we first recount the historical background of cellular automata, after which we provide a mathematical description of them, including methods for performing traffic flow measurements on their lattices when applied to vehicular traffic. We then classify the existing TCA models in on the one hand single-cell and on the other hand multi-cell models. The former include deterministic, stochastic, and slow-to-start models. The overview of the latter first sheds some light on an at-first-sight unexpected hysteresis phenomenon related to the use of a multi-cell setup. The chapter ends with a focus on multi-lane traffic, city traffic, and results obtained when converting these TCA models into an analytical form.

Chapter 5 – *"Relating the dynamics of the STCA to the LWR model"*, bridges a gap between microscopic and macroscopic models, by explaining an alternate methodology that implicitly incorporates the STCA's stochasticity into the macroscopic first-order LWR model. The innovative aspect of our approach, is that we derive the LWR's fundamental diagram directly from the STCA's rule set, by assuming a stationarity condition that converts the STCA's rules into a set of linear inequalities. These constraints define the shape of the fundamental diagram, which is then specified to the LWR model. We apply the methodology to a small theoretical case study, leading to the conclusion that, although for noise-free systems our method is exact, it becomes very important to correctly capture the capacities in both the STCA and LWR models in the presence of noise.

Chapter 6 – *"Data quality, travel time estimation, and reliability"*, gives a detailed account of the procedures followed when aggregating traffic flow measurements by means of single inductive loop detectors (SLDs) embedded in Flanders' road network. In a subsequent investigation, we uncover a significant discrepancy between the mean speeds as estimated by the SLDs, and those explicitly calculated by the presumably employed algorithm. We then implement a methodology that tracks outliers in traffic flow data, from a statistical point of view. After providing several methods for dealing with missing values, we develop a visual technique based on maps, allowing a quick assessment of structural and incidental detector malfunctioning. Furthermore, we provide a method for off-line travel time estimation based on raw traffic counts; after constructing cumulative curves, the methodology performs synchronisation of these curves, automatically taking into account systematic errors. From that point on, it is possible to estimate the distributions of the travel time on a closed section. The final part of the chapter gives a means for visualising reliability and robustness properties of traffic flow dynamics, based on tempo-spatial maps that provide an extra instrument for the analysis of recurrent congestion. Results are presented for applications to the E19 motorway and R0 ring road.

Chapter 7 – *"Dynamic traffic assignment based on cellular automata"*, elaborates upon the development of a framework that allows us to perform dynamic traffic assignment (DTA). We first describe some previous approaches towards DTA, both from an analytical and a simulation-based perspective. We then propose a methodology for performing simulation-based integrated DTA, by which we mean the sequential inclusion of both departure time choice (DTC) and dynamic route choice (DRC). The second part of the chapter elaborates upon the underlying dynamic network loading (DNL) model, which is represented as a computationally efficient cellular automaton. After providing a functional description and some code implementation details, we explain a technique that further increases the efficiency by adopting the concept of parallellisation through distributed computing, i.e., dividing the total work load over several distinct central processing nodes.

Summarising, the main contributions of this dissertation are:

- Providing a logical and consistent terminology and notation for denoting traffic flow variables (Chapter 2).

- Giving an overview of what is currently the state-of-the-art with respect to traffic flow theory, more specifically centred around relations between traffic flow characteristics, the causes of congestion, transportation planning models, and traffic flow propagation models (Chapters 2 and 3).

- Detailing the field of traffic cellular automata models with a complete survey and classification from the behavioural point of view. We focus on the historical background, a mathematical description, single- and multi-cell models (deterministic, stochastic, and slow-to-start), single-, multi-lane, and city traffic, and analytical approximations (Chapter 4).

- Explaining a possible alternate methodology that incorporates the stochasticity of a traffic cellular automaton model into a first-order deterministic macroscopic model (Chapter 5).

- Providing a method to track statistical outliers in traffic flow measurements (giving pointers for dealing with missing values) and developing a visual technique for quick assessments of structural and incidental detector malfunctioning (Chapter 6).
- Developing a methodology for deriving travel times on a closed section of the road, based on raw cumulative counts, thereby estimating the distribution of the travel time. Visualising traffic flow dynamics, based on tempo-spatial maps that indicate recurrent congestion (Chapter 6).
- Proposing a framework for dynamic traffic assignment, in which departure time choice and dynamic route choice (pre-route choice) are sequentially combined with an efficient dynamic network loading model (Chapter 7).

Part I

The Physics of Road Traffic and Transportation

Chapter 2

Traffic flow theory

The scientific field of traffic engineering encompasses a rich set of mathematical techniques, as well as researchers with entirely different backgrounds. This chapter provides an overview of what is currently the state-of-the-art with respect to traffic flow theory. Starting with a brief history, we introduce the microscopic and macroscopic characteristics of vehicular traffic flows. Moving on, we review some performance indicators that allow us to assess the quality of traffic operations. A final part of this chapter discusses some of the relations between traffic flow characteristics, i.e., the fundamental diagrams, and sheds some light on the different points of view with respect to the causes of congestion, as adopted by the traffic engineering community.

Because of the large diversity of the scientific field (engineers, physicists, mathematicians, … all lack a unified standard or convention), one of the principal aims of this chapter is to define both a logical and consistent terminology and notation. It is our strong belief that such a consistent notation is a necessity when it comes to creating order in the 'zoo of notations' that in our opinion currently exists.

> *We stimulate practitioners from all trades to adopt these conventions; as such, they have a common ground that disposes of the intrinsic hassles when reinterpretating another one's thoughts. Take for example an engineer with a background in control theory, wishing to exchange ideas with an engineer having its roots in, e.g., fluid dynamics. When talking about densities, the latter uses a letter 'k', whereas the former will frequently use the Greek letter 'ρ', because in his domain a 'k' typically means a discrete time step. This leads to possible ambiguous interpretations, as the former uses the letter 'ρ' to denote occupancies. Adopting a shared convention can therefore bridge both worlds and settle the confusion. In this respect, we believe that practitioners writing for the international field of traffic flow theory, should stick to our proposed standard, thereby putting the emphasis on the common part around traffic flow theory and not on their own specific scientific field. In the previous example, this amounts to using, e.g., $t \in \mathbb{N}$ for the time step.*

For a concise but complete overview of all abbreviations and notations proposed and adopted throughout this dissertation, we refer the reader to Appendix A.

2.1 A brief history of traffic flow theory

Historically, traffic engineering got its roots as a rather practical discipline, entailing most of the time a common sense of its practitioners to solve particular traffic problems. However, all this changed at the dawn of the 1950s, when the scientific field began to mature, attracting engineers from all sorts of trades. Most notably, John Glen Wardrop instigated the evolving discipline now known as traffic flow theory, by describing traffic flows using mathematical and statistical ideas [War52].

During this highly active period, mathematics established itself as a solid basis for theoretical analyses, a phenomenon that was entirely new to the previous, more 'rule-of-thumb' oriented, line of reasoning. Two examples of the progress during this decade, include the fluid-dynamic model of Michael James Lighthill, Gerald Beresford Whitham, and Paul Richards (or the *LWR model* for short) for describing traffic flows [Lig55; Ric56], and the car-following experiments and theories of the club of people working at General Motors' research laboratory [Cha58; Gaz59; Her59; Gaz61]. Simultaneous progress was also made on the front of economic theory applied to transportation, most notably by the publication of the 'BMW trio', Martin Josef Beckmann, Charles Bartlett McGuire, and Christopher Blake Winsten [Bec55].

From the 1960s on, the field evolved even further with the advent of the early personal computers (although at that time, they could only be considered as mere computing units). More control-oriented methods were pursued by engineers, as a means for alleviating congestion at tunnels and intersections, by, e.g., adaptively steering traffic signal timings. Nowadays, the field has been kindly embraced by the industry, resulting in what is called *intelligent transportation systems* (ITS), covering nearly all aspects of the transportation community.

In spite of the intense booming during the 1950s and 1960s, all progress seemingly came to a sudden stop, as there were almost no significant results for the next two decades (although there are some exceptions, such as the significant work of Ilya Prigogine and Robert Herman, who developed a traffic flow model based on a gas-kinetic analogy [Pri71]). One of the main reasons for this, stems from the fact that many of the involved key players returned to their original scientific disciplines, after exhausting the application of their techniques to the transportation problem [New02a]. Note that despite this calm period, the application of control theory to transportation started finding new ways to alleviate local congestion problems.

At the beginning of the 1990s, researchers found a revived interest in the field of traffic flow modelling. On the one hand, researchers' interests got kindled again by the appealing simplicity of the LWR model, whereas on the other hand one of the main boosts came from the world of statistical physics. In this latter framework, physicists tried to model many particle systems using simple and elegant behavioural rules. As an example, the now famous particle hopping (cellular automata) model of Kai Nagel and Michael Schreckenberg [Nag92b] still forms a widely-cited basis for current research papers on the subject.

In parallel with this kind of modelling approach, many of the old ‘beliefs’ (e.g., the fluid-dynamic approach to traffic flow modelling) started to get questioned. As a consequence, a plethora of models quickly found its way to the transportation community, whereby most of these models didn’t give a thought as to whether or not their associated phenomena corresponded to real-life traffic observations.

> We note here that, whatever the modelling approach may be, researchers should always compare their results to the reality of the physical world. Ignoring this basic step, reduces the research in our opinion to nothing more than a mathematical exercise !

As the international research community began to spawn its traffic flow theories, Robert Herman aspired to bring them all together in december 1959. This led to the tri-annual organisation of the *International Symposium on Transportation and Traffic Theory* (ISTTT), by some heralded as ‘the Olympics of traffic theory’ because the symposium talks about the fundamentals underlying transportation and traffic phenomena. Another example of the evolution of recent developments with respect to the parallels between traffic flows and granular media, is the bi-annual organisation of the workshop on *Traffic and Granular Flow* (TGF), a platform for exchanging ideas by bringing together researchers from various scientific fields.

Nowadays, the research and application of traffic flow theory and intelligent transportation systems continues. The scientific field has been largely diversified, encompassing a broad range of aspects related to sociology, psychology, the environment, the economy, … The global avidity of the field can be witnessed by the exponentially growing publication output. Keeping our previous comment in mind, researchers from time to time just seem to ‘add to the noise’ (mainly due to the sheer diversity of the literature body), although there occasionally exist exceptions such as the late Newell, as subtly pointed out by Michael Cassidy in [Orr02].

As a final word, we refer the reader to two personalised views on the history of traffic flow theory, namely the musings of the late Gordon Newell and Denos Gazis [New02a; Gaz02]. We furthermore invite the reader to cast a glance at the ending pages of Wardrop’s paper [War52], in which a rather colourful discussion on the introduction of mathematics to traffic flow theory has been written down.

2.2 Microscopic traffic flow characteristics

Road traffic flows are composed of drivers associated with individual vehicles, each of them having their own characteristics. These characteristics are called *microscopic* when a traffic flow is considered as being composed of such a stream of vehicles. The dynamical aspects of these traffic flows are formed by the underlying interactions between the drivers of the vehicles. This is largely determined by the behaviour of each driver, as well as the physical characteristics of the vehicles.

Because the process of participating in a traffic flow is heavily based on the behavioural aspects associated with human drivers [Gar97], it would seem important to include these human factors into the modelling equations. However, this leads to a severe increase in complexity, which is not always a desired artifact [Mae01b]. However, in the remainder of this section, we always consider a vehicle-driver combination as a single entity, taking only into account some vehicle related traffic flow characteristics.

> Note that despite our previous remarks, we do not debate the necessity of a psychological treatment of traffic flow theory. As the research into driver behaviour is gaining momentum, a lot of attention is gained by promising studies aimed towards driver and pedestrian safety, average reaction times, the influence of stress levels, aural and visual perceptions, ageing, medical conditions, fatigue, ...

2.2.1 Vehicle related variables

Considering individual vehicles, we can say that at time t, each vehicle i in a lane of a traffic stream has the following informational variables:

- a *length*, denoted by l_i,
- a *longitudinal position*, denoted by $x_i(t)$,
- a *speed*, denoted by $v_i(t) = \dfrac{dx_i(t)}{dt}$,
- and an *acceleration*, denoted by $a_i(t) = \dfrac{dv_i(t)}{dt} = \dfrac{d^2x_i(t)}{dt^2}$.

Note that the position x_i of a vehicle is typically taken to be the position of its rear bumper. In this first approach, a vehicle's other spatial characteristics (i.e., its width, height, and lane number) are neglected. And in spite of our narrow focus on the vehicle itself, the above list of variables is also complemented with a driver's *reaction time*, denoted by τ_i[1].

With respect to the acceleration characteristics, it should be noted that these are in fact not only dependent on the vehicle's engine, but also on, e.g., the road's inclination, being a non-negligible factor that plays an important role in the forming of congestion at bridges and tunnels. We do not use the derivative of the acceleration, called *jerk*, *jolt*, or *surge* (jerk is also used to represent the smoothness of the *acceleration noise* [Mon64]).

[1] Note that in most cases, a driver's reaction time is assumed to be constant (drawn from a distribution), as opposed to the more general idea that it is traffic-state dependent (e.g., people are more alert when they are following close to each other than when they are driving relaxed).

Except in the acceleration capabilities of a vehicle, we ignore the physical forces that act on a vehicle, e.g., the earth's gravitational pull, road and wind friction, centrifugal forces, ... A more elaborate explanation of these forces can be found in [Dag97b].

2.2.2 Traffic flow characteristics

Referring to Figure 2.1, we can consider two consecutive vehicles in the same lane in a traffic stream: a follower i and its leader $i + 1$. From the figure, it can be seen that vehicle i has a certain *space headway* h_{s_i} to its predecessor (it is expressed in metres), composed of the distance (called the *space gap*) g_{s_i} to this leader and its own *length* l_i:

$$h_{s_i}(t) = g_{s_i}(t) + l_i. \tag{2.1}$$

By taking, as stated before, the rear bumper as a vehicle's position, the space headway $h_{s_i}(t) = x_{i+1}(t) - x_i(t)$. The space gap is thus measured from a vehicle's front bumper to its leader's rear bumper.

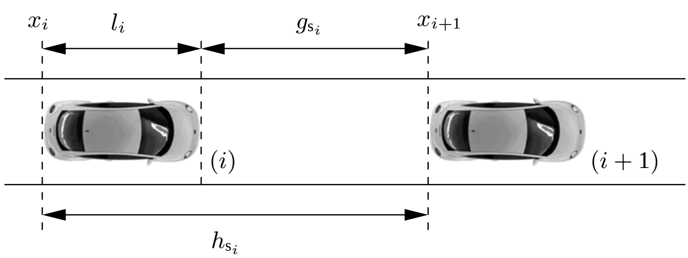


Figure 2.1: Two consecutive vehicles (a follower i at position x_i and a leader $i + 1$ at position x_{i+1}) in the same lane in a traffic stream. The follower has a certain space headway h_{s_i} to its leader, equal to the sum of the vehicle's space gap g_{s_i} and its length l_i.

Analogously to equation (2.1), each vehicle also has a *time headway* h_{t_i} (expressed in seconds), consisting of a *time gap* g_{t_i} and an *occupancy time* ρ_i:

$$h_{t_i}(t) = g_{t_i}(t) + \rho_i(t). \tag{2.2}$$

Both space and time headways can be visualised in a *time-space diagram*, such as the one in Figure 2.2. Here, we have shown the two vehicles i and $i + 1$ as they are driving. Their positions x_i and x_{i+1} can be plotted with respect to time, tracing out two *vehicle trajectories*. As the time direction is horizontal and the space direction is vertical, the vehicles' respective speeds can be derived by taking the tangents of the trajectories (for simplicity, we have assumed that both vehicles travel at the same constant speed, resulting in parallel linear trajectories). Accelerating vehicles have steep inclining trajectories, whereas those of stopped vehicles are horizontal.

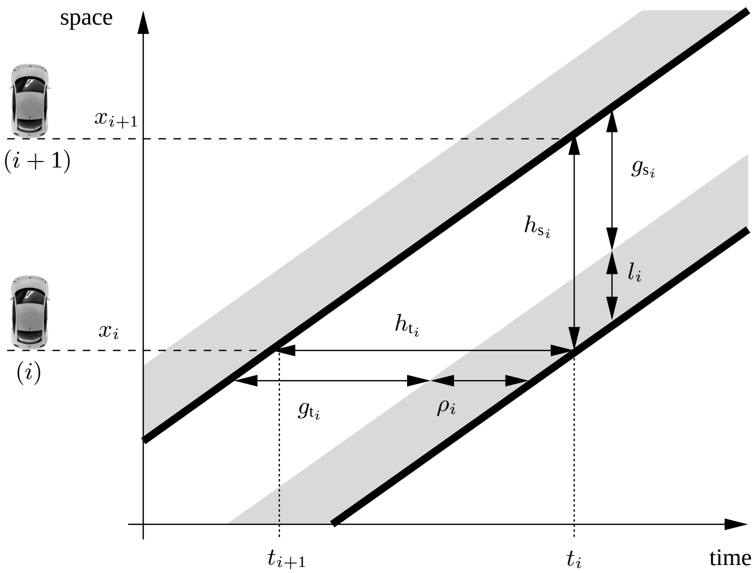


Figure 2.2: A time-space diagram showing two vehicle trajectories i and $i+1$, as well as the space and time headway h_{s_i} and h_{t_i} of vehicle i. Both headways are composed of the space gap g_{s_i} and the vehicle length l_i, and the time gap g_{t_i} and the occupancy time ρ_i, respectively. The time headway can be seen as the difference in time instants between the passing of both vehicles, respectively at t_{i+1} and t_i (diagram based on [Log03a]).

When the vehicle's speed is constant, the time gap is the amount of time necessary to reach the current position of the leader when travelling at the current speed (i.e., it is the elapsed time an observer at a fixed location would measure between the passing of two consecutive vehicles). Similarly, the occupancy time can be interpreted as the time needed to traverse a distance equal to the vehicle's own length at the current speed, i.e., $\rho_i(t) = l_i/v_i(t)$; this corresponds to the time the vehicle needs to pass the observer's location. Both equations (2.1) and (2.2) are furthermore linked to the vehicle's speed v_i as follows [Dag97b]:

$$\frac{h_{s_i}(t)}{h_{t_i}(t)} = \frac{g_{s_i}(t)}{g_{t_i}(t)} = \frac{l_i}{\rho_i(t)} = v_i(t). \tag{2.3}$$

As the above definitions deal with what is called single-lane traffic, we can easily extend them to multi-lane traffic. In this case, four extra space gaps — related to the vehicles in the neighbouring lanes — are introduced, namely $g_{s_i}^{l,f}$ at the left-front, $g_{s_i}^{l,b}$ at the left-back, $g_{s_i}^{r,f}$ at the right-front, and $g_{s_i}^{r,b}$ at the right-back. The four corresponding space headways, $h_{s_i}^{l,f}$, $h_{s_i}^{l,b}$, $h_{s_i}^{r,f}$, and $h_{s_i}^{r,b}$, are introduced in a similar fashion. The extra time gaps and headways are derived in complete analogy, leading to the four time gaps $g_{t_i}^{l,f}$, $g_{t_i}^{l,b}$, $g_{t_i}^{r,f}$, and $g_{t_i}^{r,b}$, and the four corresponding time headways $h_{t_i}^{l,f}$, $h_{t_i}^{l,b}$, $h_{t_i}^{r,f}$, and $h_{t_i}^{r,b}$.

In single-lane traffic, vehicles always keep their relative order, a principle sometimes called *first-in, first-out* (FIFO) [Dag95a]. For multi-lane traffic however, this principle is no longer obeyed due to overtaking manoeuvres, resulting in vehicle trajectories that cross each other. If the same time-space diagram were to be drawn for only one lane (in multi-lane traffic), then some vehicles' trajectories would suddenly appear or vanish at the point where a lane change occurred.

> In some traffic flow literature, other nomenclature is used: *space* for the space headway, *distance* or *clearance* for the space gap, and *headway* for the time headway. Because this terminology is confusing, we propose to use the unambiguously defined terms as described in this section.

2.3 Macroscopic traffic flow characteristics

When considering many vehicles simultaneously, the time-space diagram mentioned in Section 2.2.2 can be used to faithfully represent all traffic. In Figure 2.3 we show the evolution of the system, as we have traced the trajectories of all the individual vehicles' movements. This time-space diagram therefore provides a complete picture of all traffic operations that are taking place (accelerations, decelerations, …).

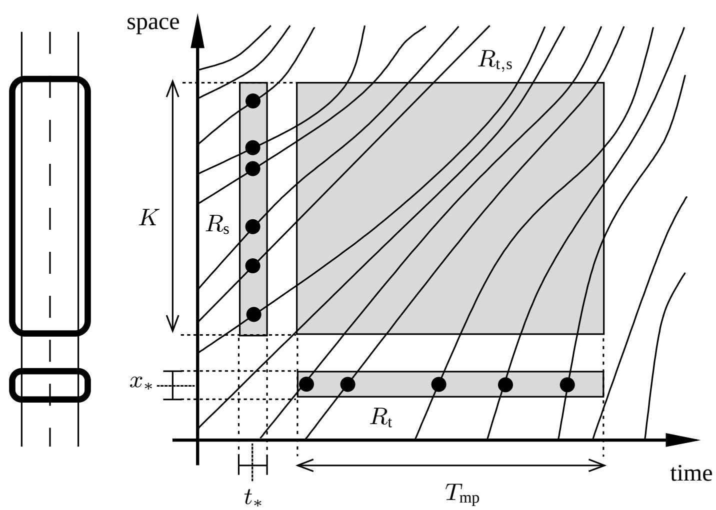


Figure 2.3: A time-space diagram showing several vehicle trajectories and three measurement regions R_t, R_s, and $R_{t,s}$. These rectangular regions are bounded in time and space by a measurement period T_{mp} and a road section of length K. The black dots represent the individual measurements.

Instead of considering each vehicle in a traffic stream individually, we now 'zoom out' to a more aggregate *macroscopic* level (e.g., traffic streams are regarded as a fluid). In the remainder of this section, we will measure some macroscopic traffic flow characteristics based on the shown time-space diagram. To this end, we define three measurement regions:

- R_{t} corresponding to measurements at a single fixed location in space (x_*), during a certain time period T_{mp}. An example of this is a single inductive loop detector (SLD) embedded in the road's concrete.
- R_{s} corresponding to measurements at a single instant in time (t_*), over a certain road section of length K. An example of this is an aerial photograph.
- $R_{\text{t,s}}$ corresponding to a general measurement region. Although it can have any shape, in this case we restrict ourselves to a rectangular region in time and space. An example of this is a continuous sequence of images made by a video camera detector.

With respect to the size of these measurement regions, some caution is advised: a too large measurement region can mask certain effects of traffic flows, possibly ignoring some of the dynamic properties, whereas a too small measurement region may obstruct a continuous treatment, as the discrete, microscopic nature of traffic flows becomes apparent.

Using these different methods of observation, we now discuss the measurement of four important macroscopic traffic flow characteristics: density, flow, occupancy, and mean speed. We furthermore give a short discussion on the moving observer method and the use of floating car data.

> With respect to some naming conventions on roadways, two different 'standards' exist for some of the encountered terminology, namely the American and the British standard. Examples are: the classical multi-lane high-speed road with on- and off-ramps, which is called a *freeway* or a *super highway* (American), or an *arterial* or *motorway* (British). A main road with intersections is called an *urban highway* (American) or a *carriageway* (British). In this dissertation, we have chosen to adopt the British standard. Finally, in contrast to Great Britain and Australia, we assume that for low-density traffic, everybody drives in the right instead of the left lane.

2.3.1 Density

The macroscopic characteristic called *density* allows us to get an idea of how crowded a certain section of a road is. It is typically expressed in number of vehicles per kilometre (or mile). Note that the concept of density totally ignores the effects of traffic composition and vehicle lengths, as it only considers the abstract quantity 'number of vehicles'.

Because density can only be *measured* in a certain spatial region (e.g., R_s in Figure 2.3), it is *computed* for temporal regions such as region R_t in Figure 2.3. When density can not be exactly measured or computed, or when density measurements are faulty, it has to be *estimated*. To this end, several available techniques exist, e.g., based on explicit simulation using a traffic flow propagation model [Muñ03b], based on a vehicle reidentification system [Coi03a], based on a complete traffic state estimator using an extended Kalman filter [Wan03], or based on a non-linear adaptive observer [AI04], ...

2.3.1.1 Mathematical formulation

Using the spatial region R_s, the density k for single-lane traffic is defined as:

$$k(t_*, x, R_\text{s}) = \frac{N}{K}, \tag{2.4}$$

with N the number of vehicles present on the road segment. If we consider multi-lane traffic, we have to sum the partial densities k_l of each of the L lanes as follows:

$$k(t_*, x, R_\text{s}) = \sum_{l=1}^{L} k_l = \frac{1}{K} \sum_{l=1}^{L} N_l, \tag{2.5}$$

in which N_l now denotes the number of vehicles present in lane l (equation (2.5) is *not* the same as averaging over the partial densities of each lane)[2].

In general, density can be defined as *the total time spent by all the vehicles in the measurement region, divided by the area of this region* [Edi65; Dag97b]. This generalisation allows us to compute the density at a point using the temporal measurement region R_t; to this end, consider the finite spatial interval ΔX surrounding x_*:

$$k(t, x_*, R_\text{t}) = \frac{\sum_{i=1}^{N} \Delta T_i}{T_\text{mp}\ \Delta X} = \frac{1}{T_\text{mp}} \sum_{i=1}^{N} \frac{\Delta T_i}{\Delta X}, \tag{2.6}$$

with ΔT_i the time taken by i^th vehicle to travel the distance ΔX. After taking the limit that reduces the interval ΔX to a single point alongside the road, we obtain the following expression for the density:

$$k(t, x_*, R_\text{t}) = \frac{1}{T_\text{mp}} \lim_{\Delta X \to 0} \sum_{i=1}^{N} \frac{\Delta T_i}{\Delta X} = \frac{1}{T_\text{mp}} \sum_{i=1}^{N} \frac{1}{v_i}, \tag{2.7}$$

[2] Note that when calculating the total density using equation (2.5), the partial densities can also correspond without loss of generality to different vehicle classes instead of just different lanes [War52; Dag97b].

with v_i the speed of the i^{th} vehicle (with $v_i \neq 0$). Extending the previous derivation to multi-lane traffic is done straightforward using equation (2.5):

$$k(t, x_*, R_\text{t}) = \frac{1}{T_\text{mp}} \sum_{l=1}^{L} \sum_{i=1}^{N_l} \frac{1}{v_{i,l}}, \tag{2.8}$$

with now $v_{i,l}$ denoting the speed of the i^{th} vehicle in lane l.

As we now can obtain the density in both spatial and temporal regions, R_s and R_t respectively, it would seem a logical extension to find the density in the region $R_{\text{t,s}}$. In order to do this, however, we need to know the travel times T_i of the individual vehicles, as can be seen in equation (2.7). Because this information is not always available, and in most cases rather difficult to measure, we use a different approach, corresponding to the temporal average of the density. Assuming that at each time step t, during a certain time period T_mp, the density $k(t)$ is known in consecutive regions R_s, the generalised definition leads to the following formulation:

$$k(t, x, R_{\text{t,s}}) = \begin{cases} \dfrac{1}{T_\text{mp}} \displaystyle\int_{t=0}^{T_\text{mp}} k(t)\, dt & \text{(continuous)}, \\[2ex] \dfrac{1}{T_\text{mp}} \displaystyle\sum_{t=1}^{T_\text{mp}} k(t) & \text{(discrete)}, \end{cases} \tag{2.9}$$

with $T_\text{mp} \in \mathbb{N}_0$ and $t \in \{1, \ldots, T_\text{mp}\}$ in the discrete case. For multi-lane traffic, combining equations (2.5) and (2.9) results in the following formula for computing the density in region $R_{\text{t,s}}$ using measurements in discrete time:

$$k(t, x, R_{\text{t,s}}) = \frac{1}{T_\text{mp}\, K} \sum_{t=1}^{T_\text{mp}} \sum_{l=1}^{L} N_l(t), \tag{2.10}$$

where $N_l(t)$ denotes the number of vehicles present in lane l at time t.

There exists a relation between the macroscopic traffic flow characteristics and those microscopic characteristics defined in Section 2.2.2. For the density k, this relation is based on the average space headway $\overline{h}_\text{s}$ [War52; Dag97b]:

$$k = \frac{N}{K} = \frac{N}{\displaystyle\sum_{i=1}^{N} h_{\text{s}_i}} = \frac{1}{\frac{1}{N}\displaystyle\sum_{i=1}^{N} h_{\text{s}_i}} = \frac{1}{\overline{h}_\text{s}}, \tag{2.11}$$

with $\overline{h}_\text{s}^{\,-1}$ the reciprocal of the average space headway.

2.3.1.2 Passenger car units

When considering heterogeneous traffic flows (i.e., traffic streams composed of different types of vehicles), operating agencies usually do not express the macroscopic traffic flow characteristics using the raw number of vehicles, but rather employ the notion of *passenger car units* (PCU). These PCUs, sometimes also called *passenger car equivalents* (PCE), try to take into account the spatial differences between vehicle types. For example, by denoting one average passenger car as 1 PCU, a truck in the same traffic stream can be considered as 2 PCUs (or even higher and fractional values for trailer trucks).

> Let us finally note that, because density is essentially defined as a spatial measurement, it is one of the most difficult quantities to obtain. It is interesting to notice that at this moment, it is theoretically possible for video cameras to measure density over a short spatial region. However, to our knowledge there currently exists no commercial implementation.

2.3.2 Flow

Whereas density typically is a spatial measurement, *flow* can be considered as a temporal measurement (i.e., region R_{t}). Flow, which we use as a shorthand for rate of flow, is typically expressed as an hourly rate, i.e., in number of vehicles per hour. Note that sometimes other synonyms such as *intensity*, *flux*, *throughput*, *current*, or *volume*[3] are used, typically depending on a person's scientific background (e.g., engineering, physics, ...).

2.3.2.1 Mathematical formulation

Measuring the flow q in region R_{t} for single-lane traffic, is done using the following equation, which is based on raw vehicle counts:

$$q(t, x_*, R_{\text{t}}) = \frac{N}{T_{\text{mp}}}, \tag{2.12}$$

with N the number of vehicles that has passed the detector's site. For multi-lane traffic, we sum the partial flows of each of the L lanes:

$$q(t, x_*, R_{\text{t}}) = \sum_{l=1}^{L} q_l = \frac{1}{T_{\text{mp}}} \sum_{l=1}^{L} N_l, \tag{2.13}$$

[3] In most cases, volume denotes the number of vehicles counted during a certain time period, as opposed to flow which is just the equivalent hourly rate.

with now N_l denoting the number of vehicles that passed the detector's site in lane l. Note that we assume that each lane has its own detector, otherwise we would be dealing with an average flow across all the lanes.

Generally speaking, flow can defined as *the total distance travelled by all the vehicles in the measurement region, divided by the area of this region* [Edi65; Dag97b]. In analogy with equation (2.7), this generalisation allows us to compute the flow at an instant in time using the spatial measurement region R_s; to this end, consider the finite temporal interval ΔT surrounding t_*:

$$q(t_*, x, R_\mathrm{s}) = \frac{\sum_{i=1}^{N} \Delta X_i}{K\,\Delta T} = \frac{1}{K} \sum_{i=1}^{N} \frac{\Delta X_i}{\Delta T}, \tag{2.14}$$

with ΔX_i the distance travelled by the i^th vehicle during the time interval ΔT. After taking the limit that reduces the interval ΔT to a single instant in time, we obtain the following expression for the flow:

$$q(t_*, x, R_\mathrm{s}) = \frac{1}{K} \lim_{\Delta T \to 0} \sum_{i=1}^{N} \frac{\Delta X_i}{\Delta T} = \frac{1}{K} \sum_{i=1}^{N} v_i, \tag{2.15}$$

with v_i the speed of the i^th vehicle. The extension to multi-lane traffic is straightforward:

$$q(t_*, x, R_\mathrm{s}) = \frac{1}{K} \sum_{l=1}^{L} \sum_{i=1}^{N_l} v_{i,l}. \tag{2.16}$$

Considering consecutive flow measurements in region $R_\mathrm{t,s}$, we can derive a formulation corresponding to the temporal average of the flow, similar to that of equation (2.9). Assuming that at each time step t, during a certain time period T_mp, the flow $q(t)$ is known in consecutive regions R_s, the generalised definition leads to the following equations:

$$q(t, x, R_\mathrm{t,s}) = \begin{cases} \dfrac{1}{T_\mathrm{mp}} \displaystyle\int_{t=0}^{T_\mathrm{mp}} q(t)\, dt & \text{(continuous)}, \\[2ex] \dfrac{1}{T_\mathrm{mp}} \displaystyle\sum_{t=1}^{T_\mathrm{mp}} q(t) & \text{(discrete)}, \end{cases} \tag{2.17}$$

with $T_\mathrm{mp} \in \mathbb{N}_0$ and $t \in \{1, \ldots, T_\mathrm{mp}\}$ in the discrete case. For multi-lane traffic, combining equations (2.16) and (2.17) results in the following formula for computing the flow in region $R_{\mathrm{t,s}}$ using measurements in discrete time:

$$q(t, x, R_{\mathrm{t,s}}) = \frac{1}{T_{\mathrm{mp}}\,K} \sum_{t=1}^{T_{\mathrm{mp}}} \sum_{l=1}^{L} \sum_{i=1}^{N_l(t)} v_{i,l}(t), \tag{2.18}$$

where $v_{i,l}(t)$ denotes the speed of the i^{th} vehicle in lane l at time t.

In analogy with equation (2.11), there exists a relation between the flow q, and the average time headway $\overline{h}_{\mathrm{t}}$ [War52; Dag97b]:

$$q = \frac{N}{T_{\mathrm{mp}}} = \frac{N}{\sum_{i=1}^{N} h_{\mathrm{t}_i}} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} h_{\mathrm{t}_i}} = \frac{1}{\overline{h}_{\mathrm{t}}}, \tag{2.19}$$

with $\overline{h}_{\mathrm{t}}^{\,-1}$ the reciprocal of the average time headway.

2.3.2.2 Oblique cumulative plots

As stated before, flows are always expressed as a rate. In contrast to this, we can also consider the raw vehicle counts at a certain location (i.e., measurement region R_{t}). If we plot the cumulative number of passing vehicles (denoted by N) with respect to time for different regions (e.g., inductive loop detectors), we get a set of curves such as the one in the left part of Figure 2.4. These curves are called *cumulative plots* (or (t, N) diagrams), and although their origins date back as far as 1954 with the work of Karl Moskowitz [Mos54], it was Gordon Newell who applied them later on to their full potential (initially in the context of queueing theory) [New82; New93a; New93b; New93c] (a similar method was applied by John Luke, in the field of continuum mechanics [Luk72; Dag95b]). Interestingly, a short while later Nam and Drew independently proposed a method to estimate the travel time based on cumulative counts at neighbouring detector stations [Nam96].

The key benefit of these cumulative plots, comes when comparing observations stemming from multiple detector stations at a closed section of the road that conserves the number of vehicles (i.e., no on- or off-ramps), in which case we also speak of *input-output diagrams*. If there are two detector stations, then the upstream and downstream stations measure the *input*, respectively *output*, of the section. Similarly like in queueing theory, the upstream curve is sometimes called the *arrival function*, whereas the downstream one is called the *departure function* [New82]. As the method is based on counting the number of individual vehicles at each observation location (whereby each vehicle is numbered with respect to a single reference vehicle), this results in a monotonically increasing function $N(t, x)$ (sometimes called the *Moskowitz function*,

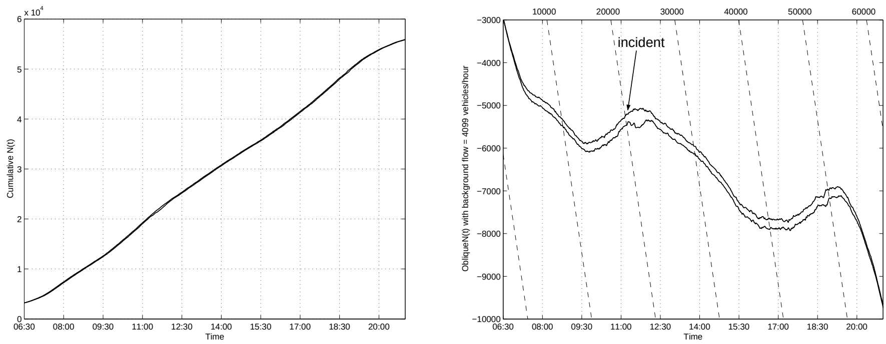


Figure 2.4: *Left:* a standard cumulative plot showing the number of passing vehicles at two detector locations; due to the graph's scale, both curves appear to lie on top of each other. *Right:* the same data but displayed using an oblique coordinate system, thereby enhancing the visibility (the dashed slanted lines have a slope corresponding to the subtracted background flow $q_b \approx 4100$ vehicles per hour). We can see a queue (probably caused due to an incident) growing at approximately 11:00, dissipating some time later at approximately 12:30. The shown detector data was taken from single inductive loop detectors [VVC03], covering all three lanes of the E40 motorway between Erpe-Mere and Wetteren, Belgium. The shown data was recorded at Monday, April 4, 2003 (the detectors' sampling interval was one minute, the distance between the upstream and downstream detector stations was 8.1 kilometres).

after its 'inventor'), which increases each time a vehicles passes by. At each time instant t, the cumulative count is defined as:

$$N(t,x) = \sum_{t'=t_0}^{t} q(t',x) = N(t-1,x) + q(t,x). \tag{2.20}$$

The time needed to travel from one location to another can easily be measured as the horizontal distance between the respective cumulative curves. Similarly, the vertical distance between these curves allows us to derive the accumulation of vehicles on the road section, which gives an excellent indication of growing and dissipating queues (i.e., congestion). Furthermore, if we compute the slope of this function at each time instant t, we obtain the flow $q(t,x) = [N(t+\Delta t,x) - N(t,x)]/\Delta t$. Finally, because $N(t,x)$ essentially is a step function, we can define a smooth approximation which is an everywhere differentiable function $\widetilde{N}(t,x)$. As $\Delta t \to 0$, this allows us to compute instantaneous flows as $q = \partial \widetilde{N}(t,x)/\partial t$. The same principle holds true for local densities, where now $k(t,x) = [N(t,x+\Delta x) - N(t,x)]/\Delta x$. As $\Delta x \to 0$, this results in $k = -\partial \widetilde{N}(t,x)/\partial x$ [Dag97b].

The main disadvantage of the method is the fact that these cumulative functions increase very rapidly, thereby masking the subtle differences between different curves. Cassidy and Windover therefore proposed to subtract a background flow q_b from these curves, resulting in functions $N(t,x) - t\ q_b$ [Cas95]. Based on this; Muñoz and

Daganzo furthermore introduced enhanced clarity by overlaying this cumulative plot with a set of oblique lines with slope $-q_{\mathrm{b}}$ [Muñ02b]. Choosing an appropriate value for q_{b}, allows us to nicely enhance the characteristic undulations that are expressed by the different oblique curves.

> Note that before using these oblique plots, the cumulative plots from different detectors stations need to be *synchronised*. To understand this, suppose a reference vehicle passes an upstream detector station at a certain time instant t_{up}; after a certain time period, the vehicle reaches the downstream detector station at a later time instant t_{down}. The amount $t_{\mathrm{down}} - t_{\mathrm{up}}$ is the time it takes to cross the distance between both detector stations, allowing the synchronisation mechanism to shift the respective cumulative curves over this time period (i.e., initialising them with the passing of the reference vehicle).
>
> One way to achieve this, is by looking at the respective shapes of both cumulative curves during light traffic conditions (e.g., the early morning period when free-flow conditions are prevailing). The idea now is to shift one curve such that the difference between the two curves' shapes is minimal [Wes95; Muñ00b; Muñ03a]. Note that other corrections may be necessary, as both detector stations can count a different number of vehicles (i.e., a systematic bias).

An example of an oblique plot can be seen in the right part of Figure 2.4: the cumulative count at each time instant can be read from an axis that is perpendicular to the oblique (slanted) overlayed dashed lines (e.g., we can see a count of some 30000 vehicles at 14:00). Note that the accumulation can still be measured by the vertical distance between two curves (i.e., at a specific time instant), but the travel time should now be measured along one of the overlayed oblique lines. Such a pair of cumulative curves can be thought of as a flexible plastic garden hose: whenever there is an obstruction on the road, the outflow of the section will be blocked, resulting in a local thickening of this 'hose' (i.e., the accumulation of vehicles on the section).

Using these oblique cumulative plots, we can now inspect the traffic dynamics in much more detail than was previously possible. For example, looking again at the right part of Figure 2.4, we can see how the specific traffic stream characteristics propagate from one detector station to another. Even more visible, is a queue that starts to grow at approximately 11:00 (i.e., the time of the appearance of a 'bulge'), dissipating at approximately 12:30. As data curves from upstream detectors lie above data curves from downstream detectors, we see a decrease in the road section's output. Careful investigation of the traffic data revealed that the detector stations recorded a rather low flow (approximately 2500 vehicles per hour as opposed to a nominal flow of 4500 vehicles per hour), whereby all vehicles drove at a low speed (between 20 and 60 km/h as opposed to 110 km/h). This gives sufficient evidence to conclude that an incident probably occurred shortly after 11:00, consequently obstructing a part of the road and leading to a build up of vehicles in the section.

Let us finally note that although oblique cumulative plots currently are not a mainstream technique used by the traffic community, we predict their rising popularity:

they are one of the most simple, yet powerful, techniques for studying local traffic phenomena, giving traffic engineers practical insight into the formation of bottlenecks. Some recent examples include the work of Muñoz and Daganzo [Muñ00a; Muñ00b; Muñ02a; Muñ03a], Cassidy and Bertini [Cas99; Ber03], Cassidy and Mauch [Cas01], Windover and Cassidy [Win01], Logghe [Log03a], Bertini et al. [Ber05], and Lindgren [Lin05].

2.3.3 Occupancy

Notwithstanding the importance of measuring traffic density, most of the existing detector stations on the road are only capable of temporal measurements (i.e., region R_{t}). If individual vehicle speeds can be measured, by double inductive loop detectors (DLD) for example, then density should be computed using equation (2.7).

However, in many cases these vehicle speeds are not readily available, e.g., when using single inductive loop detectors. The detector's logic therefore resorts to a temporal measurement called the *occupancy* ρ, which corresponds to the fraction of time the measurement location was occupied by a vehicle:

$$\rho(t, x_*, R_{\text{t}}) = \frac{1}{T_{\text{mp}}} \sum_{i=1}^{N} o_{\text{t}_i}. \tag{2.21}$$

In the previous equation, o_{t_i} denotes the i^{th} vehicle's *on-time*, i.e., the time period during which it is present above the detector (it corresponds to the shaded area swept by a vehicle at a certain location x_i in Figure 2.2). Note that this on-time actually corresponds to the effective vehicle length as seen by the detector, divided by the vehicle's speed [Coi01]:

$$o_{\text{t}_i} = \frac{l_i + K_{\text{ld}}}{v_i}, \tag{2.22}$$

with l_i the vehicle's true length and $K_{\text{ld}} > dx$ the finite, non-infinitesimal length of the detection zone. If we define $\overline{o}_{\text{t}}$ as the average on-time (based on the vehicles that have passed the detector during the observation period), then we can establish a relation between the occupancy and the flow [Dag97b] using equations (2.12) and (2.21):

$$\rho = \left(\frac{N}{T_{\text{mp}}}\right) \left(\frac{1}{N} \sum_{i=1}^{N} o_{\text{t}_i}\right) = q\,\overline{o}_{\text{t}}. \tag{2.23}$$

Furthermore, it is as before possible to define the occupancy for generalised measurement regions, using the total space consumed by the shaded areas of vehicles in a time-space diagram (e.g., Figure 2.2), divided by the area of the measurement region [Edi65; Dag97b; Cas98]. Continuing our discussion, we assume that individual vehicle lengths and speeds are uncorrelated; it can then be shown that [Dag97b; Coi01]:

$$\rho = \bar{l}\, k \Longrightarrow k = \frac{\rho}{\bar{l}}, \tag{2.24}$$

where we dropped the functional dependencies for visual clarity; in the above relation, $\bar{l}$ is the average vehicle length (note that this can correspond to the concept of passenger car units defined in Section 2.3.1). Multiplying equation (2.24) by 100, allows us to express the occupancy as a percentage. For multi-lane traffic, the occupancy is derived in analogy to equation (2.8):

$$\rho(t, x_*, R_\text{t}) = \sum_{i=1}^{L} \rho_l = \frac{1}{T_\text{mp}} \sum_{l=1}^{L} \sum_{i=1}^{N_l} o_{\text{t}_{i,l}}, \tag{2.25}$$

with now $o_{\text{t}_{i,l}}$ the on-time of the i^th vehicle in lane l. Note that the total occupancy derived in this way, can exceed 1 (but is bounded by L); if desired, it can be normalised through a division by L to obtain the *average occupancy*.

Note that if we apply equation (2.24) to measurement region R_s based on the density in equation (2.4), then the occupancy ρ can be written as:

$$\rho(t_*, x, R_\text{s}) = \left(\frac{1}{\cancel{N}} \sum_{i=1}^{N} l_i \right) \frac{\cancel{N}}{K} = \frac{1}{K} \sum_{i=1}^{N} l_i. \tag{2.26}$$

So the occupancy now represents the ‘real density’ of the road, i.e., the physical space that all vehicles occupy.

> In the past, density was sometimes referred to as *concentration*. Nowadays however, concentration is used in a more broad context, encompassing both density and occupancy whereby the former is meant to be a spatial measurement, as opposed to the latter which is considered to be a temporal measurement [Gar97].

2.3.4 Mean speed

The final macroscopic characteristic to be considered, is the *mean speed* of a traffic stream; it is expressed in kilometres (or miles) per hour (the inverse of a vehicle’s speed is called its *pace*). Note that speed is not to be confused with *velocity*; the latter is actually a vector, implying a direction, whereas the former could be regarded as the norm of this vector.

2.3.4.1 Mathematical formulation

If we base our approach on direct measurements of the individual vehicles' speeds, we can generally obtain the mean speed as *the total distance travelled by all the vehicles in the measurement region, divided by the total time spent in this region* [Edi65; Dag97b]. After taking the limit similar to equations (2.7) and (2.15), this process gives the following derivations for the spatial and temporal regions, R_{s} and R_{t} respectively (note that in the remainder of this section we have dropped the dependencies on time and space for the sake visual clarity):

$$\overline{v}_{\text{s}} = \frac{\sum_{i=1}^{N} \Delta X_i}{\sum_{i=1}^{N} \Delta T_i} = \begin{cases} \dfrac{\sum_{i=1}^{N} \Delta X_i}{N \Delta T} \Rightarrow \overline{v}_{\text{s}} = \dfrac{1}{N} \lim_{\Delta T \to 0} \sum_{i=1}^{N} \dfrac{\Delta X_i}{\Delta T} = \dfrac{1}{N} \sum_{i=1}^{N} v_i & \text{(for } R_{\text{s}}\text{)}, \\[2ex] \dfrac{N \Delta X}{\sum_{i=1}^{N} \Delta T_i} \Rightarrow \overline{v}_{\text{s}} = \dfrac{1}{\dfrac{1}{N} \lim_{\Delta X \to 0} \sum_{i=1}^{N} \dfrac{\Delta T_i}{\Delta X}} = \dfrac{1}{\dfrac{1}{N} \sum_{i=1}^{N} \dfrac{1}{v_i}} & \text{(for } R_{\text{t}}\text{)}, \end{cases} \tag{2.27}$$

with now ΔX_i and ΔT_i the distance and time travelled by the i^{th} vehicle during the time period ΔT and over the distance ΔX, respectively. N is the number of vehicles present during the measurement. The mean speed computed by the previous equations, is called the *average travel speed* (the computation also includes stopped vehicles), which is more commonly known as the *space-mean speed* (SMS); we denote it with $\overline{v}_{\text{s}}$ (note that in some engineering disciplines, the sole letter u is used to denote a mean speed, however, this is ambiguous in our opinion).

It is interesting to see that the spatial measurement is based on an *arithmetic average* of the vehicles' *instantaneous speeds*, whereas the temporal measurement is based on the *harmonic average* of the vehicles' *spot speeds*. If we instead were to take the arithmetic average of the vehicles' spot speeds in the temporal measurement region R_{t}, this would lead to what is called the *time-mean speed* (TMS); we denote it by $\overline{v}_{\text{t}}$:

$$\overline{v}_{\text{t}} = \frac{1}{N} \sum_{i=1}^{N} v_i \quad \text{(region } R_{\text{t}}\text{)}. \tag{2.28}$$

Similarly, we can compute the time-mean speed for measurement region R_{s}, by taking the harmonic average of the vehicles' instantaneous speeds. With respect to both space- and time-mean speeds, Wardrop has shown that the following relation holds [War52]:

$$\overline{v}_{\text{t}} = \overline{v}_{\text{s}} + \frac{\sigma_{\text{s}}^2}{\overline{v}_{\text{s}}}, \tag{2.29}$$

with σ_s^2 the statistical sample variance defined as follows:

$$\sigma_s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(v_i - \overline{v}_s)^2, \tag{2.30}$$

in which v_i denotes the i^{th} vehicle's instantaneous speed. One of the main consequences of equation (2.29), is that the time-mean speed always exceeds the space-mean speed (except when all the vehicles' speeds are the same, in which case the sample variance is zero and, as a consequence, the time- and space-mean speeds are equal). So a stationary observer will most likely see more faster than slower vehicles passing by, as opposed to, e.g., an aerial photograph in which more slower than faster vehicles will be seen [Dag97b]. Despite this mathematical quirk, the practical difference between SMS and TMS is often negligible for free-flow traffic (i.e., light traffic conditions); however, under congested traffic conditions both mean speeds will behave substantially differently (i.e., around 10%).

Using equation (2.29), we can also estimate the space-mean speed, based on the time-mean speed and approximating the variance of the SMS with that of the TMS [Bov00; Rak05]:

$$\begin{aligned}
\overline{v}_s &= \overline{v}_t - \frac{\sigma_s^2}{\overline{v}_s}, \\
&\approx \overline{v}_t - \frac{\sigma_t^2}{\overline{v}_s}, \\
&\Downarrow \\
\overline{v}_s - \overline{v}_t &\approx -\frac{\sigma_t^2}{\overline{v}_s}, \\
\overline{v}_s^2 - \overline{v}_s\overline{v}_t &\approx -\sigma_t^2, \\
\overline{v}_s^2 - 2\,\overline{v}_s\frac{\overline{v}_t}{2} + \frac{\overline{v}_t^2}{4} &\approx \frac{\overline{v}_t^2}{4} - \sigma_t^2, \\
\left(\overline{v}_s - \frac{\overline{v}_t}{2}\right)^2 &\approx \frac{\overline{v}_t^2}{4} - \sigma_t^2, \\
&\Downarrow \\
\overline{v}_s &\approx \frac{\overline{v}_t}{2} + \sqrt{\frac{\overline{v}_t^2}{4} - \sigma_t^2} \qquad \forall\, \overline{v}_t \geq 2\,\sigma_t.
\end{aligned} \tag{2.31}$$

In general, using the space-mean speed is preferred to the time-mean speed. However, in most cases only this latter traffic flow characteristic is available, so care should be taken when interpreting the results of a study (unless of course when SMS and TMS are negligibly different).

The extension of equation (2.27) to multi-lane is straightforward; for example, the space-mean speed is computed as follows:

$$\overline{v}_{\mathrm{s}} = \begin{cases} \displaystyle\sum_{l=1}^{L}\sum_{i=1}^{N_l} v_{i,l} \Bigg/ \sum_{l=1}^{L} N_l & \text{(region } R_{\mathrm{s}}\text{),} \\ \dfrac{1}{\dfrac{1}{\sum_{l=1}^{L} N_l} \displaystyle\sum_{l=1}^{L}\sum_{i=1}^{N_l} \dfrac{1}{v_{i,l}}} & \text{(region } R_{\mathrm{t}}\text{),} \end{cases} \tag{2.32}$$

with now $v_{i,l}$ the instantaneous (or spot) speed of the i^{th} vehicle in lane l.

2.3.4.2 Fundamental relation of traffic flow theory

There exists a unique relation between three of the previously discussed macroscopic traffic flow characteristics density k, flow q, and space-mean speed $\overline{v}_{\mathrm{s}}$ [War52]:

$$q = k\,\overline{v}_{\mathrm{s}}. \tag{2.33}$$

This relation is also called the *fundamental relation of traffic flow theory*, as it provides a close bond between the three quantities: knowing two of them allows us to calculate the third one (note that the time-mean speed in equation (2.28) does not obey this relation). In general however, there are two restrictions, i.e., the relation is only valid for (1) continuous variables[4], or smooth approximations of them, and (2) traffic composed of substreams (e.g., slow and fast vehicles) which comply to the following two assumptions:

Homogeneous traffic

There is a homogeneous composition of the traffic substream (i.e., the same type of vehicles).

Stationary traffic

When observing the traffic substream at different times and locations, it 'looks the same'. Putting it a bit more quantitatively, all the vehicles' trajectories should be parallel and equidistant [Dag97b]. A stationary time period can be seen in a cumulative plot (e.g., Figure 2.4) where the curve corresponds to a linear function.

[4] Note that the hypothesis also assumes that the variables are spatially measured, e.g., space-mean speed.

The latter of the above two conditions[5], is also referred to as traffic operating in a *steady state* or at *equilibrium*. Based on equations (2.5) and (2.13) using partial densities and flows for different substreams (e.g., vehicle classes with distinct travel speeds, macroscopic characteristics of different lanes, ...), we can now calculate the space-mean speed, using relation (2.33), in the following equivalent ways:

$$\begin{aligned}
\overline{v}_{\mathrm{s}} &= q \,/\, k, \\
&= \sum_{c=1}^{C} q_c \Bigg/ \sum_{c=1}^{C} k_c, && (2.34) \\
&= \sum_{c=1}^{C} q_c \Bigg/ \sum_{c=1}^{C} \frac{q_c}{\overline{v}_{\mathrm{s}_c}}, && (2.35) \\
&= \sum_{c=1}^{C} k_c \, \overline{v}_{\mathrm{s}_c} \Bigg/ \sum_{c=1}^{C} k_c, && (2.36)
\end{aligned}$$

in which C denotes the number of substreams, q_c, k_c, and $\overline{v}_{\mathrm{s}_c}$ the flow, density, and space-mean speed, respectively, of the c^{th} substream. In the above derivations, equation (2.34) should be used when both the flows and densities are known, equation (2.35) should be used when both the flows and space-mean speeds are known, and equation (2.36) should be used when both the densities and space-mean speeds are known.

As can be seen in equation (2.36), the space-mean speed is calculated by averaging the substreams' space-mean speeds using their densities as weighting factors. Similarly, the time-mean speed can be derived by using the flows as weighting factors for the substreams' time-mean speeds $\overline{v}_{\mathrm{t}_c}$:

$$\overline{v}_{\mathrm{t}} = \sum_{c=1}^{C} q_c \, \overline{v}_{\mathrm{t}_c} \Bigg/ \sum_{c=1}^{C} q_c, \tag{2.37}$$

Because density can not always be easily measured, we can compute it using the fundamental relation (2.33). Density can then be directly derived from flow and space-mean speed measurements, or if the latter are not available, they can be estimated from occupancy measurements; in [Coi01; Coi03b; Coi03a], Coifman et al. provide a nice set of techniques for dealing with these difficulties.

2.3.5 Moving observer method and floating car data

When measuring and/or computing the macroscopic traffic flow characteristics in the previous sections, we always assumed a fixed measurement region. There exists how-

[5]A variable z is said to be homogeneous when $z(t,x) = z(t)$, i.e., independent of space. Similarly, it is said to be stationary when $z(t,x) = z(x)$, i.e., independent of time.

ever yet another method, based on what is called a *moving observer* [War54]. The idea behind the technique is to have a vehicle drive in both directions of a traffic flow, each time recording the number of oncoming vehicles and the net number of vehicles it gets overtaken by, as well as the times necessary to complete the two trips. Note that the assumption of stationary traffic still has to hold, i.e., the round trip should be completed before traffic conditions change significantly.

Using this method, it is then possible to derive the flow and density of the traffic stream in the direction of interest [Gar97; Dag97b]. However, the main disadvantage of this method is that, in order to obtain an acceptable level of accuracy on a road with a low flow, a very large number of trips are required [War54; Gar97; Mul02].

One of the techniques that has entered the picture during the last decade, is the use of so-called *floating cars* or *probe vehicles*. They can be compared to the moving observer method, but in this case, the vehicles are equipped with GPS and GSM(C)/GPRS devices that determine their locations based on the USA's NAVSTAR-GPS (or Europe's planned GNSS Galileo), and transmit this information to some operator. Initially, this allows an agency, e.g., a parcel delivery service or a transportation firm, to track its vehicles throughout a network, based on their locations. Nowadays, the technique has evolved, resulting in several completed field tests of which the main goal was to estimate the traffic conditions based on a small number of probe vehicles. During field measurements, floating cars can mimic several types of behaviour, most notably by travelling at the traffic flows' mean speed[6], or by trying to travel at the road's speed limit, or even by chasing another randomly selected vehicle from the traffic stream.

Another technique is based on the information obtained by mobile devices such as GSMs. The idea is that, when a driver with a cell phone has a conversation, his location is continuously monitored by so-called base transceiver stations (BTS); these latter are in fact modelled as a grid of hexagonal cells that are each centred around an antenna post. As the cell phone moves from one cell to another, a handover is executed. Based on two consecutive handovers, it is possible to determine the travel time between these two zones. In a subsequent step, this travel time is accurately matched onto a map containing the underlying road network. The upshot of this is that we can now obtain cost-effective real-time traffic information on roads with a low sensor coverage (e.g., in the absence of inductive loop detectors, cameras, ...). Note that a crucial aspect here, is that due to the many delicate privacy concerns involved with tracking individual people's units, the mobile information should remain anonymously. This implies that all personal information (i.e., identification of the caller et cetera) gets stripped from the data.

Some examples of studies and experiments with *floating car data*[7] (FCD) are given in the following. Firstly, Fastenrath gives an overview of a telematic field trial (*VEhicle Relayed Dynamic Information*, VERDI) that addresses issues such as economical, political, and technical constraints [Fas]. Secondly, Westerman provides an overview

[6] This can be accomplished when the probe vehicle overtakes as many vehicles as are overtaking the vehicle itself.

[7] Sometimes also described as *floating vehicle data* (FVD).

of available techniques for obtaining real-time road traffic information, with the goal of controlling the traffic flows through telematics [Wes95], and Wermuth et al. describe a ‘TeleTravel System’ used for surveying individual travel behaviour [Wer00]. Then, Taale et al. compare travel times from floating car data with measured travel times (using a fleet of sixty equipped vehicles driving around in Rotterdam, The Netherlands), concluding that they correspond reasonably well [Taa00]. Next, Michler derives the minimum percentage of vehicles necessary, in order to estimate traffic stream characteristics for certain traffic patterns (e.g., free-flow and congested traffic) based on rigid statistical grounds [Mic01], and Linauer and Leihs measure the travel time between points in a road network, based on a high number of users that submit a low number of GSM hand-over messages [Lin03]. In addition, Demir et al. accurately reconstruct link travel times during periods of traffic congestion, using only a very limited number of FCD-messages with a small number of users [Dem03]. Furthermore, Fontaine advanced the field of wireless location technology-based (WLT) by enhancing map-matching algorithms and developing guidelines for system design and application of WLT-based systems [Fon05].

In conclusion, we can state that the use of probe vehicles provides an effective way to gather accurate current travel times in a road network, thereby allowing good up-to-date estimations of traffic conditions. The technique will continue to grow and evolve, already by introducing personalised traffic information to drivers, based on their location and the surrounding traffic conditions. This development is furthermore stimulated by the fact that GSM market penetration still rises above 70% [Lin03], and it is our belief this will also be the case for personal GPS devices and in-vehicle route planners. Already, many private companies are entering the market, with the goal of providing real-time traffic information and route guidance in close cooperation with the mobile operators. Recent examples include the company LogicaCMG, which developed the *Mobile Traffic Services* (MTS), converting anonymous GSM-handover data into travel times by means of a map-matching algorithm. A validation study was performed for the cities of Breda and Tilburg in The Netherlands [Log05b; Log05c]. Another similar company is ITIS, which even patented its *Cellular Floating Vehicle Data* (CFVD) technology [Sim02].

2.4 Performance indicators

After considering the previously mentioned macroscopic traffic flow characteristics, we now take a look at some of the popular performance indicators used by traffic engineers when assessing the quality of traffic operations. Such performance indicators allow to compare the current progress with respect to some predefined goals, e.g., when assessing the impact of operational strategies. In general, they can be classified as follows:

- quantitative indicators (e.g., volume counts, vehicle miles travelled (VMT), person miles travelled (PMT), total time spent in the system, . . .),

- qualitative indicators (e.g., travel times and the associated delays as lost vehicle hours, average travel speed, level of service (LOS), volume-over-capacity, ...),
- and other types of indicators (e.g., safety indicators, reliability, fuel consumption, emissions and immissions[8] of air pollutants and noise, socio-economic costs, ...).

In the remainder of this section, we concisely discuss (i) the peak hour factor, (ii) the level of service, (iii) the reliability of travel times, and (iv) a measure of efficiency of a road. Of these four indicators, the first and second ones are still in common use today, whereas the family embodied by the third one is steadily gaining appreciation and the fourth one is of a qualitative nature but based on quantitative indicators. For a more complete overview, we refer the reader to the synthesis report made by Shaw [Sha03].

2.4.1 Peak hour factor

During high flow periods in the peak hour, a possible indicator for traffic flow fluctuations is the so-called *peak hour factor* (PHF). It is calculated for one day as the average flow during the hour with the maximum flow, divided by the peak flow rate during one quarter hour within this hour [May90]:

$$\text{PHF} = \frac{\overline{q}_{|60}}{\overline{q}_{|15}}. \tag{2.38}$$

For example, suppose we measure flows on a main unidirectional road with three lanes, during a morning peak: from 07:00 to 08:00 we measure consecutively 3500, 6600, 6200, and 4500 vehicles/hour during each quarter. The total average flow $\overline{q}_{|60}$ is 5200 vehicles/hour, with a peak 15 minute flow rate $\overline{q}_{|15} = 6600$ vehicles/hour. The PHF is therefore equal to $5200/6600 = 0.\overline{78}$.

Note that some manuals express the peak 15 minute flow rate as the number of vehicles during that quarter hour, necessitating an extra multiplication by 4 in the denominator of equation (2.38) to convert the flow rate to an hourly rate.

We can immediately see that the PHF is constrained to the interval [0.25,1.00]; the higher the PHF, the flatter the peak period (i.e., a longer sustained state of high flow). Typically, the PHF has values around 0.7 – 0.98. Note that two of the obvious problems with the PHF are, on the one hand, the question of when to pick the correct 15 minute interval, and on the other hand the fact that some peak periods may last longer than one hour.

[8]The difference between emissions and immissions is that the former are defined and measured locally, while the latter include dispersal effects within the environment.

2.4.2 Level of service

Historically, one of the main performance indicators to assess the quality of traffic operations, was the *level of service* (LOS), introduced in the 1960s. It is represented as a grading system using one of six letters (A – F), whereby LOS A denotes the best operating conditions and LOS F the worst. These LOS measures are based on road characteristics such as speed, travel time, ..., and drivers' perceptions of comfort, convenience, ... [HCM00]. As is customary among traffic engineers, these representative statistics of these characteristics are collectively called *measures of effectiveness* (MOE).

Levels A through D are representative for free-flow conditions whereby LOS A corresponds to free flow, LOS B to reasonable free flow, LOS C to stable traffic operations, and LOS D to bordering unstable traffic operations. LOS E is reminiscent of near-capacity flow conditions that are extremely unstable, whereas LOS F corresponds to congested flow conditions (caused by either structural or incidental congestion) [May90].

As an example, we provide an overview of the different levels of service in Table 2.1 (based on [May90], in similar form originally published in the *Highway Capacity Manual* (HCM) of 1985 as the *Transportation Research Board's* (TRB)[9] special report #209.

LOS	*Density (veh/km)*	*Occupancy (%)*	*Speed (km/h)*
A	$0 \rightarrow 7$	$0 \rightarrow 5$	≥ 97
B	$7 \rightarrow 12$	$5 \rightarrow 8$	≥ 92
C	$12 \rightarrow 19$	$8 \rightarrow 12$	≥ 87
D	$19 \rightarrow 26$	$12 \rightarrow 17$	≥ 74
E	$26 \rightarrow 42$	$17 \rightarrow 28$	≥ 48
F	$42 \rightarrow 62$ > 62	$28 \rightarrow 42$ > 42	< 48

Table 2.1: Level of service (LOS) indicators for a motorway (adapted from [May90], in similar form originally published in the 1985 HCM).

Calculating levels of service can be done using a multitude of methods; some examples include using the density (at motorways), using the space-mean speed (at arterial streets), using the delay (at signallised and unsignallised intersections), ... [HCM00]. The distinction between different LOS is primarily based on the measured average speed, and secondly on the density (or occupancy). Furthermore, as traditional analyses only focus on a select number of hours, a new trend is to conduct *whole year analyses* (WYA) based on aggregated measurements such as, e.g., the *monthly average daily traffic* (MADT) and the *annual average daily traffic* (AADT) [Bri00]. The MADT is calculated as the average amount of traffic recorded during each day of the week, averaged over all days within a month. Averaging the resulting twelve MADTs

[9]The TRB was formerly known as the *Highway Research Board* (HRB).

gives the AADT.

> Regarding the use of the LOS, we note that it is a rather old-fashioned method for evaluating the quality of traffic operations. In general, it is difficult to calculate, mainly because the defined standards at which the different levels are set, always depend on the specific type of traffic situation that is studied (e.g., type of road, ...). This makes the LOS more of an engineering tool, used when assessing and planning operational analyses. Instead of using the LOS, we therefore propose to adopt the more suited approach based on oblique cumulative plots (we refer the reader to Section 2.3.2.2). These allow for example to assess the differences between travel times under free-flow and congested conditions, thereby giving a more meaningful and intuitive indication of the quality of traffic operations to the drivers.

2.4.3 Travel times and their reliability

When travelling around, people like to know how long a specific journey will take (e.g., by public transport, car, bicycle, ...). This notion of an expected travel time, is one of the most tangible aspects of journeying as perceived by the travellers. When people are travelling to their work, they are required to arrive on time at their destinations. Based on this premise, we can naturally state that people reason with a built-in safety margin: they consider the *average time* it takes to reach a destination, and use this to decide about their departure time.

Aside from the above obvious human rationale, there is also an increased interest in obtaining precise information with respect to travel times in the context of *advanced traveller information systems* (ATIS). Here, an essential ingredient is the accurate prediction of future travel times. Coupled with incident detection for example, drivers can obtain correct travel time information, thereby staying informed of the actual traffic conditions and possibly changing their journey. The requested information can reach the driver by means of a cell-phone (e.g., as a feature offered by the mobile service provider), it can be broadcasted over radio (e.g., the *Traffic Message Channel* – TMC), or it can be displayed using *variable message signs* (VMS) above certain road sections (e.g., *dynamic route information panels* – DRIPs), ...

2.4.3.1 Travel time definitions

The travel time of a driver completing a journey, can be defined as 'the time necessary to traverse a route between any two points of interest' [Tur98]. In this context, the *experienced dynamic travel time*, starting at a certain time t_0, over a road section of length K is defined as follows [Bov00]:

$$T(t_0) = \int_0^K \frac{1}{v(t,x)} dx \qquad \forall\, t \geq t_0, \tag{2.39}$$

for which it is assumed that all local *instantaneous* vehicle speeds $v(t,x)$ are known at all points along the route, and at all time instants (hence the term *dynamic* travel time). In most cases however, we do not know all the $v(t,x)$, but only a finite subset of them, defined by the locations of the detector stations (demarcating section boundaries). The travel time can then be approximated using the recorded speeds at the beginning and end of a section (there is an underlying assumption here, namely that vehicles travel at a more or less constant speed between detector locations). As stated earlier, the experienced travel time requires the knowledge of local vehicle speeds at *all* time instants after T_0. Because this is not always possible, a simplification can be used, resulting in the so-called *experienced instantaneous travel time*:

$$\widetilde{T}(t_0) = \int_0^K \frac{1}{v(t_0,x)} dx, \tag{2.40}$$

In general, we can derive the travel time using equation (2.27), i.e., the total distance travelled by all the vehicles, divided by their space-mean speed:

$$T(t_0) = \frac{K}{\overline{v}_{\mathrm{s}}(t_0)}, \tag{2.41}$$

in which an accurate estimation of the space-mean speed $\overline{v}_{\mathrm{s}}(t_0)$ at time t_0 is necessary (e.g., by taking the harmonic average of the recorded spot speeds).

2.4.3.2 Queueing delays

Traffic congestion nearly always leads to the build up of queues, introducing an increase (i.e., the *delay*) in the experienced travel time. The congestion itself can have originated due to traffic demand exceeding the capacity, or because an incident occurred (e.g., road works, a traffic accident, ...)[10]. This can create *incidental* (non-recurrent) or *structural* (recurrent) congestion. Congestion can thus be seen as a loss in travel time with respect to some base line reference. Two such commonly used references are the travel time under free-flow conditions, and the travel time under maximum (i.e., capacity) flow. The delay is typically expressed in vehicle hours. As stated earlier, there are several ways to inform a driver of the current and predicted travel time. Using DRIPs it is possible to advertise the extra travel time (the delay is now typically expressed in vehicle minutes), as well as queue lengths. We note that in our opinion it is more intuitive to advertise a temporal estimation (i.e., the travel time or the delay), than a spatial estimation (e.g., the queue length on a motorway).

[10]Note that in a broader sense, queueing delays also encompass delays at signalised and unsignalised intersections.

2.4.3.3 An example of travel time estimation using cumulative plots

There exist several techniques for estimating the current travel time; one method for directly 'measuring' the travel time, is by using a probe vehicle (we refer the reader to Section 2.3.5 for more details). This way, it is possible to extract actual travel times from a traffic stream. Note that as traffic conditions get more congested, more probe vehicles are required in order to obtain an accurate estimation of the travel time.

Another method for measuring the travel time, is based on historical data, namely cumulative plots (introduced in Section 2.3.2.2). As mentioned earlier, the travel time can then be measured as the distance along the horizontal (or oblique) time axis; any excess due to delays can then easily be spotted on a set of oblique cumulative plots.

Based on cumulative plots of consecutive detector stations, we can calculate the travel time between the upstream and downstream end of a road section. To illustrate this, let us reconsider the cumulative curves shown in Figure 2.4 of Section 2.3.2.2. The evolution of the travel time during the day for these curves, is depicted in the top part of Figure 2.5. The derived histogram (indicative of the underlying travel time probability density function), in the bottom part of the figure, shows that the mean travel time during the day is approximately 4 minutes.

We already mentioned the likely occurrence of an incident at 11:00, resulting in the formation of a queue. During this period, the travel time shot up, reaching first 5, then 7 minutes. Looking at the top part of Figure 2.5, we furthermore notice a slight increase in the travel time at approximately 18:45, for a short period of some 10 minutes. Investigation of the detector data, revealed that the flow remained constant at about 4500 vehicles per hour, but the speed dropped to some 90 km/h (as opposed to 110 km/h); we can conclude that all vehicles were probably simultaneously slowing down during this period (perhaps a rubbernecking effect). Another possibility is a platoon of slower moving vehicles, but then it would seem to have dissipated rather quickly after 10 minutes.

Using ample historical data, we can analyse the travel time over a period of many weeks, months, or even years. This would allow us to make *intuitive* statements such as:

> *"The typical travel time over this section of the road during a working Monday, lies approximately between 4 and 6 minutes. There is however an 8% probability that the travel time increases to some 22 minutes (e.g., due to an occurring incident)."*

Finally note that, besides the two previously mentioned techniques for estimating travel times, an extensive overview can be found in the *Travel Time Data Collection Handbook* [Tur98]. Another concise but more theoretically-oriented overview is provided by Bovy and Thijs [Bov00].

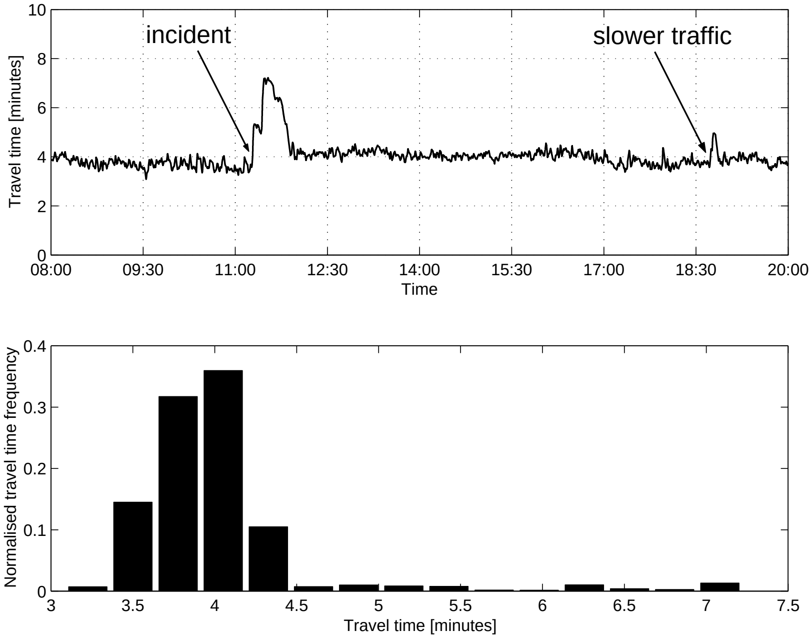


Figure 2.5: *Top:* The evolution of the travel time during one day, based on the cumulative plots from Section 2.3.2.2. As can be seen, an incident likely occurred at 11:00, increasing the travel time from 4 to 7 minutes. Furthermore, at approximately 18:45 in the evening, all traffic seemed to simultaneously slow down for a period of some 10 minutes. *Bottom:* Based on the calculated travel times during the day, we can derive a histogram that is an approximation of the underlying travel time probability density function. The mean is located around 4 minutes.

2.4.3.4 Reliability and robustness properties

As mentioned in the introduction of this section, people reason about their expected travel times based on a built-in safety margin. Central to this is the concept of the *average travel time*. The *reliability* of such a travel time is then characterised by its *standard deviation*. Drivers typically accept (and sometimes expect) a small delay in their expected travel time. A traveller knows the *expected travel time* because of the familiarity with the associated trip. To the traveller, this is personal historical information, for instance obtained by learning the trip's details (e.g., the traffic conditions during a typical morning rush hour) [Bal04b].

Directly linked to the reliability of a certain expected travel time, is its variability. They are said to be unreliable when both expected and experienced travel times differ sufficiently. A typical characterisation of reliability involves the mean and standard deviation (i.e., the variance, which is a measure of variability) of a travel time distribution [Che02]. An example of such a travel time distribution for one day is shown in the histogram in the bottom part of Figure 2.5.

Both first- and second-order measures of a distribution are by themselves insufficient to capture the complete picture. In order to grasp the notion of the previously mentioned safety margin, another typical statistical measure is considered, namely the *90th percentile*. The rationale behind the use of this percentile is that travellers adopt a certain 'safe' threshold with respect to their expected journey times. Considering the 90th percentile, this means that only one out of ten times the experienced travel time will differ significantly from the expected travel time. Travel time reliability can thus be viewed upon as a measure of service quality (similar to the concept of 'quality of service' (QoS) in telecommunications).

There has been some research into the analytic form of travel time distributions (e.g., the work of Arroyo and Kornhauser, concluding that a lognormal distribution seems the most appropriate [Arr05]). There exist however significant differences between travel time distributions: in general, a smaller standard deviation indicates a better service quality and reliability. In contrast to this, a large standard deviation is indicative of chaotic behaviour of the traffic flow, the latter being totally unstable. Furthermore, travel time distributions can have a long tail; this signifies seldom events (e.g., incidents), that can have significant repercussions on the quality of traffic operations.

> Let us finally note that there is an increased interest in the *reliability of complete transportation networks*, and their *robustness* against incidents. To this end, Immers et al. consider reliability as a user-oriented quality, whereas robustness is more a property of the system itself [Imm04]. Among several characterising factors for robustness of transportation systems, they also introduce the following practical notions in this context: *redundancy*, denoting a spare capacity, and *resilience*, which is the ability to repeatedly recover from a temporary overload. Their conclusion is that the key element in securing transportation reliability lies in a good network design.

2.4.4 Efficiency

In [Che01b], Chen et al. state that the main reason for congestion is not demand exceeding capacity (i.e., the number of travellers who *want* to use a certain part of the transportation network, exceeds the available infrastructure's capacity), but is in fact the inefficient operation of motorways during periods of high demand. In order to quantify this efficiency, they first look at what the prevailing speed is when a motorway is operating at its maximum efficiency, i.e., the highest flow (corresponding to the effective capacity, which is different from the HCM's capacity which is calculated from the road's physical characteristics). Based on the distribution of 5-minute data samples from some 3300 detectors, they investigate the speed during periods of very high flows. This leads them to a so-called sustained speed $\overline{v}_{\text{sust}} = 60$ miles per hour (which corresponds to 60 mi/h $\times$ 1.609 $\approx$ 97 km/h).

The performance indicator they propose, is called the *efficiency* η and it based on the ratio of the *total vehicle miles travelled* (VMT), divided by the *total vehicle hours* *travelled* (VHT). Note that as the units of VMT and v_{sust} should correspond to each other, we propose to use the terminology of *total vehicle distance travelled* (VDT) instead of the VMT, in order to eliminate possible confusion. Both VDT and VHT are defined as follows:

$$\text{VDT} = q\,K, \tag{2.42}$$

$$\text{VHT} = \frac{\text{VDT}}{\overline{v}_{\text{s}}}, \tag{2.43}$$

with, as before, q the flow, K the length of the road section, and $\overline{v}_{\text{s}}$ the space-mean speed. Using the above definitions, we can write the efficiency of a road section as:

$$\eta = \frac{\text{VDT}/\text{VHT}}{\overline{v}_{\text{sust}}}. \tag{2.44}$$

The efficiency is expressed as a percentage, and it can rise above 100% when the recorded average speeds surpass the sustained speed $\overline{v}_{\text{sust}}$. In general, the discussed efficiency can also easily be calculated for a complete road network and an arbitrary time period. It can furthermore be seen as the ratio of the actual productivity of a road section (the *output* produced by this section during one hour), to its maximum possible production (the *input* to the section) under high flow conditions.

> Note that as a solution to their original claim (*“congestion arises due to inefficient operation”*), Chen et al. propose to increase the operational efficiency, mainly through the technique of suitable ramp metering (using an idealised ramp metering control practice that maintains the occupancy downstream of an on ramp to its critical level). But in our opinion, they neglect to take into account the entire situation, i.e., they fail to consider the extra effects induced by holding vehicles back at some on ramps (e.g., the total time travelled by *all* the vehicles, including delays), thus rendering their statement practically worthless by giving a feeble argument. Careful examination of their reasoning, reveals that these extra effects are dealt with by shifting demand during the peak periods, but this just confirms our hypothesis that congestion occurs when demand exceeds capacity, even when this capacity is for example controlled through ramp metering !

In contrast to the work of Chen et al., Brilon proposes another definition for the *efficiency* (now denoted as E): it is expressed as the number of vehicle kilometres that are produced by a motorway section per unit of time [Bri00]:

$$E = q\,\overline{v}_{\text{s}}\,T_{\text{mp}}, \tag{2.45}$$

with now q the total flow recorded during the time interval T_{mp}. Brilon concludes that in order for motorways to operate at maximum efficiency, their hourly flows typically have to remain *below* the capacity flow (e.g., at 90% of q_{cap}). Brilon also proposes to use this point of maximum efficiency as the threshold when going from LOS D to LOS E.

2.5 Fundamental diagrams

Whereas the previous sections dealt with individual traffic flow characteristics, this section discusses some of the relations between them. We first give some characterisations of different traffic flow conditions and the rudimentary transitions between them, followed by a discussion of the relations (which are expressed as fundamental diagrams) between the traffic flow characteristics, giving special attention to the different points of view adopted by traffic engineers.

2.5.1 Traffic flow regimes

Considering a stream of traffic flow, we can distinguish different types of operational characteristics, called *regimes* (two other commonly used terms are traffic flow *phases* and *states*). As each of these regimes is characterised by a certain set of unique properties, classification of them is sometimes based on occupancy measurements (see for example the discussion about levels of service in Section 2.4.2), or it is based on combinations of different macroscopic traffic flow characteristics (e.g., the work of Kerner [Ker04]).

In the following sections, we discuss the regimes known as *free-flow* traffic, *capacity-flow* traffic, *congested*, *stop-and-go*, and *jammed* traffic. Our discussion of these regimes is in fact based on the commonly adopted way of looking at traffic flows, as opposed to for example Kerner's three-phase traffic theory that includes a regime known as *synchronised* traffic (we refer the reader to Section 2.5.4 for more details). We conclude the section with a note on the transitions that occur from one regime to another.

2.5.1.1 Free-flow traffic

Under light traffic conditions, vehicles are able to freely travel at their desired speed. As they are largely unimpeded by other vehicles, drivers strive to attain their own comfortable travelling speed (we assume that in case a vehicle encounters a slower moving vehicle ahead, it can easily change lanes in order to overtake the slower vehicle). Notwithstanding this ability for unconstrained travelling, drivers have to take into account the maximum allowed speed (denoted by v_{max}), as well as road-, engine-, and other vehicle characteristics. Note that in some cases, depending on the country under scrutiny, drivers perform speeding.

In essence, the previous description of *free-flow traffic* considers a traffic flow to be unrestricted, i.e., no significant delays are introduced due to possible overtaking manoeuvres. As a consequence, the *free-flow speed* (by some called the *nominal speed*) is the mean speed of all vehicles, travelling at their own pace (e.g., 100 km/h); it is denoted by $\overline{v}_{\text{ff}}$.

Free-flow traffic occurs exclusively at low densities, implying large average space headways according to equation (2.11). As a result, small local disturbances in the temporal and spatial patterns of the traffic stream have no significant effects, hence traffic flow is *stable* in the free-flow regime.

2.5.1.2 Capacity-flow traffic

When the traffic density increases, vehicles are driving closer to each other. Considering the number of vehicles that pass a certain location alongside the road, an observer will notice an increase in the flow. At a certain moment, the flow will reach a maximum value (which is determined by the mean speed of the traffic stream and the current density). This maximum flow is called the *capacity flow*, denoted by q_{c}, q_{cap}, or even q_{max}. A typical value for the capacity flow on a three-lane Belgian motorway with v_{max} equal to 120 km/h, can reach a maximum of some 7000 vehicles [VVC03]. According to equation (2.19), the average time headway is minimal at capacity-flow traffic, indicating the (local) formation of tightly packed clusters of vehicles (i.e., platoons), which are moving at a certain *capacity-flow speed* $\overline{v}_{\text{c}}$ (or $\overline{v}_{\text{cap}}$) which is normally a bit lower than the free-flow speed. Note that some of these fast platoons are very unstable when they are composed of tail-gating vehicles: whenever in such a string a vehicle slows down a little, it can have a cascading effect, leading to exaggerate braking of following vehicles. Hence, these latter manoeuvres can destroy the local state of capacity-flow, and can in the worst case lead to multiple rear-end collisions. At this point, traffic becomes *unstable*.

> The calculation of the capacity flow is a daunting task, holding traffic engineers occupied for the last six decades. The fact of the matter is that there exists no rigourous definition for the concept of 'capacity'. As a result, after many years of research, this culminated in the publication of the fourth edition of the already previously mentioned *Highway Capacity Manual*. It contains an impressive overview, spanning methodologies for assessing the capacity at specific types of road infrastructures (motorway facilities, weaving sections, on- and off-ramps, signallised and unsignallised urban intersections, ...) [HCM00]. Despite this impressive amount of literature, the HCM's approach remains firmly rooted in the level-of-service concept, still adopting a static view on congestion. Only quite recently (i.e., the 2000 version) have simulations entered the picture in order to account for the temporal nature of the formation and dissolution of congestion at motorway complexes.

2.5.1.3 Congested, stop-and-go, and jammed traffic

Considering the regime of capacity-flow traffic, it is reasonable to assume that drivers are more mentally aware and alert in this regime, as they have to adapt their driving style to the smaller space and time headways under high speeds. However, when more vehicles are present, the density is increased even further, allowing a sufficiently large disturbance to take place. For example, a driver with too small space and time headways, will have to brake in order to avoid a collision with the leader directly in front; this can lead to a local chain of reactions that disrupts the traffic stream and triggers a breakdown of the flow. The resulting state of saturated traffic conditions, is called *congested traffic*. The moderately high density at which this breakdown occurs, is called the *critical density*, and is denoted by k_c or k_{crit} (for a typical motorway, its value lies around 25 vehicles (PCUs) per kilometre per lane, [VVC03]). From this knowledge, we can derive the optimal driving speed for single-lane traffic flows as $\overline{v}_s = q_{\text{cap}}/k_{\text{crit}} = \frac{7000}{3} \div 25 \approx 93$ km/h.

Higher values for the density indicate almost always a worsening of the traffic conditions; congested traffic can result in *stop-and-go traffic*, whereby vehicles encounter so-called *stop-and-go waves*. These waves require them to slow down severely, or even stop completely. When traffic becomes motionless, the space headway reaches a minimum as all vehicles are standing bumper-to-bumper; this extreme state is called *jammed traffic*. Clearly, there exists a maximum density at which the traffic seems to turn into a 'parking lot', called the *jam density* and it is denoted by k_j, k_{jam}, or k_{max}. For a typical motorway, its value lies around 140 vehicles (PCUs) per kilometre per lane [VVC03].

> Note that the jam density is typically expressed in vehicles per kilometre. As already stated in the introduction of Section 2.3.1, density ignores the effects of traffic composition and vehicle lengths. For a typical value of some 140 vehicles/km/lane for the jam density, this means that we express the density by using passenger car units (see Section 2.3.1.2 for more details). Suppose now for example that an average trailer truck equals 4.5 PCUs, then the jam density would decrease to some 140 $\div$ 4.5 $\approx$ 31 trucks for this class of vehicles. *As a consequence, the value of the jam density is different for each vehicle class.*

2.5.1.4 A note on the transitions between different regimes

Streams of traffic flows can be regarded as many-particle systems (e.g., gasses, magnetic spin systems, ...); as they have a large number of degrees of freedom, it is often intractable when it comes to solving them exactly. However, from a physical point of view, these systems can be described in the framework of *statistical physics*, whereby the collective behaviour of their constituents is approximately treated using statistical techniques.

Within this context, the changeover from one traffic regime to another, can be looked upon as a *phase transition*. Within thermodynamics and statistical physics, an *order parameter* is often used to describe the phase transition: when the system shifts from one phase to another (e.g., at a *critical point* for liquid-gas transitions), the order parameter expresses a different qualitative behaviour. Two examples of such an order parameter that is applicable to traffic flows, can be found in Schadschneider et al. who considered nearest neighbour correlations [Sch02a], and in Jost and Nagel who devised a measure of inhomogeneity [Jos03a] (we refer the reader to our work in [Mae04f] for an example in which they are used and compared when tracking phase transitions).

There exists a difference in which a phase transition can express itself. This difference is designated by the order of the transition; generally speaking, the two most common phase transitions are *first-order* and *second-order transitions*. According to *Ehrenfest's classification*[11], first-order transitions have an abrupt, discontinuous change in the order parameter that characterises the transition. In contrast to this, the changeover to the new phase occurs smoothly for second-order transitions [Yan52; Lee52]. Note that higher-order phase transitions also exist, e.g., in superconducting materials [Cro01].

With respect to the description of regimes in traffic flows, it is commonly agreed that there exists a first-order phase transition when going from the capacity-flow to the congested regime. The point at which this transition occurs, is the critical density. Studying the phase transitions encountered in fluid dynamics, there exists a transition from the *laminar* flow (i.e., a fluid flowing in layers, each moving at a different velocity) to the *turbulent* flow (i.e., the disturbed random and unorganised state in which vortices form). However, the transition here is triggered by an increase in the velocity of the fluid, as opposed to the transition in traffic flows where a change in the density can lead to a cascading instability. In this respect, the analogy for traffic flows holds better when comparing them to gas-liquid transitions. Here free-flow traffic corresponds to a gaseous phase, in which particles are evenly spread out in the system. At the point of the phase transition, liquid droplets will form, coagulating together into bigger droplets. This leads to a state where both gaseous and liquid phases coexist, typically in the form of a big liquid droplet surrounded by gas particles. For even higher densities, particles are so close to each other, and the only remaining state is the liquid phase [Eis98; Kra99; Kay01; Jos02; Jos03a; Nag03a; Mah05].

In conclusion, we refer the reader to the work of Tampère, where an excellent overview is given, detailing the different traffic flow regimes, their transitions, and mechanisms with respect to jamming behaviour [Tam04a].

2.5.2 Correlations between traffic flow characteristics

Whereas the previous sections all treated the macroscopic traffic flow characteristics on an individual basis, this section considers some of the relations between them. We start our discussion with a look at the historic origin of fundamental diagrams, after which we shed some light on the different classical approaches. The section concludes with some considerations with respect to empirical measurements.

[11] Note that the modern approach to the classification of phase transitions, relates first-order transitions to systems that have *mixed phases*, releasing and/or absorbing energy. In contrast to this behaviour, second-order phase transitions, also called continuous phase transitions, have no associated latent heat.

2.5.2.1 The historic origin of the fundamental diagram

As in many scientific disciplines, the resulting statements and theories are often preceded by an investigation of obtained experimental data, which serves as empirical evidence for them. In this line of reasoning, Greenshields was among the first to provide — as far back as 1935 — a basis for most of the classical work on, what are called, *empirical fundamental diagrams*. In his seminal paper, he sketched a *linear relation between the density and the mean speed*, based on empirically obtained data [Gre35]:

$$\overline{v} = \overline{v}_{\text{ff}} \left(1 - \frac{k}{k_{\text{j}}} \right). \tag{2.46}$$

As can be seen from Greenshields' relation, when increasing the density from zero to the jam density k_{j}, the mean speed will monotonically decrease from the free-flow speed $\overline{v}_{\text{ff}}$ to zero (note that we dropped the 's' or 't' subscript from the mean speed, as it is not sure whether or not Greenshields used space- or time-mean speed, respectively). The relation can be understood intuitively, by assuming that drivers will tend to slow down in crowded traffic, because this naturally gives them more time to react to changes (e.g., sudden braking of the lead vehicle). As it is reasonable to assume that the mean speed remains unaffected for very low densities, Greenshields furthermore flattened the upper-left part of the regression line (corresponding to the free-flow speed), although this effect is not incorporated in equation (2.46).

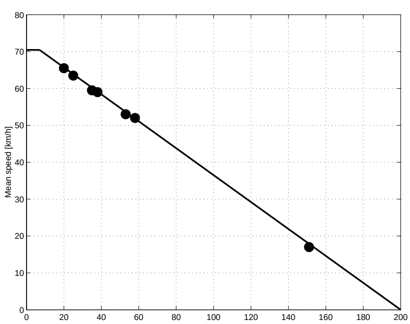


Figure 2.6: Greenshields' original linear relation between the density and the mean speed. Note that the regression line is based on only seven measurements points, and that there is an artificial flattening of its upper-left part (figure based on [Lig55] and [Gar97]).

Although Greenshields' derivation of the linear relation between density and space-mean speed appears elegant and simple, it should nevertheless be taken with a grain of salt. The fact of the matter is that his hypothesis is, as can be seen in Figure 2.6, based on *only seven measurement points*, which comprise aerial observations taken on September 3, (Monday, Labor Day), 1934 [May90]. One of the problems is that these observations are *not independent*. An even more serious problem is that six of these observations were obtained for free-flow conditions, whereas the one single point that indicates congested conditions, was obtained at an entirely *different road*, on a *different day* [Gar97] !

Some twenty years later, Lighthill and Whitham developed a theory that describes the traffic flows on long crowded roads using a first-order fluid-dynamic model (see also Section 3.2.1.2) [Lig55]. As one of the main ingredients in their theory, they postulated the following fundamental hypothesis: *"at any point of the road, the flow q is a function of the density k"*. They called this function the *flow-concentration curve* (recall from Section 2.3.3 that density in the past got sometimes referred to as concentration).

Continuing their reasoning, Lighthill and Whitham then referred to Greenshields' earlier work, relating the space-mean speed to the density, and, by means of equation (2.33), thus relating the flow to the density. The *existence* of the concept of the flow-concentration curve mentioned above, was justified on the grounds that it describes traffic operating under steady-state conditions, i.e., homogeneous and stationary traffic as explained in Section 2.3.4.2. In this context, the flow-concentration curve therefore describes the *average characteristics* of a traffic flow. So Greenshields first fitted a regression line to scarce data, after which his functional form seemed to be taken for granted for the following seventy years. The key aspect in Lighthill and Whitham's (and also Richards' [Ric56]) approach, lay in the fact that they broadened the flow-concentration curve's validity, including also conditions of non-stationary traffic. They also stated that, because of, e.g., changes in the traffic composition, the curve can vary from day to day, or even within a day (e.g., rush hours, ...). The same statement holds also true when considering the flow-concentration curves of different vehicle classes (e.g., cars and trucks).

The term *fundamental diagram* itself, is historically based on Lighthill and Whitham's *fundamental* hypothesis of the existence of such a *one-dimensional* flow-concentration curve. As traffic engineers grew accustomed to the graphical representation of this curve, they started talking about the *diagram* that represents it, i.e., the 'fundamental diagram' [Hai63].

> In its original form, the fundamental diagram represents an *equilibrium relation* between flow and density, denoted by $q_{\mathrm{e}}(k)$. But note that, because of the fundamental relation of traffic flow theory (see Section 2.3.4.2), is it equally justified to talk about the $\overline{v}_{\mathrm{s_e}}(k)$ or the $\overline{v}_{\mathrm{s_e}}(q)$ fundamental diagrams. Due to this equilibrium property, the traffic states (i.e., the density, flow, and space-mean speed) can be thought of as 'moving' over the fundamental diagrams' curves.

2.5.2.2 The general shape of a fundamental diagram

We now give an overview of some of the qualitative features of the different possible fundamental diagrams, representing the equilibrium relations between density, space-mean speed, and average space headway, and flow. Note that in each example, we consider a *possible* fundamental diagram, as they can take on many (functional) shapes.

Space-mean speed versus density
We start our discussion based on the equilibrium relation between space-mean speed and density, i.e., the $\overline{v}_{\mathrm{s_e}}(k)$ fundamental diagram. The main

reason for starting here, is the fact that this diagram is the easiest to understand intuitively. Complementary to the example of Greenshields in Figure 2.6, we give a small overview of its most prominent features:

- the density is restricted between 0 and the maximum density, i.e., the jam density k_j,
- the space-mean speed is restricted between 0 and the maximum average speed, i.e., the free-flow speed $\overline{v}_\mathrm{ff}$,
- as density increases, the space-mean speed monotonically decreases,
- there exists a small range of low densities, in which the space-mean speed remains unaffected and corresponds more or less to the free-flow speed,
- and finally, the flow (equal to density times space-mean speed), can be derived as the area demarcated by a rectangle who's lower-left and upper-right corners are the origin and a point on the fundamental diagram, respectively.

Space-mean speed versus average space headway
Microscopic and macroscopic traffic flow characteristics are related to each other by means of equations (2.11) and (2.19). According to the former, density k is inversely proportional to the average space headway $\overline{h}_\mathrm{s}$. We can therefore derive a fundamental diagram, similar to the previous one, by substituting the density with the average space headway. As as result, the abscissa gets 'inverted', resulting in the fundamental diagram as shown in Figure 2.7.

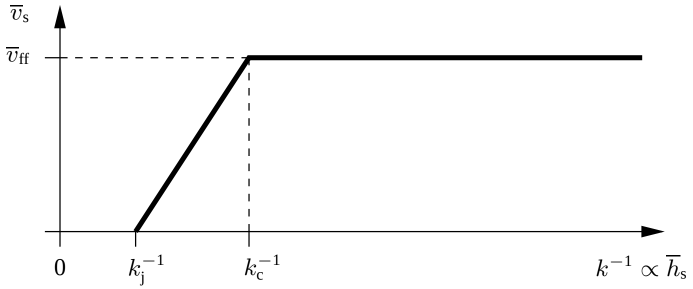


Figure 2.7: A fundamental diagram relating the average space headway $\overline{h}_\mathrm{s}$ to the space-mean speed $\overline{v}_\mathrm{s}$. Note that the average space headway is proportional to the inverse of the density, i.e., k^{-1}.

The interesting features of this type of fundamental diagram, can be summed as follows:

- the curve starts not in the origin, but at k_j^{-1}, corresponding to the average space headway when the jam density is reached (i.e., all vehicles are standing nearly bumper to bumper),

- as the average space headway increases, its inverse (the density) decreases, and the space-mean speed increases,
- the space-mean speed continues to rise with an increasing average space headway, until it reaches the maximum average speed, i.e., the free-flow speed $\overline{v}_{\text{ff}}$; this happens at the inverse of the critical density k_{c}^{-1},
- from then on, the space-mean speed remains constant with increasing average space headway.

The above features can be understood intuitively: at large average space headways, a driver experiences no influence from its direct frontal leader. However, there exists a point at which the driver comes 'close enough' to this leader (i.e., in crowded traffic), so that its speed will decrease. This slowing down will continue to persist as traffic gets more dense (this the same reasoning behind Greenshields' derivation in Section 2.5.2.1).

Flow versus density

Probably the most encountered form of a fundamental diagram, is that of flow versus density. Its origins date back to the seminal work of Lighthill and Whitham who, as described earlier, referred to it as the flow-concentration curve. An example of the $q_{\text{e}}(k)$ fundamental diagram is depicted in Figure 2.8.

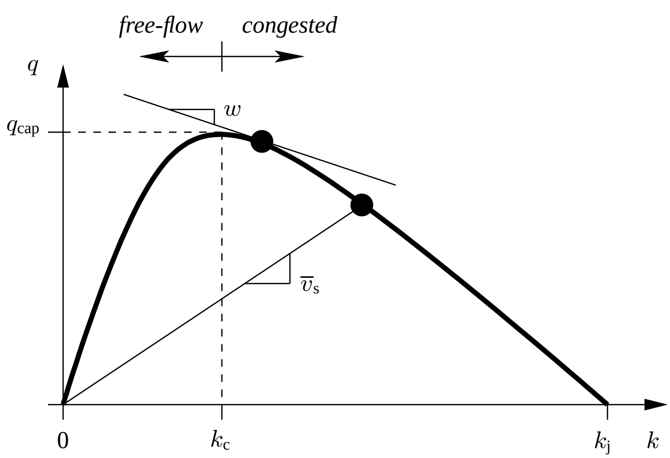


Figure 2.8: A fundamental diagram relating the density k to the flow q. The capacity flow q_{cap} is reached at the critical density k_{c}. The space-mean speed $\overline{v}_{\text{s}}$ for any point on the curve, is defined as the slope of the line through that point and the origin. Taking the slope of the tangent to points on the curve, gives the characteristic wave speed w.

Noteworthy features of this type of fundamental diagram are:

- for moderately low densities (i.e., below the critical density k_c), the flow increases more or less linearly (this is called the *free-flow branch* of the fundamental diagram),
- near the critical density k_c, the fundamental diagram can bend slightly, due to faster vehicles being obstructed by slower vehicles, thereby lowering the free-flow speed [New93b],
- at the critical density k_c, the flow reaches a maximum, called the capacity flow[12] q_cap,
- in the congested regime (i.e., for densities higher than the critical density), the flow starts to degrade with increasing density, until the jam density k_j is reached and traffic comes to a stand still, resulting in a zero flow (this is called the *congested branch* of the fundamental diagram),
- the space-mean speed $\overline{v}_\text{s}$ for any point on the $q_\text{e}(k)$ fundamental diagram, can be found as the slope of the line through that point and the origin,

There is one more piece of information revealed by the $q_\text{e}(k)$ fundamental diagram: When taking the slope of the tangent in any point of the diagram, we obtain what is called the *kinematic wave speed*. These speeds w correspond to *shock waves* encountered in traffic flows (e.g., the stop-and-go waves). As can be seen from the figure, the shock waves travel forwards, i.e., *downstream*, in free-flow traffic ($w \geq 0$), but backwards, i.e., *upstream*, in congested traffic ($w \leq 0$).

The above shape of the $q_\text{e}(k)$ fundamental diagram is just one possibility. There exist many different flavours, originally derived by traffic engineers seeking a better fit of these curves to empirical data. After the work of Greenshields, another functional form — based on a logarithm — was proposed by Greenberg [Gre59]. Another possible form was introduced by Underwood [Und61]. All of the previous diagrams are called *single-regime models*, because they formulate only one relation between the macroscopic traffic flow characteristics for the entire range of densities (i.e., traffic flow regimes) [May90]. In contrast to this, Edie started developing *multi-regime models*, allowing for discontinuities and a better fit to empirical data coming from different traffic flow regimes [Edi61]. We refer the reader to the work of Drake et al. [Dra67] and the book of May [May90] for an extensive comparison and overview of these different modelling approaches (note that Drake et al. used time-mean speed).

During the last two decades, other, sometimes more sophisticated, functional relationships between density and flow have been proposed. Examples are the work of Smulders who created a non-differentiable point

[12] Note that this capacity flow is not an extreme value, i.e., it can be different from the maximum observed flow. The reason is that, with respect to the nature of the fundamental diagram, the capacity flow is taken to be an *average* value [Gre35; Lig55].

at the critical density in a two-regime fundamental diagram [Smu89], the METANET model of Messmer and Papageorgiou who's single-regime fundamental diagram contains an inflection point near the jam density [Mes90], the work of De Romph who generalised Smulders' functional description of his two-regime fundamental diagram [Rom94], the typical triangular shape of the fundamental diagram introduced by Newell, resulting in only two possible values for the kinematic wave speed w [New93b], … As can be seen, these fundamental diagrams sometimes take on non-concave forms, depending on the existence of inflection points in the functional relation between flow and density. In general, they can be convex, concave, (dis)continuous, piecewise-linear, everywhere differentiable, have inflection points, … Variations in shape will continu to be proposed, as it is for certain that there is no general consensus among traffic engineers regarding the correct shape of this fundamental diagram. To illustrate this, a more exotic approach is based on *catastrophe theory*, which is, in a sense, a three-dimensional model that jointly treats density, flow, and space-mean speed. Acha-Daza and Hall applied the technique, resulting in a satisfactory fit with empirical data [AD94].

The most extreme argument with respect to the shape of the fundamental diagram, came from Kerner who questioned its validity, and consequently rejected it altogether by replacing it with his fundamental hypothesis of *three-phase traffic flow theory* (refer to Section 2.5.4 for more details) [Ker04].

Space-mean speed versus flow

An often spotted shape is that of the $\overline{v}_{s_e}(q)$ fundamental diagram, depicted in Figure 2.9. As opposed to the earlier discussed $\overline{v}_{s_e}(k)$ fundamental diagram, the space-mean speed versus flow curve no longer embodies a function in the strict mathematical sense: for each value of the flow, there exists two different mean speeds, namely one in the free-flow regime (upper branch) and one in the congested regime (lower branch).

Some people, e.g., economists who use the flow to represent traffic demand as will be explained in Section 3.1.4.1, find this kind of fundamental diagram easy to cope with. But in our opinion, we are convinced however, that this diagram is rather difficult to understand at first sight. We believe the $\overline{v}_{s_e}(k)$ fundamental diagram is a much better candidate, because density can intuitively be understood as a measure for how crowded traffic is, as opposed to some flow giving rise to two different values for the space-mean speed.

As a final comment, we would like to point out that the previously discussed bivariate functional relationships between the traffic flow characteristics (e.g., density and flow), are based on observations. More importantly, this means that there is *no direct causal relation* assumed between any two variables. Fundamental diagrams sketch

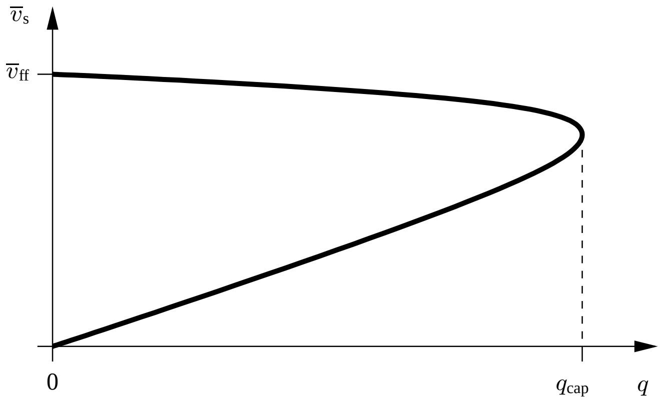


Figure 2.9: A fundamental diagram relating the flow q to the space-mean speed $\overline{v}_s$. The capacity flow q_{cap} is located at the right edge of the diagram, i.e., it is defined as the maximum average flow. Note that there are two possible speeds associated with each value of the flow.

only possible correlations, implying that the *nature of the transitions* between different traffic regimes thus remains to be explored (see Section 2.5.1.4 for a discussion).

2.5.2.3 Empirical measurements

As mentioned earlier, the fundamental diagrams discussed in the previous section represent equilibrium relations between the macroscopic traffic flow characteristics of Section 2.3. In sharp contrast to this, real empirical measurements from detector stations do not describe such nice one-dimensional curves corresponding to the functional relationships.

As an illustrative example, we provide some scatter plots in Figure 2.10. The shown data comprises detector measurements (the sampling interval was one minute) during the entire year 2003; they were obtained by means of a video camera [VVC03] located at the E17 three-lane motorway near Linkeroever[13], Belgium. Because of the nature of this data, we only obtained flows, occupancies, and time-mean speeds. After calculating the average vehicle length, the occupancies were converted into densities using equation (2.24). Using these recorded time series, we then constructed scatter plots of the density, time-mean speed, flow, and average space headway. Note that no substantial changes are introduced in these plots due to, e.g., our using of densities calculated from occupancies, instead of using real measured densities.

As the dimension of time is removed in these scatter plots, Daganzo calls them *time-independent models* [Dag97b]. It is important to understand that *these scatter plots are not fundamental diagrams*, because the latter represent one-dimensional equilibrium curves. According to Helbing, a better designation would be *regression models*

[13]The detector station is called CLO3, which is an acronym for ‘Camera Linkeroever’.

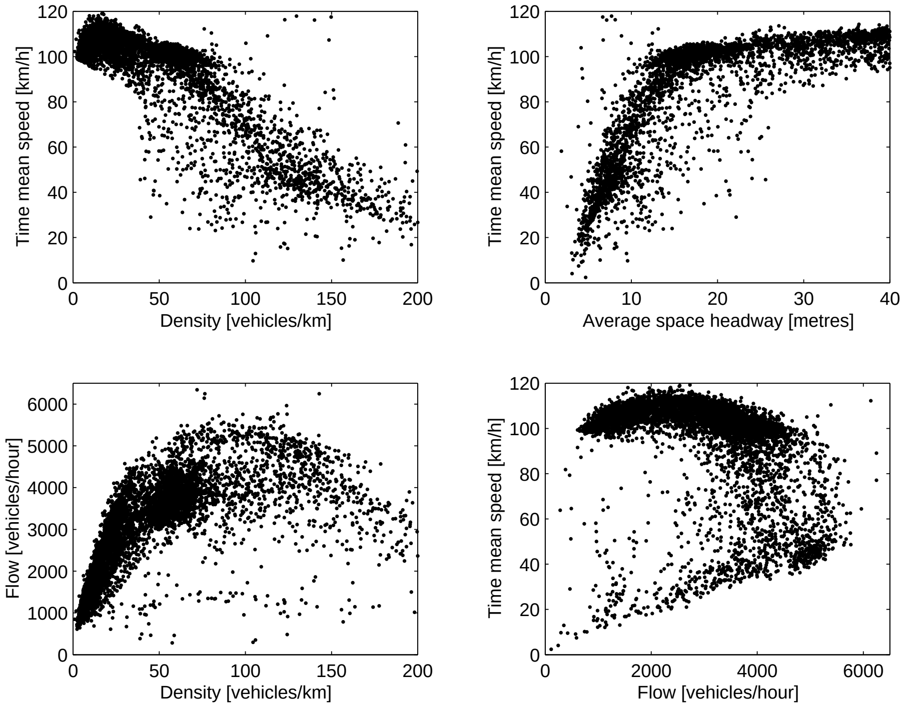


Figure 2.10: Illustrative scatter plots of the relations between traffic flow characteristics as measured by video camera CLO3 located at the E17 three-lane motorway near Linkeroever, Belgium. The measured occupancies were converted into densities, the time-mean speed remained unchanged. Shown are scatter plots of a (k,$\overline{v}_{\text{t}}$) diagram (*top-left*), a ($\overline{h}_{\text{s}}$,$\overline{v}_{\text{t}}$) diagram (*top-right*), a (k,q) diagram (*bottom-left*), and a (q,$\overline{v}_{\text{t}}$) diagram (*bottom-right*).

[Hel97]. In this dissertation, we introduce a terminology based on *phase spaces* (or equivalently *state spaces*), resulting in, e.g., the (k,q) diagram (note that we dropped the adjective ‘fundamental’).

In reality, traffic is not homogeneous, nor is it stationary, thus having the effect of a large amount of scatter in the presented diagrams. In free-flow traffic, interactions between vehicles are rare, and their small local disturbances have no significant effects on the traffic stream. As a result, all points are somewhat densely concentrated along a line — representing the free-flow speed — in all four diagrams. However, in the congested regime, a wide range of scatter is visible due to the interactions between vehicles. Furthermore, vehicle accelerations and decelerations lead to large fluctuations in the traffic stream, as can be seen by the thin, but large, cloud of data points. The effect is especially pronounced for intermediate densities, leading to large fluctuations in the time-mean speed and flow.

The occurrence of all this scatter in the data, leads some traffic engineers to question the validity of the fundamental diagram. More specifically, the behaviour in congested traffic seems ill-defined to some. As stated earlier, Kerner is the most intense opponent in this debate, as he outright rejects Lighthill and Whitham's hypothesis that remained popular over the last fifty years. Despite this criticism, the fundamental diagram remains, to the majority of the community, a fairly accurate description of the average behaviour of a traffic stream. Cassidy and Coifman even provided quantifiable evidence of the existence of well-defined bivariate relations between traffic flow characteristics. The key here was to separate stationary periods from non-stationary ones in the detector data (i.e., stratifying it) [Cas97; Dag97b; Cas98]. Prior work of Del Castillo and Benítez resulted in a more mathematically justified method, for fitting empirical curves in data regions of stationary traffic, after construction of a rigid set of properties that all fundamental diagrams should satisfy [Del95a; Del95b]. Finally, note that Nishinari et al. and Treiber et al. attribute the scatter in the data to the acceleration behaviour of individual vehicles in a traffic stream, based on single-vehicle data of time headways and local speed variances [Nis03c; Nis03b; Tre03; Tre05b].

As a final note, we remark that the distribution of the cloud-like data points of the diagrams in Figure 2.10, is a result of various kinds of phenomena. First and foremost, there is the heterogeneity in the traffic composition (fast passenger cars, slow trailer trucks, ...). Secondly, as already mentioned, the non-stationary behaviour of traffic introduces a significant amount of scatter in the congested regime. Thirdly, each scatter plot is dependent on the type of road, and the time of day at which the measurements were collected. In this respect, the influence of (changing) weather conditions is not to be underestimated (e.g., rain fall results in different diagrams). In conclusion, it is clear that if we want these scatter plots to better fit the fundamental diagrams, all data points should be collected under similar conditions. Even more so, the relative location on the road at which the data points were recorded plays a significant role: e.g., a jam that propagates upstream, passing an on-ramp will show different effects, depending on where the observations were gathered (upstream, right at, or downstream of the on-ramp) and on whether or not the particular bottleneck was active [May90].

2.5.3 Capacity drop and the hysteresis phenomenon

In the early sixties, traffic engineers frequently observed a discontinuity in the measurements near the capacity flow. To this end, Edie proposed a two-regime model that included such a discontinuity at the critical density [Edi61]. Nowadays, this typical form of the $q_{\text{e}}(k)$ fundamental diagram is known as a *reversed lambda shape* (the name was originally suggested by Koshi et al. [Kos83]).

An example of such a reversed λ fundamental diagram, is shown in the left part of Figure 2.11. Note however, that the depicted discontinuity apparently leads to *overlapping* branches of the free-flow and congested regimes, resulting in a *multi-valued* fundamental diagram.

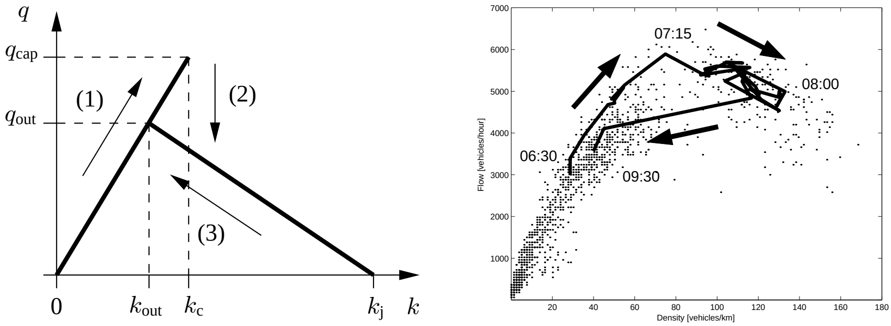


Figure 2.11: *Left:* the typical inverted λ shape of the (k,q) fundamental diagram, showing a capacity drop from q_{cap} to below $q_{\text{out}} \ll q_{\text{cap}}$ (i.e., the queue discharge flow). The hysteresis effect occurs when going from the congested to the free-flow branch, as indicated by the three arrows (1) – (3). *Right:* a (k,q) diagram based on empirical data of one day, obtained by video camera CLO3, at the E17 three-lane motorway near Linkeroever, Belgium. The black dots denote minute measurements, whereas the thick solid line represents the time-traced evolution of traffic conditions. The observed hysteresis loop was based on consecutive 5-minute intervals covering a period that encompasses the morning rush hour between 06:30 and 09:30.

Considering the left part of Figure 2.11, it appears the flow can take on two different values (hence the name 'two-regime, two-capacity' model) depending on the traffic conditions, i.e., whether traffic is moving from the free-flow to the congested regime on the equilibrium curve or vice versa. In order to comprehensively understand this *hysteretic* behaviour, we consider the following intuitive sequence of events:

(1) In the free-flow regime, the flow steadily rises with increasing density, small perturbations in the traffic flow have no significant effects (see Section 2.5.1.1).

(2) At the critical density k_{c}, traffic is said to be *metastable*: for small disturbances, traffic is stable, but when these disturbances are sufficiently large, they can lead to a cascading effect (see Section 2.5.1.2), resulting in a breakdown of traffic and kicking it onto the congested branch. The state of capacity flow at q_{cap} is destroyed, due to a sudden decrease of the flow, called the *capacity drop*.

(3) In order to recover from the congested to the free-flow regime, the traffic density has to be *reduced substantially* (in comparison with the reverse transition), i.e., well below the critical density k_{c}. After this recovery, the flow will *not* be equal to q_{cap}, but to $q_{\text{out}} \ll q_{\text{cap}}$, which is called the *outflow from a jam* or the *queue discharge capacity*.

The above sequence signifies a *hysteresis loop* in the flow versus density fundamental diagram: going from the free-flow to the congested regime occurs via the capacity flow, but the reverse transition proceeds via another way. The phenomenon was first observed by Treiterer and Meyers, who used aerial photography to calculate densities and space-mean speeds, extracted from a platoon of moving vehicles [Tre74]. Hall et al. later observed a similar phenomenon [Hal86].

The right part of Figure 2.11 shows a (k,q) diagram, obtained with empirical data collected at Monday, September 10, 2001. The data was recorded by video camera CLO3, at the E17 three-lane motorway near Linkeroever, Belgium. The small dots represent minute-based measurements, whereas the thick solid line represents the time-traced evolution of traffic conditions. The observed hysteresis loop was based on consecutive 5-minute intervals covering a period that encompasses the morning rush hour between 06:30 and 09:30.

Zhang is among the few who try to give a possible *rigourous mathematical explanation* for the occurrence of this hysteresis phenomenon [Zha99]. His exposition is based on the behaviour of individual drivers during car-following: central to his interpretation is the existence of an *asymmetry* between accelerating and decelerating vehicles (a related notion was already explored by Newell back in 1963 [New63b]). The former are associated with larger space headways, whereas the latter typically have smaller space headways. Both observations can be understood when considering the characteristic ‘harmonica’ effect of a string of consecutive vehicles: when the next stop-and-go wave is encountered, a driver is more alert as he typically has to brake rather hard in order to avoid a collision. But once this wave has passed, a driver gets more relaxed, resulting in a larger response time when applying the gas pedal. The deceleration reaction leads to a sudden decrease of the space headway, whereas the acceleration reaction leads to a gradually developing larger space headway. To this end, Zhang introduces three distinct traffic phases, respectively called the *acceleration phase*, the *deceleration phase*, and a *strong equilibrium* (indicating a constant speed). Because the space headway is thus treated differently under these qualitatively different circumstances, the result is that there are now different functional relations for the $\overline{v}_{\text{s}_\text{e}}(\overline{h}_\text{s})$ fundamental diagram. As a consequence, a hysteresis loop can appear in the (density,flow) state space. Note that Zhang's work describes a *continuous* loop in state space, whereas in most cases hysteresis is assumed to follow a discontinuous fundamental diagram. Furthermore, as there are three different ways for vehicles to reside in a traffic stream (i.e., Zhang's traffic phases), there are now three different capacities related to these conditions; it is the capacity under a stationary equilibrium flow that should be considered as the ideal capacity of a roadway [Zha01b].

Note that depending on the location where the traffic stream measurements were performed, the transition from the free-flow to the congested regime and vice versa does not always have to pass via the capacity flow. Instead, observations can indicate that the traffic state can jump abruptly from one branch to another in the diagram [Gar97]. A possible explanation is that upstream of a jam, vehicles arrive with high speeds, resulting in strong decelerations; a detector station located at this point would observe traffic jumping from the uncongested branch immediately to the congested branch, without necessarily having to pass via the capacity [New82]. This has led Hall et al. to believe the reversed lambda shape is more correctly replaced by a continuous but non-differentiable *inverted V shape* [Hal86].

Continuing this latter train of thought, Daganzo believes that many of these 'extravagant' phenomena (e.g., a multi-valued fundamental diagram) are uncalled for. Applying the stratification methodology of Cassidy [Cas98], the scatter in the empirical data may vanish, restoring a smooth continuous equilibrium relation between density and flow. One way of explaining the high tip of the lambda, is to assume that it is caused by statistical fluctuations that comprise platoons of densely packed vehicles [Dag97b]. However, quite recently, Laval and Daganzo constructed a traffic flow model that incorporates lane changes; they postulate that the lane-changing process limits the vehicles' accelerations, thereby inducing a capacity drop [Lav06].

> During the last seventy years, there has been a continuing quest to find the 'correct' form of the fundamental diagrams. In this respect, we like to stress the fact that 'only looking at the measurements' is not sufficient: traffic engineers wanting to mine the gigabytes of empirical data, should always look at the global picture. This means that the typical driving patterns, *as well as the local geometry/infrastructure*, should also be taken into account, so that the local measurements can be interpreted with respect to the traffic flow dynamics. If this is neglected, the danger exists that traffic is only sampled at discrete locations, giving a sort of 'truncated' view of the occurring dynamical processes.
>
> Finally, we like to agree with Zhang's comments: the root cause of most of the differences in the construction of fundamental diagrams, is the erroneous treatment of data (e.g., mixing data stemming from different traffic flow regimes) [Zha99]. Because fundamental diagrams imply the notion of an equilibrium, care should be taken when using the data, i.e., only considering stationary periods after removing the transients.

2.5.4 Kerner's three-phase theory

In the mid-nineties, Kerner and other fellow researchers, studied various traffic flow measurements stemming from detector stations along German motorways. Initially, they agreed with the classical notion of Lighthill and Whitham's fundamental hypothesis of the existence of one-dimensional equilibrium relation between the macroscopic traffic flow characteristics (see Section 2.5.2.1 for more details). However, upon discovery of a rich and complex set of empirical tempo-spatial patterns in congested traffic flow, Kerner decided to abolish this hypothesis, as it could not adequately capture all of these observed patterns. As a consequence, Kerner rejects *all* traffic flow theories and models that are based on this one-dimensional equilibrium relation [Ker04].

In the search for a more correct theory that could accurately describe empirical traffic flow observations, Kerner developed what is known as the *three-phase theory* of traffic flow.

2.5.4.1 Free flow, synchronised flow, and wide-moving jam

In Section 2.5.1, we elaborated on a classical approach to traffic flow, general assuming two qualitatively different regimes, namely free-flow and congested traffic. Based on empirical findings, Kerner and Rehborn in 1996 proposed three different regimes, separating the congested regime into two other regimes. This led them to the introduction of the following regimes [Ker96b]:

- free flow,
- synchronised flow,
- and wide-moving jam.

The main difference between *synchronised flow* and the *wide-moving jam*, is that in the former low speeds but high flows (comparable to free-flow traffic) can be observed, whereas in the latter both low speeds and low flows are observed. The description by the term ‘synchronised’ was based on the discovery that the time series of flows, densities, and mean speeds exhibited large degrees of correlation among neighbouring lanes. And although synchronised flow is treated as a form of congestion, it nevertheless is characterised by a high continuous flow. Furthermore, a typical tempo-spatial region of synchronised flow has a fixed downstream front (that could be located at a bottleneck’s position), whereas both the upstream and downstream fronts of a wide-moving jam can propagate undisturbed in the upstream direction of a traffic stream [Ker96a].

Kerner distinguishes several congestion patterns with respect to traffic flows. A first typical pattern is a *synchronised-flow pattern* (SP), which can be further classified as a *moving SP* (MSP), a *widening SP* (WSP), and a *localised SP* (LSP). An SP can only contain synchronised flow; as we will shortly mention in Section 2.5.4.3, a moving jam can only occur inside such an SP. When such a jam transforms into a wide-moving jam, the resulting pattern is called a *general pattern* (GP); a GP therefore contains both synchronised flow and wide-moving jams. Just as with the SP, there exist different types of GP. These are a *dissolving GP* (DGP), a *GP under weak congestion*, and a *GP under strong congestion*. A final often encountered pattern occurs when two bottlenecks are spatially close to each other, resulting in what is called an *expanded congested pattern* (EP).

Taking the above considerations into account, the discovery and distinction between both types of congested traffic patterns should be made on the basis of tempo-spatial plots of the speed, rather than the flow (because the flow in synchronised traffic is difficult to differentiate from that of free-flow traffic) [Ker04]. To this end, Kerner et al. developed two applications that are capable of accurately estimating, automatically tracking, and reliably predicting the above mentioned congested traffic patterns. Their models are the *Forecasting of Traffic Objects* (FOTO) and *Automatische StauDynamikAnalyse* (ASDA) [Ker01].

2.5.4.2 Fundamental hypothesis of three-phase traffic theory

Central to Kerner's theory, is the *fundamental hypothesis of three-phase traffic theory*, which basically states that hypothetical steady states of synchronised flow, cover a *two-dimensional region* in a flow versus density diagram (as opposed to the classical notion of a one-dimensional equilibrium relation). An example of such a diagram can be seen in Figure 2.12.

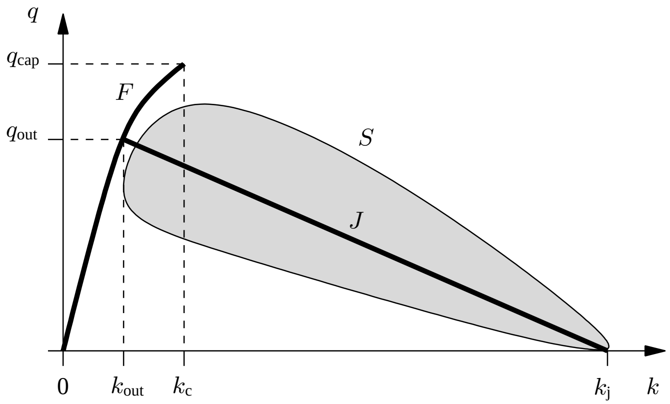


Figure 2.12: The flow versus density relation according to Kerner's three-phase traffic theory. The curve of free flow (denoted by F) is reminiscent of observations in the classical free-flow regime. It levels of a bit towards the capacity flow q_{cap} at the critical density k_{c}. As a result of Kerner's fundamental hypothesis, the region of synchronised flow (denoted by S) covers a large two-dimensional part of the density-flow phase space. It is intersected by the line J, denoting the steady propagation of wide-moving jams. The line J also intersects the curve of free flow in the outflow from a jam $q_{\text{out}} \ll q_{\text{cap}}$ at the associated density k_{out}.

In the flow versus density diagram in Figure 2.12, the three regimes are depicted: the curve of free flow (denoted by F), the region of synchronised flow (denoted by S) and the wide-moving jam (denoted by the empirical line J). Just as in the classical fundamental diagrams, the observations in free-flow traffic lie on a sharp line that linearly increases the flow with higher densities (note the levelling of the curve near the capacity flow q_{cap} associated with the critical density k_{c}). The region of synchronised flow spans a large part of the density-flow phase space; an important remark here is that consecutive measurement points are scattered within this region, meaning that an increase in the flow can happen with both higher *and* lower densities (as opposed to the free-flow regime) [Ker96b].

The characteristic line J denotes the steady, undisturbed propagation of wide-moving jams. Its slope corresponds to the speed of a wide-moving jam's downstream front, which typically lies around $w \approx$ -15 km/h [Ker96a]. The upper-left point of the line J is located at a density k_{out} corresponding to the outflow $q_{\text{out}} \ll q_{\text{cap}}$ from a wide-moving jam. This is an illustration of the capacity drop phenomenon, elucidated in

Section 2.5.3. The line J is defined as follows:

$$q(k) = \frac{1}{T}\left(1 - \frac{k}{k_{\text{jam}}}\right), \tag{2.47}$$

with T the time gap in congested traffic flows; it is used to tune the outflow from a jam. Because wide-moving jams travel undisturbed, their outflow — caused by vehicles that leave the downstream front — can be either free flow or synchronised flow. Typical values for this outflow range from 1500 to 2000 vehicles/hour/lane [Ker04]. The average flow rate within such a wide-moving jam can be almost zero, meaning that vehicles continuously encounter stop-and-go waves.

Related to the wild scatter in the (k,q) diagram of three-phase traffic theory, is the microscopic behaviour of individual vehicles. The explanation given by Kerner and Klenov, is that vehicles in synchronised flow do not assume a fixed preferred distance to their direct frontal leader, but rather accept a certain *range of distances*. Within this range, drivers have both the tendency to over-accelerate when they think there is the ability to overtake, and the tendency for drivers to adjust their speed to that of their leader, when this overtaking can not be fulfilled [Ker03; Ker04].

2.5.4.3 Transitions towards a wide-moving jam

The breakdown of traffic from the free-flow to the wide-moving jam state, is nearly always characterised by two successive $F \rightarrow S$ and $S \rightarrow J$ transitions, between free flow and synchronised flow, and synchronised flow and wide-moving jam respectively. In the first stage, a state of free flow changes to synchronised flow by the $F \rightarrow S$ transition. Central to the idea of this phase transition, is the fact that there is no explicit need for an external disturbance for its occurrence. A sufficiently large (i.e., *supercritical*) internal disturbance inside the traffic stream (e.g., a lane change) causes a *nucleation effect* that instigates the $F \rightarrow S$ transition. Once it has set in, the onset of congestion is accompanied by a sharp drop in the mean vehicle speed. During the second stage, a set of narrow-moving jams can grow inside the tempo-spatial region of synchronised flow. A narrow-moving jam is different from a wide-moving jam, in that vehicles typically do not on average come to a full stop inside the jam. But, due to a compression of synchronised flow (an effect termed the *pinch effect*), these narrow-moving jams can coalesce into a wide-moving jam, thereby completing the cascade of the $F \rightarrow S \rightarrow J$ transition, resulting in stop-and-go traffic [Ker98].

With respect to the flow versus density diagram in Figure 2.12, it can be seen that the line J actually divides the region of synchronised flow in two parts. Points that lie underneath this line, characterise stable traffic states where no $S \rightarrow J$ transition can occur. Points above the line J however, characterise metastable traffic states, meaning that sufficiently large disturbances can trigger a $S \rightarrow J$ transition [Kno02a; Ker04].

Note that the direct $F \rightarrow J$ transition between free flow and wide-moving jam can also occur, but it has a very small probability, i.e., the critical perturbation needed, is much higher than that of the frequently occurring $F \to S$ transition between free flow and synchronised flow. So in general, wide-moving jams do not emerge spontaneously in free flow, but a situation where such a transition may occur, is when an off-ramp gets filled with slow-moving vehicles. This results in a local obstruction at the motorway's lane directly adjacent to the off-ramp, which can cause a local breakdown of the upstream traffic, resulting in a wide-moving jam. Finally, it is important to distinguish the nature of this transition from that of the $F \to S$ transition: the former is a transition *induced* by an *external disturbance* of the local traffic flow, whereas the latter is considered as a *spontaneous* transition due to an *internal disturbance* within the local traffic flow (e.g., a lane change) [Ker04].

2.5.4.4 From descriptions to simulations

As Kerner himself describes his three-phase theory, it is a *qualitative* theory. In essence, it gives no explanation of *why* certain transitions occur, as it only *describes* them [Ker04]. However, several exemplary microscopic traffic flow models have already been developed (i.e., treating all vehicles and their interactions individually). These models can reproduce the different empirical tempo-spatial patterns described by Kerner's theory. As examples, we mention two models based on cellular automata: a first attempt was made by Knospe et al., who developed a model that takes into account a driver's reaction to the brake-lights of his direct frontal leader [Kno00]. Kerner et al. refined this approach by extending it; their work resulted in a family of models based on the notion of a *synchronisation distance* for individual vehicles; they are commonly called the KKW-models (from its three authors, Kerner, Klenov, and Wolf) [Ker02].

The theory can describe most of the encountered tempo-spatial features of congested traffic. And at the moment, successful microscopic models have been developed, but the work is not yet over: an important challenge that remains for theoreticians, is the mathematical derivation of a consistent macroscopic theory (i.e., one that treats traffic at a more aggregate level as a continuum) [Ker04]. In pursuit of such a model, Kim incorporated Kerner's traffic regimes into a broader framework, encompassing six different possible states: the transitions between these states are tracked with a modified macroscopic model that uses concepts from fuzzy logic theory [Kim02].

2.5.5 Theories of traffic breakdown

A central question that is often asked in the field of traffic flow theory, is the following: *"What causes congestion ?"* Clearly, the answer to this question should be a bit more detailed than the obvious *"Because there are too many vehicles on the road !"* With respect to the phase transitions that signal a breakdown of the traffic flow, various — seemingly contradicting — theories exist. Are they merely a matter of belief, or can they be rigourously 'proven' ? Opinions are divided, but nowadays, two qualitatively different mainstream theories exist, attributed to different schools of thought [Bud00; Mae04c; Tam04a]:

The European (German) school
In the early seventies, Treiterer and Meyers performed some aerial observations of a platoon of vehicles. As they constructed individual vehicle trajectories, they could observe a growing instability in the stream of vehicles, leading to an apparently emerging *phantom jam* (i.e., a jam ‘out of nothing’) [Tre74].

Some twenty years later, in the mid-nineties, Kerner and Konhäuser made detailed studies of traffic flow measurements, obtained at various detector stations along German motorways. Their findings indicated that phantom jams seemed to emerge in regions of unstable traffic flow [Ker94]. This stimulated Kerner and Rehborn to further research efforts directed towards the behaviour of propagating jams [Ker96a; Ker96b]. They proposed a different set of traffic flow regimes, culminating in what is now called *three-phase traffic theory* (see Section 2.5.4 for more details) [Ker97; Ker98; Ker04]. The main idea supported by followers of this school of researchers, is that traffic jams *can* spontaneously emerge, without necessarily having an infrastructural reason (e.g., on-ramps, incidents, . . .)[14]. In dense enough traffic, phase transitions from the free-flow to the synchronised-flow regime can occur, after which a local instability such as, e.g., a lane change can grow (the so-called *pinch effect*), triggering a stable jam leading to stop-and-go behaviour [Ker98]. Kerner’s three-phase theory stands out as an archetypical example of these modern views. But although his theory has, in our opinion, been worked out well enough, he more than frequently encounters harsh criticisms when conveying it to most audiences (perhaps the main cause for this human behaviour is the fact that Kerner always mentions the same view, i.e., *“all existing traffic flow theories are wrong”*).

Inspired by Kerner’s work, Helbing et al. gave in 1999 an extended treatise on the different types of congestion patterns that can be observed in the vicinity of spatial inhomogeneities (e.g., on-ramps). Their work resulted in a universal phase diagram, containing a whole plethora of patterns of congested traffic states (called *homogeneously congested traffic* – HCT, *oscillatory congested traffic* – OCT, *triggered stop-and-go traffic* – TSG, *pinned localised cluster* – PLC, and *moving localised cluster* – MLC), each one having unique characteristics [Hel99a][15]. In that same year, Lee et al. studied the patterns that emerge at on-ramps, thereby agreeing with the findings of Helbing et al. [Lee99]. As the previous research into congestion patterns was largely based on the use of analytical traffic flow models and computer simulations, the need for validation with empirical data grew. In 2000, the work of Treiber et al. among others, proved the

[14] But note that bottleneck-induced traffic flow breakdowns are not excluded by the theory of Kerner et al.

[15] In addition, they also provided a link with Kerner’s three-phase theory, whereby synchronised flow can correspond to HCT, OCT, or PLC, and moving jams can correspond to TSG or MLC states [Hel02b].

existence of the previously mentioned congestion patterns [Tre00].

At this point, it is noteworthy to mention the seminal work of Nagel and Schreckenberg [Nag92b], who in 1992 developed a model that describes traffic flows in which local jams can form spontaneously. As many variations on this model have been proposed (see Chapter 4 for more details), later work also focussed on the stability of traffic flows in these models, e.g., the work of Jost and Nagel [Jos03a].

The Berkeley school

Including names such as the late Newell, Daganzo, Bertini, Cassidy, Muñoz, . . . , the 'Berkeley school' (University of California) supports the theory that all congestion is strictly induced by bottlenecks. The hypothesis holds for both recurrent and, in the case of an incident, non-recurrent congestion.

The main starting point states that there is *always* a 'geometrical' explanation for the breakdown. This explanation is based on the presence of road inhomogeneities such as on- and off-ramps, tunnels, weaving areas, lane drops, sharp bends, elevations, . . . Once a jam occurs due to such a (temporary) bottleneck, it does not dissipate immediately; as a result, drivers can wonder why they enter and exit a congestion wave, without there being an apparent reason for its presence (since it happened earlier and the cause, e.g., an incident, already got cleared). Daganzo uses this line of reasoning as an explanation for the dismissal of phantom jams [Dag02c].

The school uses a specific terminology with respect to bottlenecks (being road inhomogeneities). Two qualitatively different regimes exist: the *free-flow regime* and the *queued regime*. The latter occurs when a bottleneck becomes *active*, which will result in a queue growing upstream of the bottleneck while a free-flow regime exists downstream. The *bottleneck capacity* is then defined as the maximum *sustainable* flow downstream (which is different from the maximum flow that can be observed prior to the bottleneck's activation).

The location of these bottlenecks has some peculiarities involved: one of them is the concept of a *capacity funnel* [Buc74]. It assumes that drivers are at times more alert, e.g., when they are driving on a motorway and nearing an on-ramp in rather dense traffic conditions [Zha01b]. This impels them to accept shorter headways, so they are driving closely behind each other at a relatively high speed. Once they have passed the on-ramp's location, they tend to relax, resulting in larger headways. The effect is that the bottleneck's *actual* position is located more downstream.

Shortly after the publication of Kerner and Rehborn's findings about the peculiar phase transitions that seemed to occur on German motorways, Daganzo et al. provided a swift response where they stated that the occurring phase transitions *could* also be caused by bottlenecks in a predictable way [Dag99c]. They implied that no spontaneously emerging traffic jams are suggested, and that the observed traffic data from both German and North American motorways did not contradict their own statements about the cause of the phase transitions [Nag05]. Another study aimed at the data stemming from the German motorways, was performed by Lindgren who analysed bottleneck activation, discharge flows, …; he concluded that all features were reproducible from day to day [Lin05]. In short, the subtle difference between their work and that of Kerner and Rehborn, is that instabilities in the traffic stream are the *result* and not the cause of the queues that emerge at active bottlenecks. With respect to a spontaneous breakdown of traffic flow at on- and off-ramps (i.e., bottlenecks), Daganzo also states that this can be explained using a simple traffic flow model operating under the assumption of a too high inflow from the on-ramp or a caused by blocking of the off-ramp [Dag96].

The studies undertaken by this school, are heavily based on the researchers' use of cumulative plots and elegantly simple traffic flow models (as will be explained in Section 3.2.1.7), as opposed to the classical methodology that investigates time series of recorded counts and speeds. As stated earlier (see Section 2.3.2.2), some recent examples include the work of Muñoz and Daganzo [Muñ00a; Muñ00b; Muñ02a; Muñ03a] and Cassidy and Bertini [Cas99; Ber03].

Recently, Tampère argued that both theories, as enunciated by the two schools, are not entirely contradictory. His statement is based on the fact that the mechanisms behind the bottleneck-induced breakdown and spontaneous breakdown are approximately the same, only differing in the *probability* of such a breakdown (which is related to the instability of a traffic flow) [Tam04a].

> In our view, both theories are sufficiently different, *but compatible*, in that the first school elaborately describes traffic flow breakdown more or less as having an inherently *probabilistic nature*, whereas the second school treats breakdown a strictly *deterministic process*. The former introduces a complex variety of congestion patterns, while the latter primarily focusses on an elegantly simple description of traffic flow breakdown. Even more characteristically, is the observation that most adepts of the European school, *inherently need stochasticity in the models* in order to produce their sought phantom traffic jams (note that notwithstanding the fact that stochastic models are in a strict sense also deterministic, we nevertheless adopt in this dissertation, the convention that deterministic means 'non-stochastic'). Our argument is in a way also supported by Nagel and Nelson, who state that the purpose of the traffic flow model (e.g., the effect of moving bottlenecks versus predicting mean traffic behaviour) decides whether or not stochasticity in the model is required [Nag05]. Furthermore, there might be some room for stochasticity in the Berkeley models after all, with the work of Laval which suggests that (disruptive) lane changes form the main cause for instabilities in a traffic stream [Lav04]. Deciding which school is right, is therefore in our opinion a matter of personal taste, but in the end, we agree with Daganzo when he states that research into bottleneck behaviour is the most important in the context of traffic flow theory [Dag99b].

2.6 Conclusions

In this chapter, an extensive account was given, detailing several aspects related to the description of traffic flows. Most importantly, we have introduced a nomenclature convention, built upon a consistent set of notations. Our discussion of traffic flow characteristics centred around the space and time headways as microscopic characteristics, with densities and flows as their macroscopic counterparts. Several noteworthy highlights are the technique of oblique cumulative plots and the derivation of travel times based on these plots. A finally large part of this chapter reviewed some of the relations between traffic flow characteristics, i.e., the fundamental diagrams, and clarified some of the different points of view adopted by the traffic engineering community.

Whereas this chapter described the *properties* of real-life traffic flows, the next chapter will introduce some of the models that can be used to *simulate* the behaviour of traffic flows.

Chapter 3

Transportation planning and traffic flow models

Whereas the previous chapter dealt with the notations and terminology that are associated with traffic flow characteristics, this chapter focusses on the different traffic flow models that exist. Due to our frequently encountered confusion among traffic engineers and policy makers, this chapter goes into more detail about transportation planning models on the one hand, and traffic flow models on the other hand. The former deal with households that make certain decisions which lead to transportation and the use of infrastructure, as opposed to the latter which explicitly describe the physical propagation of traffic flows in a road network.

Our goal is not to give a full account (as that would be a dissertation of its own, given the broadness of the field), but rather to impose upon the reader a thorough feeling for the differences between transportation planning and traffic flow models. Because of the high course of progress over the last decade (or even during the last five years), this chapter tries to chronicle both past models, as well as some of the latest developments in this area.

3.1 Transportation planning models

Before going into detail about the possible mathematical models that describe the physical propagation of traffic flows, it is worthwhile to cast a glance at a higher level, where transportation planning models operate. The main rationale behind transportation planning systems, is that travellers within these systems are motivated by making certain decisions about their wishes to participate in social, economical, and cultural activities. The ensemble of these activities is called the *activity system*. Because these activities are spatially separated (e.g., a person's living versus work area), the need for transportation arises. In such a system, the so-called *household activity patterns* form the main explanation for what is seen in the transportation network.

These models have as their primary intent the performing of impact and evaluation studies, and conducting 'before and after' analyses. The fact that such transportation studies are necessary, follows from a counter-intuitive example whereby improving the transportation system (e.g., by making extra infrastructure available), can result in an *increase* of the travel times. This phenomenon, i.e., allowing more flexible routing that results in more congestion, is known as Braeß' paradox, after Dietrich Braeß [Bra69]. The underlying reason for this counter-intuitive behaviour, is that people generally only *selfishly* try to minimise their own travel times, instead of considering the effects they have on other people's travel times as well [Pas97].

As transportation is inherently a temporal and spatial phenomenon, we first take a look at the concept of land-use models and their relation to the socio-economical behaviour of individual people. In the two subsequent sections, we consider two types of transportation planning models, i.e., the classical trip-based models, and the class of activity-based models, respectively. The section concludes with a brief reflection on the economist's view on transportation systems.

3.1.1 Land use and socio-economical behaviour

As already stated, transportation demand arises because of the desire to participate in a set of activities (e.g., social, economical, cultural, ...). In order to deduce this *derived* transportation demand, it is necessary to map the activity system and its spatial separations. This process is commonly referred to as *land use*, mainly playing the role of forging a relation between economical and geographical sciences. In general, land-use models seek to explain the growth and layout of urban areas (which is not strictly determined by economical activities alone, i.e., ethnic considerations et cetera can be taken into account),

Because transportation has spatial interactions with land use and vice versa, it can lead to a kind of chicken-and-egg problem [Rod05]. For example, building a new road will attract some economical activity (e.g., shopping malls et cetera), which can lead to a possible increase of the travel demand. This in turn, can lead to an increase of extra economical activity (because of the well-suited location), and so on, resulting in a local reorganisation of the spatial structure. Resolving this chicken-and-egg paradox, is typically done by means of feedback and iterations between land-use and transportation models, whereby the former provide the basic starting conditions for the latter models (with sometimes a reversal of the models' roles).

In the following two sections, we first shed some light on several of the archetypical land-use models, after which we take a look at some of the more modern models for land use in the context of geosimulation.

3.1.1.1 Classical land-use models

The discussion given in this section, talks about several kinds of land-use models that — at their time — were considered as landmark studies. That said, the models presented here should be judged as being general in that they deal with (pre-)industrial American societies in the first part of the 20th century. They are devised to gain insight into the general patterns that govern the growth and evolution of a city. As such, they almost never 'fit' perfectly, leading to the obvious criticism that they are more applicable to American cities than elsewhere. Notwithstanding these objections, the models remain very useful as explanations for the mechanisms underpinning the socio-economical development of cities.

One of the oldest known models describing the relation between economic markets and spatial distances, is that of Johann Heinrich von Thünen [Thü26]. As the model was published in 1826, it presents a rather 'pre-industrial' approach: the main economical ingredients are based on agricultural goods (e.g., tomatoes, apples, wheat, ...), whereas the transportation system is composed of roads on which carts pulled by horses, mules, or oxen ride. The spatial layout of the model, assumes an *isolated state* (self-sustaining and free of external influences), in which a central city location is surrounded by concentric regions of respectively farmers, wilderness, field crops, and meadows for grazing animals. All farmers aim for maximum profits, with transportation costs proportionally with distance, thus determining the land use around the city centre.

Some 100 years later, inspired by von Thünen's simple and elegant model, Ernest W. Burgess developed what is known as the *concentric zone model* [Bur25]. It was based on observations of the city of Chicago at the beginning of the 20th century. As can be seen in the left part of Figure 3.1, Burgess considered the city as growing around a central business district (CBD), with concentric zones of respectively the industrial factories and the low-, middle-, and high-class residents. The outermost ring denotes the commuter zone, connecting the CBD with other cities. As time progresses, the city develops and the radii of these concentric zones would grow by processes of 'invasion' and 'succession': an inner ring will expand, invading an outer ring that in turn has to grow, in order to make space.

Fifteen years after Burgess' theory, Homer Hoyt introduced refinements, resulting in the *sector model* [Hoy39]. One of the main incentives, was the observation that low-income residents were typically located in the vicinity of railroads. His model accommodates this kind of observation, in that it assumes that a city expands around major transportation lines, resulting in wedge-shaped patterns (i.e., sectors), stretching outward from the CBD. A typical example of this development, can be seen in the right part of Figure 3.1.

Halfway the previous century, Chauncy D. Harris and Edward L. Ullman were convinced that the previous types of models did not correspond to many of the encountered cities. The main reason for this discrepancy, was to be found in the stringent condition

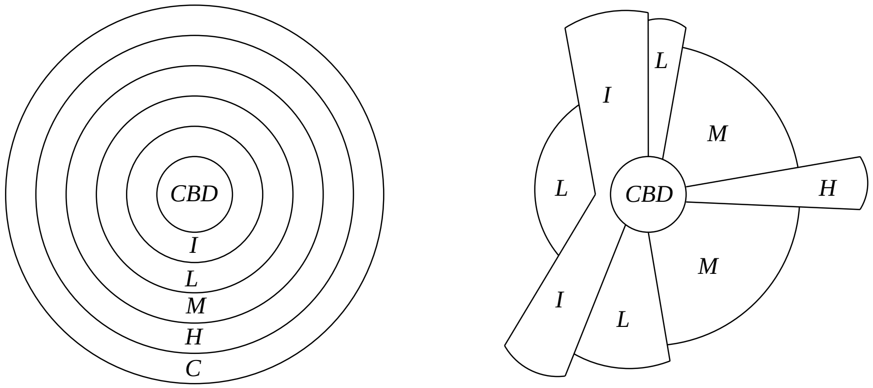


Figure 3.1: Typical examples of two models relaying the evolution of land use. *Left:* the concentric zone model of Burgess. *Right:* the sector model of Hoyt. In both figures, *CBD* corresponds to the central business district, *I* to the industrial factories, *L*, *M*, and *H* to the low-, middle-, and high-class residents respectively. In the Burgess model at the left, *C* denotes the commuter zone.

of a central area being surrounded by different zones. As a solution to this shortcoming, Harris and Ullman presented their *multiple nuclei model* [Har45]. Their theory assumed that in larger cities, small suburban areas could develop into fully fledged business districts. And although Harris and Ullman did not dispose of the CBD as the most important city centre, their smaller 'nuclei' would take on roles of being areas for specialised socio-economical activities.

To end our discussion of classical land-use models, we highlight the work of Peter Mann in 1965 [Man65], who considered a *hybrid model* for land-use representation. He combined both Burgess's and Hoyt's models, when deriving a model that described a typical British city. In his model that studied the cities of Huddersfield, Nottingham, and Sheffield, the CBD still remained the central location, surrounded by zones of pre- and post 1918 housing respectively. Dispersed around the outer concentric zone, the low-, middle-, and high-class residents would live. A most notable feature of Mann's model, is the fact that he considered the industrial factories to be on one side of the city, with the high-class residents diametrically opposed (the rationale being that high-class residents would prefer to stay upwind of the factories' smoke plumes).

3.1.1.2 The modern approach to land-use models

In the current time of living, most modern citizens have a different behaviour than their former counterparts at the beginning of the 20$^{\text{th}}$ century. It seems there is an increased trend towards expansion, as people are feeling more comfortable about covering larger distances, e.g., working in a busy city centre or at a remote industrial facility, coupled with living on the countryside). The activities related to working, living, and recreation appear to occur at substantially different spatial locations. Furthermore, several urban regions are composed of unique ethnic concentrations, among other things leading to the conclusion that the emphasis on the geographical aspect of a city gets less important during its evolution.

Recognising these radical changes in the development, modern land-use models approach the integration of an activity system from a completely different perspective. The growth of a city is represented as the evolution of a *multi-agent system*, in which a whole population of individual households is simulated. Due to the tremendous increase in computational power over the last two decades, these large-scale simulations are now possible. As an example, it is feasible to consider residential segregation in urban environments: within these environments (e.g., the city and housing market), individual agents (i.e., households) interact locally in a well-defined manner, leading to emergent structures, i.e., the evolving city. Besides data surveys that try to capture the households' behaviour, the basic landscape and mapping data is fed into *geographical information systems* (GIS) that is coupled with a *computer aided design* (CAD) representational model of the real world (although the difference between the traditional GIS and CAD concepts is slowly fading away) [Wad04]. A recent example of such an all-encompassing approach, is the work related to the UrbanSim project, where researchers try to interface existing travel models with new land use forecasting and analysis capabilities [Wad02]. It is being developed and improved by the Center for Urban Simulation and Policy Analysis at the University of Washington.

To conclude this section, we refer to the work of Benenson and Torrens, who adopted the terminology of *geosimulation* [Ben04]. Their methodology is based on what they call the 'collective dynamics of interacting objects'. As such, geosimulation hinges on the representation of what we would call a *socio-economy* that is simulated, taking into account hitherto neglected dynamic effects (e.g., demographic changes, shifts of the economic activities, . . .).

3.1.2 Trip-based transportation models

The relation between activity patterns and the transportation system has a long history, starting around 1954 with the seminal work of Robert B. Mitchell and Chester Rapkin [Mit54]. They provided the first integrated study, establishing a link that introduced a framework for transportation analysis, primarily intended for studying large scale infrastructure projects [McN00b]. Their methodology was based on four consecutive steps (i.e., submodels), collectively called the *four step model* (4SM). In 1979, Manheim casted the model's structure into a larger framework of transportation systems analysis, encapsulating both activity and transportation systems [Man79]. Central to this framework, was the notion of 'demand and performance procedures', which we can validly call *demand* and *supply procedures*. In a typical setup, they respectively represent the traffic that wants to use this infrastructure and the road infrastructure. For a more historically tinted recollection of the trip-based approach, we refer the reader to the outstanding overview of Boyce [Boy04a].

> With respect the 4SM's history, a subtle — almost forgotten — fact is that the classical four step model was actually conceived independently from the integrated network equilibrium model proposed by Beckmann, McGuire, and Winsten in the mid-fifties; the 4SM can actually be perceived as a trimmed-down version of this latter model [Boy04a]. Intriguingly, over the years, the work of the 'BMW trio' has had profound impacts on the mathematical aspects of determining network equilibria, optimal toll policies, algorithms for variational inequalities, stability analyses, supply chains, ... [Alt03; Nag03b; Boy04a; Boy04b; Boy06].

In the next four sections, we consider the basic entities and assumptions of the four step model, followed by a brief overview of the four individual submodels with some more detail on the fourth step (traffic assignment), concluding with some remarks on the criticisms often expressed against the four step model. For a more extensive survey of the four step model, we refer the reader to the books of Sheffi [She85] and Ortuzar and Willumsen [Ort01].

3.1.2.1 Basic entities and assumptions

The basic ingredients on which the four step model is rooted, are the *trips*. These trips are typically considered at the household level, and relate to aggregate information (individuals are no longer explicitly considered). This level of detail, essentially collapses the whole tempo-spatial structure of transportation planning based on individual travellers into bundles of trips, going from one point in the transportation network to another.

In the four step model, one of the most rigid assumptions is that all trips describe departure and arrival within the planning period (e.g., the morning commute). Furthermore, the usage of the model's structure is intended for large-scale planning purposes, excluding small infrastructural studies at, e.g., a single intersection of urban roads. Another assumption is based on the fact that an entity within the four step model has to make certain decisions, e.g., what is the departure time, which destination is picked, what kind of transportation (private or public) will be used, which route will be followed, ... In many cases, these decisions are considered concurrently, but the four step model assumes they are made independently of each other. And finally, as each submodel needs input, most of the data is aggregated into *spatial zones* (often presumed to be distinguished by socio-economic characteristics) in order to make the model computationally feasible. These zones are typically represented by their centrally located points, called *centroids*.

3.1.2.2 The four steps

Within the four step model, the first three steps (I) – (III) can collectively be seen as a methodology for setting up the travel demand, based, e.g., on land use and other socio-economical activities. This travel demand is expressed as *origin-destination* (OD) pairs (by some respectively called 'sources' and 'sinks'), reflecting the amount of traffic that *wants* to travel from a certain origin to a certain destination (these are typically the zones mentioned in the previous section). The last step (IV) then consists of loading this travel demand onto the network, thereby assigning the *routes* that correspond to the trips.

(I) Trip generation
In an essential first step, transportation engineers look at all the trips that on the one hand originate in certain zones, and on the other hand arrive in these zones. As such, the first step comprises what are called the *productions* and *attractions*. Central to the notion of a trip, is the *motive* that instigated the trip. An example of such a motive is a home-based work trip, i.e., a trip that originates in a household's residential area, and arrives in that household's work area. Other examples include recreational and social motives, shopping, … and the *chaining of activities*. Based on these intentions, productions and attractions consist of absolute counts, denoting the number of trips that depart from and arrive in each zone. Because of this, productions and attractions are in fact *trip ends*. Both of them are derived using techniques based on regression analysis, category analysis, or even logit models. As different models can be used for the derivations of the number of productions and attractions, an a posteriori balancing is performed that equalises both results. In the end, step (I) gives the magnitude of the total travel demand on the network. Note that all activities (i.e., the original motives) are at this point in effect transformed and *aggregated* into trips. More importantly, these trips are *only considered for a specific time* period (e.g., the morning rush hour).

(II) Trip distribution
Once the total number of productions and attractions for all zones in the transportation network is known, the next step then consists of deriving how many trips, originating in a certain zone, arrive at another zone. In other words, step (II) *connects* trip *origins* to their *destinations* by distributing the trips. The result of step (II) is then the construction of a complete *origin-destination* table (*OD table*). In such an OD table (or *OD matrix* as some people say), an element at a row i and a column j denotes the total number of trips departing from origin zone O_i and arriving in destination zone D_j. Diagonal elements denote intra-zonal trips. Note that step (II) does not state anything about the different *routes* that can be taken between two such zones; this is something that is derived in the final step (IV). Because of the implicit assumption in step (I), namely that all trips are considered for a specific time period, the same premise holds for all the derived OD tables. Consequently, the four step model is applied for different time periods, e.g., during rush hours or off-peak periods. In this context, we advise to use the nomenclature of *time-dependent* or *dynamic* OD tables, denoting OD tables that are specified for a certain period, e.g., from 07:00 until 08:00 (or even tables given for consecutive quarter-hours).

Considering the fact that an OD table contains a large amount of unknown variables (i.e., generating them from known link flows entails a considerably under determined system of equations, as there are more unknowns than constraints), several techniques have been introduced to deal with this problem by introducing additional constraints.

If an OD table for a previous period (called a *base table*) is known, then a new OD table can be derived by using a so-called *growth factor model*. Another method is by using *gravity models* (also known as *entropy models*, see, e.g., the discussion in by Helbing and Nagel [Hel04]), which are based on *travel impedance functions*. These functions reflect the relative attractiveness of a certain trip, e.g., based on information retrieved from household travel surveys. In most cases, they are calibrated as power or exponential functions. One of the harder problems that still remains to be solved, is how to deal with so-called *through trips*, i.e., trips that originate or end outside of the study area. Horowitz and Patel for example, directly incorporate rudimentary geographical information and measured link flows into a model that allows to derive through-trip tables, using a notion of external stations located in an external territory. Application of his methodology to regions in Wisonsin and Florida, result in reasonable estimates of link flows that are comparable with empirically obtained data [Hor99].

Besides using results from productions and attractions, gathering the necessary information for construction of OD tables can also be done using other techniques. An equivalent methodology is based on the consideration of *turning fractions* at intersections. The process can be largely automated when using video cameras coupled with image recognition software. Furthermore, there literally exist thousands of papers devoted to the estimation of origin-destination matrices, mostly applicable to small-scale vehicular transportation networks and local road intersections. Some past methodologies used are the work of Nihan and Davis who developed a recursive estimation scheme [Nih87], the review Cascetta and Nguyen who casted most earlier methods into a unified framework [Cas88], and Bell who estimated OD tables based on constrained generalised least squares [Bel91]. An example of a more recent technique is the work of Li and De Moor who deal with incomplete observations [Li02].

(III) Mode choice / modal split

Once the origin-destination table for the given network and time period is available, the next step deals with the different *modes of transportation* that people choose between. Typical examples are the distinction between private and public transportation (both vehicular and railroad traffic). The ‘split’ in this step, refers to the fact that the OD table obtained from step (II), is now divided over the supported transportation modes. To this end, *discrete choice theory* is a popular tool that allows a disaggregation based on the choice of individual travellers, e.g., by using utility theory based on a nested logit model [BA85]. A modern trend in this context is to work with fully *multi-modal transportation networks*; these multi-layered networks provide access points for changing from one layer (i.e., mode of transportation) to another [Nes02; Car05].

Historically, steps (I) — production and attraction — and (III) were executed simultaneously, but nowadays they are considered separate from step (I): the main reason is the fact that the modal choice is not only dependent on, e.g., a household’s income, but also on the type of trip to be undertaken, as well as the trip’s destination. As a result, the modal split can be intertwined with step (II), trip distribution, or it can be executed subsequently after step (II). In the former case, the same kind of travel impedance functions are used in combination with an adjusted gravity model, whereas in the latter case, a hierarchic logit model can be used.

(IV) Traffic assignment

At this point in the four step model, the total amount of trips undertaken by the travellers is known. The fourth and final step then consists of finding out *which routes* these travellers follow when going from their origins to their destinations, i.e., which sequence of consecutive links they will follow ? In a more general setting, this process is known as *traffic assignment*, because now the total travel demand (i.e., the trips) are assigned to routes in the transportation network. Note that in some approaches, an iteration is done between the four steps, e.g, using the traffic assignment procedure to calculate link travel times that are fed back as input to steps (II) and (III).

It stands to reason that all travellers will endeavour to take the *shortest route* between their respective origins and destinations. To this end, a suitable measure of distance should be defined, after which a shortest path algorithm, e.g., Dijkstra's algorithm [Dij59], can calculate the possible routes. Such a notion of distance typically includes both spatial and temporal components, e.g., the physical length of an individual link and the travel time on this link, respectively. The use of the travel time is one of the most essential and tangible components in travellers' route choice behaviour. Note that in a more general setting, the distance can be considered as a *cost*, whereby travellers then choose the *cheapest route* (i.e., the quickest route when time is interpreted as a cost). Daganzo calls these formulations the *forward shortest path problem*, as opposed to the *backward shortest path problem* that tries to find the cheapest route for a given arrival time [Dag02d].

The basic principles that underlay route choice behaviour of individual travellers, were developed by Wardrop in 1952, and are still used today. In his famous paper, relating space- to time-mean speed, Wardrop also stated two possible criteria governing the distribution of traffic over alternative routes [War52]:

> **User equilibrium (W1):** *"The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route."*
>
> **System optimum (W2):** *"The average journey time is a minimum."*

The above two criteria are based on what is called the *Nash equilibrium* in game theory [Nas51], albeit that now a very large number of individuals are considered[1]. In the first criterion (W1), it is assumed that all individuals' decisions have a negligible

[1] The difference between a Wardrop and a Nash equilibrium is a subtle but important one. In the Wardrop case, an infinite number of individuals is considered, each seeking their own optimum. Note that the concept 'infinite number of individuals' can in practice be approximated by 'a large amount of individuals'. The Nash case also considers an infinite number of individuals, but they are now grouped into a finite number of classes, with each class seeking its own optimum. If in this latter case the number of classes goes to infinity, then the Nash equilibrium converges to a Wardrop equilibrium [Hau85].

effect on the performance of others. Two, more important, fundamental principles here are the fact that in the equilibrium situation, there is *no cooperation between individuals* assumed, and that all individuals make their decisions in an *egoistic and rational* way [Hag01]. In both cases, all individuals are on equal footing; an exception to this is, e.g., the presence of a centralised authority that acts as a leading player, resulting in a *Stackelberg equilibrium* in game theory. In real-life traffic, everybody is expected to follow the first criterion (W1), such that the whole system can settle in an equilibrium in which no one is better off by choosing an alternative route. In this respect, the work of Roughgarden on *selfish routing* is interesting because it provides a mathematical basis for the quantification of the worst loss of social welfare due to the selfish behaviour of drivers when choosing routes[2]. In continuation of this, Roughgarden also considered the design and management of networks that limit these effects in order to obtain a socially desirable outcome [Rou02]. In contrast to this user equilibrium situation, the second criterion (W2) is unlikely to occur spontaneously. However, when the perceived cost of a route by a traveller is changed to a *generalised* or *marginal cost* (i.e., including the costs of the effects brought on by adding an extra vehicle to the travel demand), then a system optimum is achieved with respect to these latter costs. In any case, as some people will be better off, others will be worse off, but the transportation system as a whole will be best off.

The above two principles, are a bit idealistic, in the sense that there are many exceptions to these behavioural guidelines. For example, in urban city centres, a significant part of the congestion can be brought on by vehicles looking for parking space. Furthermore, many drivers just follow their *usual route*, because this is the route they know best, and they know what to expect with respect to travel time. In a broader setting, this make these 'standard' routes more *appealing* to road users than other unfamiliar alternative routes. In some cases however, travellers will opt for these less known routes, thereby possibly entering the *risk* of experiencing a higher travel time as has been concluded in the work of Chen and Recker [Che01a]. Another fact that we expect to have a non-negligible effect on the distribution of traffic flows, is that nowadays more and more people use *intelligent route planners* to reach their destinations. These planners take into account congestion effects, as the trip gets planned both spatially and temporally. This will result in a certain percentage of the population that is informed either *pre-route* or *en-route*, and these people can consequently change their departure time or actual route (e.g., through route guidance), respectively. Another interesting research problem arises because transportation infrastructure managers should then be able to adapt their policies to the changing travel patterns. For example, how should a policy maker optimally control the traffic when only 20% of the population will follow the proposed route guidance ?

Due to the importance of the subject, we have devoted two separate sections in this dissertation to the concept of traffic assignment. In these sections, we discuss the traffic assignment procedure in a bit more detail, considering two prominent methodologies from a historic perspective, namely *static* versus *dynamic* traffic assignment.

[2]In his work, Roughgarden also introduced *the price of anarchy*, which is defined as the total travel time at the user equilibrium divided by the total travel time at the system optimum [Rou03].

3.1.2.3 Static traffic assignment

The classical approach for assigning traffic to a transportation network, assumes that all traffic flows on the network are *in equilibrium*. In this context, the *static traffic assignment* (STA) procedure can be more correctly considered as dealing with *stationary* of *steady-state* flows: the travel demand and road infrastructure (i.e., the supply) are supposed to be time-independent, meaning that the calculated link flows are the result of a constant demand. In a typical setup, this entails the assignment of an hourly (or even daily) OD table to the network (e.g., during on- and off-peak periods), resulting in *average* flows for the specified observation period. Because the STA methodology neglects time varying congestion effects (it assumes constant link flows and travel times), various important phenomena such as queue spill back effects are not taken into account.

In general, several possible techniques exist for achieving an STA. The first one assumes (i) that all drivers will choose the same cheapest route between a pair of origins and destinations, (ii) that they all have the same *perfect information* about the links' impedances, and (iii) that these impedances are considered to be constant, i.e., independent of a link's traffic load (so no congestion buildup is taken into account). As the methodology implies, this is called an *all-or-nothing assignment* (AON). A second technique refines this notion, whereby differences among drivers are introduced (i.e., giving rise to imperfect information), resulting in a *stochastic assignment*. In this methodology, the link travel impedances are assumed to be probabilistically distributed: for each link in the network, an impedance is drawn from the distribution after which an AON assignment is performed on the resulting network. This Monte Carlo process is repeated until a certain termination criterion is met.

Both previous methods carry a significant drawback with respect to link capacities, that is to say, no effects are taken into account due to the fact that an increased flow on a link will generally result in an increase of the travel time (i.e., the link's impedance). To this end, a third method introduces *capacity restraints* such that an increase of the travel demand on a link, will result in a higher cost (thereby possibly changing the route with the cheapest cost). This method is called an *equilibrium assignment*, and just like as in the second method, a *stochastic equilibrium assignment* version can be derived, taking into account travellers' imperfect knowledge. The underlying assumption is that all travellers behave according to Wardrop's user equilibrium (W1). Furthermore, the capacity restraints are included in the travel impedance functions, as they are now synonymously called *travel time (loss) functions*, *congestion functions*, *volume delay functions*, *link impedance functions*, *link cost functions*, or even *link performance functions*. A popular form of these functions that express the travel time T in function of the flow q on a link, is the *Bureau of Public Roads* (BPR) power function[3] [BPR64]:

$$T = T_{\text{ff}} \left(1 + \alpha \left(\frac{q}{q_{\text{pc}}} \right)^{\beta} \right). \tag{3.1}$$

In this BPR relation, the coefficients α and β determine the shape of the function. An example of such a function is depicted in Figure 3.2. For low flows, the BPR function is rather flat and the travel time corresponds to the travel time T_{ff} under free-flow conditions. When higher flows occur on the link, the coefficient β determines the threshold at which the BPR function rises significantly (in some formulations it *asymptotically* approaches the capacity flow). The travel time will increase with the ratio of the flow q and the so-called *practical capacity* q_{pc}. This latter characteristic is derived from the value of the travel time under congested conditions. As a result, the practical capacity is different from the maximum capacity of a link as defined by a fundamental diagram. Finally, note that a serious disadvantage associated with these BPR functions in combination with static traffic assignment, is the fact that the travel demand on the network at a certain time does not always correspond to the actual physical flows that can be sustained. Under congested conditions, this implies that the flows in the STA approach can be higher than the physically possible link capacities (which are different from the previously mentioned practical capacities), leading to an incorrect assignment with faulty oversaturated links.

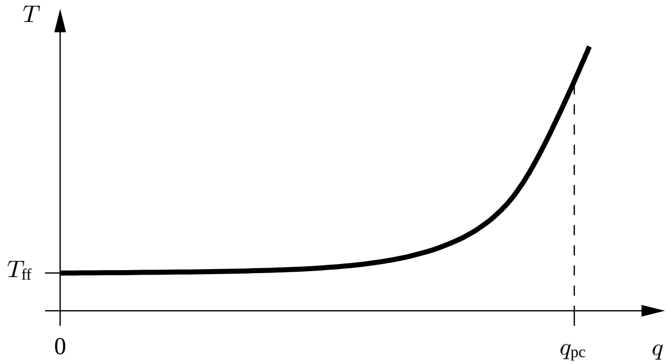


Figure 3.2: The Bureau of Public Roads (BPR) function, relating the travel time to the flow. It is based on the travel time T_{ff} under free-flow conditions and the practical capacity q_{pc} of the link under consideration.

Once the travel time of a link can be related to its current flow using, e.g., a BPR function, an iterative scheme is adopted to calculate the equilibrium traffic assignment.

[3] Other possible forms for the travel time function are a *linear delay function* [Fra99], a *Kleinrock function* which has a vertical asymptote [Kle72], a *conical function* which has, in contrast to the BPR function, smaller travel times above the capacity flow [Spi90], and the *Davidson's function* which has its roots in queueing theory and is similar to the Kleinrock function [Dav66].

Popular implementations are the *Frank-Wolfe algorithm* [Fra56] and the *method of successive averages* (MSA) [Smo62]. The former method is based on principles of optimisation theory, as demonstrated by Beckmann et al. [Bec55; Boy04b] who reformulated the Wardrop equilibrium as a convex optimisation problem, consisting of a single objective function with linear inequality constraints based on the Karush-Kuhn-Tucker (KKT) conditions, thereby resulting in a global minimum. Because travellers in general do not have perfect information, Daganzo and Sheffi formulated a variation on Wardrop's first criterion (W1), whereby all traffic distributes itself over the network with respect to a *perceived travel time* of the individual drivers [Dag77]. The resulting state of flows on the network is called a *stochastic user equilibrium* (SUE), as opposed to the *deterministic user equilibrium* (DUE)[4]. Note that a further discrimination is also possible, as proposed by the work of Chen and Recker, who make a distinction between travellers' perception errors on the one hand, and network uncertainty (i.e., stochasticity of the travel times) on the other hand [Che01a]. For a thorough overview of the STA approach, we refer the reader to the work of Patriksson [Pat94].

Although, as mentioned earlier, time varying congestion effects are not taken into account, the STA approach does fit nicely into the concept of long-term transportation planning. For short-term analyses however, these effects can have a significant impact on the end results, thus requiring a more detailed approach to traffic assignment.

3.1.2.4 Dynamic traffic assignment

As explained in the previous paragraphs, the static traffic assignment heavily relies on simple travel time functions (e.g., BPR). An associated problem with these is the difficulty in capturing the concept of 'capacity of a road'. In reality, congestion is a dynamic phenomenon, whereby its temporal character is not to be neglected. To tackle these problems inherent to the STA approach, a more dynamic treatment of traffic assignment is necessary [Mae04b]. A fundamentally important aspect in this *dynamic traffic assignment* (DTA) procedure, is the fact that congestion has a temporal character, meaning that its buildup and dissolution play an important role: the *history of the transportation system* should be taken into account (e.g., congestion that occurs due to queue spill back) [Dag95c]. Neglecting this time dependency by assuming that the entry of a vehicle to a link instantaneously changes the flow on that link, results in what is called *Smeed's paradox*. This leads to incorrect behaviour as a result of a violation of the FIFO property (see Section 2.2.2), because now a vehicle can leave link earlier then a vehicle that enters it later (i.e., arriving earlier by departing later) [Sme67]. The methodology of dynamic traffic assignment was now designed to deal with all these particular aspects. The DTA technique is composed of two fundamental components:

Route choice
Just as in the STA approach, each traveller in the transportation network

[4] Note that in some contexts, DUE is also used to denote a *dynamic user equilibrium* [Sze04; Han06].

follows a certain route based on an instinctive criterion such as, e.g., Wardrop's (W1). The associated component that takes care of travellers' route choices, can be complemented to allow for imperfect information. Another, more important, aspect related to the route choice, is a traveller's *choice of departure time*. An STA approach assumes that all traffic of a given OD table is simultaneously assigned to the network, whereas DTA coupled with departure time choice can spread the departures in time (leading to, e.g., realistically spreading of the morning peak's rush hour).

Dynamic network loading (DNL)
Instead of using simple travel time functions, a DTA approach typically has a component that loads the traffic onto the network. In fact, this step resembles the *physical propagation* of all traffic in the network. In order to achieve reliable and credible results, a good description of the network's links is necessary, as well as the behaviour of traffic at the nodes connecting the links within this network (i.e., this is a mandatory requirement to achieve correct modelling of queue spill back). The DNL component in the DTA approach has been an active field of research during the last decade, and it still continues to improve the state-of-the-art. Testimonies include the use of *analytic models* that give correct representations of queueing behaviour, as well as detailed simulations that describe the propagation of individual vehicles in the transportation network. Note that in the case of *simulation-based* (also called *heuristic*) traffic assignment, the route choice and DNL components can be iteratively executed, whereby the former establishes a set of routes to follow, and the latter step feedbacks information to the route choice model until a certain termination criterion is met (e.g., a relaxation procedure). Furthermore, using simulation-based traffic assignment with very large road networks is not always computationally feasible to calculate all shortest paths. As a result, it might be beneficial to resort to simplifications of either the simulation model (e.g., using faster queueing models), or the number of paths to consider (e.g., based on the hierarchy inherently present in the road network) [Ros01]. Finally, we mention the work of Astarita who provides an interesting classification of DNL models, based on the discretisation with respect to the spatial and temporal dimensions, as well as with respect to the modelling of the traffic demand [Ast02].

Despite the appealing nature of simulation-based DTA, there is in contrast to the STA approach, no unified framework that deals with the convergence and stability issues [Gaw98a; Gaw98b; Pee03].

Some examples of these DTA mechanisms are: Gawron who uses a queueing model to develop a simulation-based assignment technique that is able to deal with large-scale networks and is proven to be empirically stable [Gaw98b; Gaw98a], Bliemer who developed a macroscopic analytical DTA approach (with different user-classes) based on a variational inequality approach [Bli01], Bliemer's work furthermore culminated in the development of *INteractive DYnamic traffic assignment* (INDY) [Mal03; Bli04] which — in combination with the OmniTRANS[5] commercial transportation planning software — can be used as a fully fledged DTA analysis tool [Ver03a], Lo and Szeto who developed a DTA formulation based on a variational inequality approach leading to a dynamic user equilibrium [Lo02], the group of Mahmassani who is actively engaged in the DTA scene with the development of the DYNASMART (*DYnamic Network Assignment-Simulation Model for Advanced Roadway Telematics*) simulation suite [DYN03], … An excellent comprehensive overview of several traditional DTA techniques is given by Peeta and Ziliaskopoulos [Pee01].

Another important field of research, is how individual road travellers react to the route guidance they are given. In his research, Bottom considered this type of *dynamic traffic management* (DTM), providing route guidance to travellers whilst taking into account their *anticipated behaviour* during, e.g., incidents [Bot00]. Taking this idea one step further, it is possible to study the interactions between the behaviour of travellers in a road network, and the management of all the traffic controls (e.g., traffic signal lights) within this network. An example of such a *dynamic traffic control* (DTC) and DTA framework, is the work of Chen who considers the management from a theoretic perspective based on a non-cooperative game between road users and the traffic authority [Che98].

3.1.2.5 Critique on trip-based approaches

Considering its obvious track record of the past several decades, the *conventional* use of the trip-based approach is — to our feeling — running on its last legs. By 'conventional' we denote here the fact that the current *state-of-the-practice* is still firmly based on the paradigm of static traffic assignment, despite the recent (academic) progress on the front of dynamic traffic assignment techniques. The four step model still largely dominates the commercial business of transportation planning, although its structure remained largely unchanged since its original inception. As mentioned earlier, in the case of STA, all trips are assumed to depart and arrive within the specified planning period. This leads to an unnatural discrepancy between models and reality in congested areas during, e.g., a morning rush hour: some travellers *want* to make a trip and, in the former case, are perfectly allowed to achieve this trip, whereas in the latter case they are in fact physically *unable* to make the trip due to dynamical congestion effects.

In order to facilitate this disagreement between the balancing of travel demand versus supply (i.e., the transportation infrastructure), the DTA approach is gaining importance as more features are provided. An example of such a feature includes the framework of congestion pricing, where we have an incorporation of departure time choice models coupled with the derivation of optimal road tolls. Some noteworthy studies that have been carried out in this respect, are the work of de Palma and Marchal

[5] http://www.omnitrans-international.com

who present the METROPOLIS toolbox, allowing the simulation[6] of large-scale transportation networks [Pal02; Mar03], the work of Lago and Daganzo who combined a departure time equilibrium theory with a fluid-dynamic model in order to assess congestion policy measures [Lag03b], the work of Szeto and Lo who coupled route choice and departure time choice with the goal of numerically handling large-scale transportation networks [Sze04; Lo04]. Closely related to Lago's and Daganzo's work is that of Yperman et al., who determined an optimal pricing policy, describing the demand side with a bottleneck model and an analytical fluid-dynamic model as the DNL component [Ype05a].

At this point, we should mention some of the complications associated with the traditional method of modelling traffic flow propagation using queue-based analogies. Historically, there have been two different queueing techniques with FIFO discipline that describe this aspect in a trip-based assignment procedure:

- Point-queue models (PQM, also called *vertical queues*): this type of queue has no spatial extent. Because vehicles can *always* enter the queue, and leave it after a certain delay time, congestion is incorrectly modelled. A well-known example of a model based on this queueing policy is Vickrey's bottleneck model [Vic69].
- Spatial/physical-queue models (SQM, also called *horizontal queues*): a queue of this type has an associated *finite capacity*, i.e., a buffer storage. Vehicles can only enter the queue when there is enough space for them available.

The correct modelling of congestion effects such as queue spill back, is of fundamental importance when assessing certain policy measures like, e.g., road pricing schemes. To this end, the use of vertical queues should be abandoned, in favour of horizontal queues. However, even horizontal queues have problems associated with them: the buildup and dissolution of congestion in a transportation network are flawed, e.g., vehicles that are leaving the front of a queue *instantly open up a space at the back of this queue*, thus allowing an upstream vehicle to enter. This leads to shorter queue lengths, because the physical queueing effect of individual vehicles (i.e., the upstream propagation of the empty spot) is absent [Gaw98a; Sim99; Cet03; Lag03b]. In order to alleviate this latter issue, a more realistic wave velocity (see Section 2.5.4.2) should be adopted, thus calling for more advanced modelling techniques that *explicitly* describe the propagation of traffic (e.g., fluid-dynamic approaches, models with dynamical vehicle interactions, . . .).

As often is the case, a model's criticisms can be found in its underlying assumptions. In the case of the four step approach, it is obvious that all information regarding individual households is lost because of its aggregation to a trip level. As was already recognised from the start, the individual itself loses value during this conversion process. This opened the door towards another approach to transportation planning, more precisely *activity-based modelling* (ABM) which is discussed in the next section.

[6] Note that in the case of queue-based models, the simulations are typically implemented as efficient event-based schemes.

A final complaint that is more common around many of these grotesque models, is their requirement of a vast amount of specific data. In many cases, a diverse range of national studies are carried out, having the goal of gathering as much data as possible. Regardless of this optimism, some of the key problems remain, e.g., it is still not always straightforward to properly interpret and adapt this data so it can be used as input to a transportation planning tool.

3.1.3 Activity-based transportation models

As it was widely accepted that the rationale for travel demand can be found in people's motives for participating in social, economical, and cultural activities, the classical trip-based approach nevertheless kept a strong foothold in the transportation planning community. Instead of focussing attention elsewhere, the typical institutional policy was to ameliorate the existing four step models [McN00a]. However, some problems persistently evaded a solution with the trip-based approach, e.g., shops that remain open late, employers who introduce flexible working hours, the consideration of joint activities by members of a household, ...

In the next few sections, we illustrate how all this changed with the upcoming field of activity-based transportation planning. We first describe its historic origins, after which we move on to several of the approaches taken in activity-based modelling. The concluding section gives a concise overview of some of the next-generation modelling techniques, i.e., large-scale agent-based simulations.

3.1.3.1 Historic origins

The historic roots of the activity-based approach can probably be traced back to 1970, with the querulous work of Torsten Hägerstrand [Häg70]. He asserted that researchers in regional sciences should focus more on the intertwining of both disaggregate spatial and temporal aspects of human activities, as opposed to the more aggregate models in which the temporal dimension was neglected. This scientific field got commonly termed as *time geography*; it encompasses all time scales (i.e., from daily operations to lifetime goals), and focusses on the constraints that individuals face rather than predicting their choices [Mil04].

Central to Hägerstrand's work was the notion of so-called *space-time paths* of individuals' activity and travel behaviour. In a three-dimensional space-time volume, two spatial dimensions make up the physical world plane, with the temporal dimension as the vertical axis. The journey of an individual is now the path traced out in this space-time volume: consecutive visits to certain locations are joined by a curve, with vertical segments denoting places where the individual remained stationary during a certain time period. The complete chain of activities (called a *tour*) is thus joined by individual *trip legs*. In this respect, the space-time path represents the *revealed outcome* of an unrevealed behavioural process [McN00a]. An example of such a path can be seen in Figure 3.3: we can see a woman going from her home in Boulder (Colorado, USA), to the university's campus, followed by a visit to the post office and grocery store, and finally returning home [Det01]. Note that Hägerstrand extended his notion in the space-time volume to include *space-time prisms* that encapsulate and effectively *constrain* all of a person's reachable points (i.e., all his/her possible space-time paths), given a certain maximum travel speed as well as both starting and ending points within the volume [Cor05]. This environment is sometimes also referred to as a person's *action space*, enveloping that person's *time budget*.

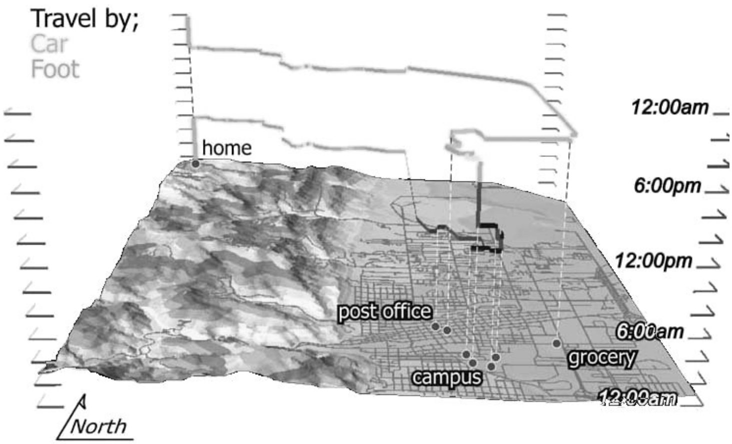


Figure 3.3: An example of a space-time path showing an individuals' activity and travel behaviour in the space-time volume: the two spatial dimensions make up the physical world plane, with the vertical axis denoting the temporal dimension. In this case, we can see a woman going from her home in Boulder (Colorado, USA), to the university's campus, followed by a visit to the post office and grocery store, and finally returning home (image reproduced after [Det01]).

Contrary to the belief that the field of activity-based transportation planning found its crux with the dissatisfaction of trip-based modelling, it grew and emerged spontaneously as a separate research study into human behaviour [McN00a]. The underlying idea however remained the same as in the trip-based approach, namely that travel decisions arise from a need to participate in social, economical, and cultural activities. But as opposed to the more aggregated trip-based view, the basic units here are individual activity patterns, commonly referred to as *household activity patterns*. In this context, the activity-based approach then studies the interactions between members of a household, and the relation to their induced travel behaviour [Axh00].

3.1.3.2 Approaches to activity-based modelling

Departing from Hägerstrand's initial comments, activity-based research progress has been slowly but steadily. In contrast with the development of the trip-based approach that culminated in the four step model, *there is no explicit general framework that encapsulates the activity-based modelling scheme*. There were however early comprehensive studies into human activities and their related travel behaviour, e.g., the synopsis provided by Jones et al. [Jon83]. As the field began to mature, certain ingredients could be recognised, e.g. [Axh00]:

- the *generation of activities*, which can be regarded as the equivalent of the production/attraction step in trip-based modelling,
- the modelling of *household choices*, i.e., with respect to their activity chains; this includes choosing starting time and duration of the activity, its location as well as a modal choice,
- the *scheduling of activities*, outlining how a household plans and executes the tasks of its members for long-, middle-, and short-term activities, going from year- and lifetime-long commitments, to daily operations.

During the last three decades, many research models that encompass activity scheduling behaviour have been developed. An excellent overview is given by Timmermans, who makes a distinction between *simultaneous* and *sequential models* [Tim01]. The former class is based on full activity patterns (e.g., for one whole day), whereas the latter is based on an explicit modelling of the activity scheduling process. Simultaneous models comprise utility-maximisation models and mathematical programming models (e.g., Recker's household activity pattern problem – HAPP [Rec95] and Bowman and Ben-Akiva's discrete choice model [Bow95]). Sequential models are frequently implemented as so-called *computational process models* (CPM), acknowledging the belief that individuals do not arrive at optimal choices, but rather employ context-dependent heuristics.

As an example of a CPM, we illustrate the seminal *Simulation of Travel/Activity Responses to Complex Household Interactive Logistic Decisions* (STARCHILD) model, which was originally a simultaneous model based on the maximisation of individuals' utilities. Based on Hägerstrand's notion and derivatives thereof, i.e., the central idea that an individual's travel behaviour is constrained by its space-time prism, Recker et al. developed the STARCHILD research tool addressing activity-based modelling [Rec86a; Rec86b]. The model hinges on three interdependent consecutive steps: (i) the generation of household activities, (ii) constructing choice sets for these activities, as well as scheduling them, and (iii) constraining these choices within the boundaries of the space-time prism [McN00a]. Another example of such a complete activity-based system, is *A Learning BAsed TRansportation Oriented Simulation System* (ALBATROSS) of Arentze and Timmermans [Are00]. A fundamental assumption in their work, is that activity participation is located at the level of a household, i.e., not at the level of its individual members (although noting the fact that these latter form a household's basic needs).

Note that the principal critique on these models' operations, was — and today still is — their need for an extensive amount of specifically tailored data that encompasses

Hägerstrand's concepts. Just as with the four step model, these data are arduous to come by. In short, most of the data are based on and transformed from, e.g., conventional trip-based surveys, *travel diaries* (e.g., the MOBEL (Belgium) and MOBIDRIVE (Germany) surveys of Cirillo and Axhausen [Cir02]) and the like, although more recently passive GPS-based information is collected [Axh00; McN00a; Mar02; Mil04].

In the future, a complete integration of activity generation, scheduling, and route choice (DTA) is expected to take place, on the condition that suitable empirical data will become available. We must however be careful not to be too optimistic, e.g., as Axhausen states that depending on the 'research-political' adoption of the activity-based approach, *"both a virtuous circle of progress or a vicious circle of stagnation are a possibility for the future"* [Axh00]. An even more harsh argumentation was voiced by Timmermans, who looked back at the development of the integration between land-use models and transportation planning [Tim03]. In his overview, he identified three waves, i.e., (i) aggregate spatial interaction-based models, (ii) utility-maximising multinomial logit-based models, and (iii) activity-based detailed microsimulation models. His final conclusion states that, despite the advances in finer levels of spatial detail, the scientific field has not undergone any significant theoretical progress. And although there exists a pronounced need for better behavioural models, the critique remains that this implies a tremendous complexity, hence the insinuation that many of the approaches are in fact based on black-box models.

3.1.3.3 Towards elaborate agent-based simulations

One of the most notable critiques often expressed against classical trip-based approaches such as the four step model, is the fact that all eye for detail at the level of the individual traveller is lost in the trip aggregation process. Activity-based modelling schemes try to circumvent this disadvantage by starting from a fundamentally different basis, namely individual household activity patterns. To this end, it is necessary to retain all information regarding these individual households during the planning process.

As hinted at earlier, an upcoming technique that fits nicely in this concept, is the methodology of *multi-agent simulations*. In such models, the individual households are represented as *agents*; the models then allow these agents to make independent decisions about their actions. These actions span from long-term lifetime residential housing decisions, the mid-term planning of daily activities, to even short-term decisions about an individual's driving behaviour in traffic. The following description of such a simulation system is based on the work of the group of Nagel et al. [Nag02a; Bal04a; Bal04b; Bal04c; Nag04; Hel04]:

- As a first step, a *synthetic population* of agents is generated. There is a close relation with the common land-use models, as these agents come from populations that should be correctly seeded, i.e., they should entail a correct demographic representation of the real world. Once the synthetic population is available to the model, the next step is to *generate activity patterns* (i.e., activity chains), generate these activities' *locations*, and finally the *scheduling* of the activities, as described in the previous section. Finally, *mode* and *route choice* form the bridge between the activity-based model and the transportation layer. As a consequence, it is beneficial to deal with agents' plans directly, rather than to rely on the information contained in OD tables [Bal04a].

Note that the generation of activity patterns has attracted a lot of research interest by itself, leading to quite sophisticated models. This should come as no surprise, as it forms the corner stone of the activity-based approach. In this respect, we note the work of Wang who devised a methodology for producing trips (i.e., step (I) of the four step model as explained in Section 3.1.2.2) based on the analyses of complete activity patterns [Wan97], the work of Kulkarni, who linked the trip-based approach to the generation of activity patterns and subsequently used Monte Carlo simulations of each household in the population [Kul02]. Finally, also note the work of Venter who investigated the decision on when people decide about their activities and the role this plays in regular activities and activity disturbances [Ven98].

- The component that represents the *physical propagation of agents* throughout, e.g., the road network, sits at the lowest level of the model. In this case, the necessary ingredients constitute the physical propagation of individual vehicles in the traffic streams. Popular models are *traffic cellular automata* and/or *queueing models*, allowing a fast and efficient simulation of individual agents in a network. Higher level models such as, e.g., pure fluid-dynamic models are inherently not suitable because they operate on a more aggregated basis and consequently ignore the individuality of each agent in the system. Note that this latter type of model can be deemed appropriate, on the condition that they can incorporate the tracking of individual particles by, e.g., a smoothed particle hydrodynamics method [Bal04b].

- An important issue that revolves around the two previous aspects, is the *clear absence of a rigidly defined direction of causality*, i.e., when exactly do people choose their travel mode, is it before the planning of activities, or is it rather a result from the planning process ? This problem can be dealt with in a broader context, wherein agents make certain plans about their activities, and *iteratively learn* by replanning and rescheduling (either on a *day-to-day* or *within-day* basis). This process of *systematic relaxation* continues until, e.g., a Wardrop equilibrium (W1) is reached (see step (IV) in Section 3.1.2.2 for more details). However, note that the question of whether or not people in reality strive to reach such an equilibrium, and whether or not such an equilibrium is even reached, remains an open debate. At this stage in the model, we are clearly dealing with aspects from evolutionary game theory, be it cooperative or non-cooperative. In this context, the concept of within-day replanning by agents is getting more attention, as it constitutes a necessary prerequisite for intelligent transportation systems, i.e., when and how do travellers react (e.g., *en-route choice*) to changes (e.g., control signals, incidents, . . .) in their environment [Bal04b] ?

The above description of a multi-agent activity-based simulation system may seem straightforward, nevertheless, no complete practical implementation exists today. The model suite that comes the closest to reaching the previously stated goals, is the *TRansportation ANalysis and SIMulation System* (TRANSIMS[7]) project. This project is one part of the multi-track Travel Model Improvement Program of the U.S. Department of Transportation, the Environmental Protection Agency, and the Department of Energy in the context of the Intermodal Surface Transportation Efficiency Act and the Clean Air Act Amendments of 1990. Its original development started at Los Alamos National Laboratory, but a commercial implementation was provided by IBM Business Consulting.

Since its original inception, TRANSIMS has been applied to a various range of case studies. One of the most notable examples, is the truly country-wide agent-based detailed microsimulation of travel behaviour in Switzerland (see also Figure 3.4) [Voe01; Ran03]. A similar study encompassing the iteration and feedback between a simulation model and a route planner, was carried out for the region of Dallas [Nag98a; Nag98b]. In the context of large-scale agent-based simulations, queueing models were employed as a TRANSIMS component by the work of Simon et al. for the city of Portland [Sim99], as well as the work of Gawron [Gaw98a; Gaw98b] and Cetin et al. [Cet03]. Because of the computational complexity involved in dealing with the enormous amount of agents in real-world scenarios, a popular approach is the use of parallel computations, as described in the work of Nagel and Rickert [Nag01]. Another example of this last type of simulations, is the work of Chopard and Dupuis who applied the methodology of large-scale simulations to the city of Geneva [Cho95; Cho97; Dup98].

As a final remark, we would like to draw attention to some more *control-oriented aspects of multi-agent simulations*. In this respect, the transportation system is considered as a whole, whereby the agents are now the local controllers within the system (e.g., traffic lights, variable message signs, . . .), instead of individual households as was previously assumed. Using a coordinated control approach, is it then possible to achieve a system optimum. An example of such a control system is the *Advanced Multi-agent Information and Control for Integrated multi-class traffic networks* (AMICI[8]) project from The Netherlands. As one of its goals, it strives to provide routing information to different classes of road users, as well as controlling them by means of computer simulated agents who operate locally but can be steered hierarchically. Related to this, is the work of van Katwijk et al., who developed a test bed that allows to perform road traffic management by controlling a multi-agent system using a rule-based decision model [Kat04].

[7] http://www.transims.net

[8] http://www.amici.tudelft.nl

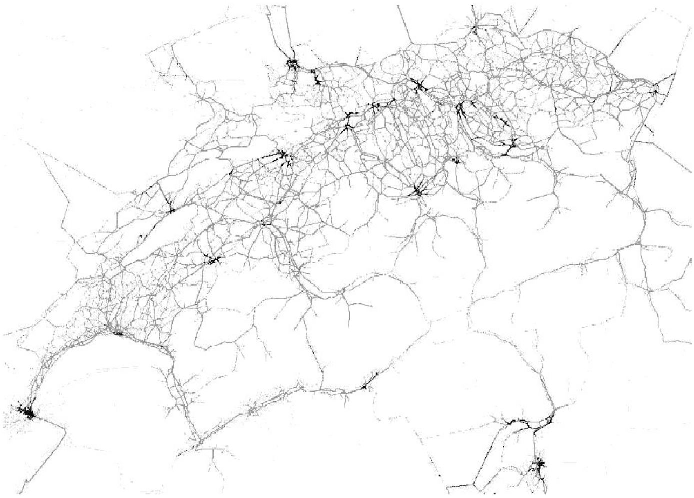


Figure 3.4: An example of a multi-agent simulation of the road network of Switzerland, around 08:00 in the morning: each vehicle is indicated by a single grey pixel, with low-speed vehicles coloured black. The image clearly reveals more vehicular activity (and congestion) in the city centres than elsewhere in the country (image reproduced after [Voe01]).

3.1.4 Transportation economics

Most of the work related to traffic flow theory has been considered by researchers with roots in engineering, physics, mathematics et cetera. With respect to transportation planning, the scene has shifted somewhat during the last couple of decades towards policy makers who test and implement certain strategies, based on, e.g., the four step modelling approach. Around 1960 however, another branch of scientists entered the field of transportation planning: economists developed standard models that viewed transportation as a market exchange between demand and supply.

Generally stated, the economics of transportation does not exclusively focus on traffic as a purely physical phenomenon (i.e., the theory of traffic flows as explained in the previous chapter), but also takes into account the fact that *transportation incurs certain costs*, both to the individual as well as to the society as a whole [Lin00].

In the following sections, we describe the setting in which economists view transportation, after which we discuss the concept of road pricing.

3.1.4.1 The economical setting

In the context of economic theory, a transportation system can be seen as an interaction between *demand* (profits) and *supply* (costs). In a static setting, both demand and supply are frequently described by means of functions: they are expressed as the *price* for a good associated with the *quantity* of that good. In transportation economics, quantity is frequently identified as the *number of trips made* (e.g., by the macroscopic characteristic of traffic flow q as defined in Section 2.3.2) [Bec55]. In the remainder of this section, we use the term *travel demand* to denote the demand side, as opposed to the supply side which is composed of the transportation infrastructure (including changes due to incidents, ...). In a more broader context, travel demand is typically described as the amount of traffic volume that *wants* to use a certain infrastructure (i.e., the supply): when demand thus exceeds the infrastructure's capacity under congested conditions (implying queueing), this supply effectively acts as a constraint to the present volume of traffic flow.

According to the aforementioned conventions, a demand side function is expressed as a certain cost associated with a level of flow (i.e., number of trips). We call such a curve a *travel demand function* (TDF), and it is typically decreasing with increasing flow; an example of such a function can be seen as the dotted curve in Figure 3.5. Note that, to be correct, the depicted curve actually represents a *marginal travel demand*: it gives the additional profit that is received with the obtaining of one extra unit (the total amount of profit is just the area under (i.e., the integral of) the demand curve). Translated to a transportation system, this means that the benefits of a traveller tend to decrease with increasing travel demand (i.e., congestion).

In similar spirit, we can consider a supply side curve, i.e., price (costs) versus quantity (flow). One of the most used approaches for describing traffic flow operations at the supply side from an economical point of view, is the use of an *average cost function* (ACF) [Ver98] as proposed by A.A. Walters in 1961 [Wal61]. The theory was based on the description of traffic flow by means of fundamental diagrams (see Section 2.5). Consider for example Greenshields' simple linear relation between density k and space-mean speed $\overline{v}_{\text{s}}$, as given by equation (2.46) on page 46. The corresponding $\overline{v}_{\text{s}_{\text{e}}}(q)$ fundamental diagram (e.g., Section 2.5.2.2), consists of a tilted parabola, lying on its side. From equation (2.41) on page 37, it follows that the travel time T is inversely proportional to the space-mean speed $\overline{v}_{\text{s}}$. Walters' idea now was to assume certain *costs* related to the travel demand. Some examples of these costs are those associated with [Bor01; Hau05a; Hau05b]:

- (i) the construction of the transportation infrastructure,
- (ii) vehicle ownership and use,
- (iii) taxes,
- (iv) travel time, i.e., *value of time* (VOT),
- and (v) environmental and social costs.

Based on these costs, and using the relation between travel time and travel demand, Walters derived a functional relationship for the economical cost C associated with the travel demand q. This relationship (i.e., the ACF) denotes the supply side of transportation economics; an example is the thick solid curve in Figure 3.5.

Once both travel demand and average cost functions are known, they can be used to determine the *equilibrium points* that occur at their intersection(s): given both curves, the transportation system is assumed to settle itself at these equilibria, with a certain travel cost associated with the equilibrium traffic demand. Note that because of the nature of the analysis procedure, i.e., using static (stationary) curves, the resulting travel costs are *average* costs, hence the name of the average cost function.

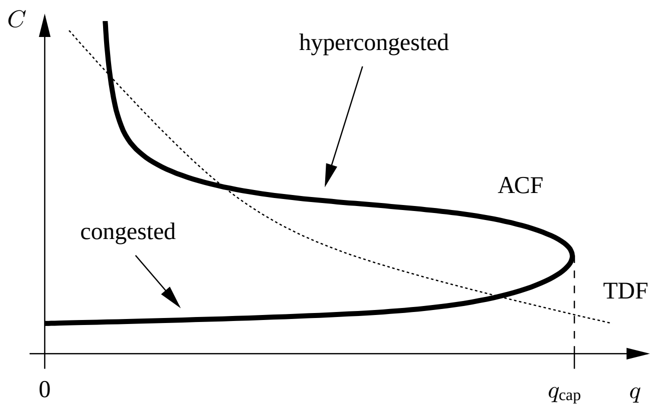


Figure 3.5: A graphical illustration of the economics of transportation operations: the dotted curve represents the demand side, i.e., the travel demand function (TDF), whereas the thick solid curve represents the supply side, i.e., the average cost function (ACF). Both curves express the cost C associated with the number of trips made (e.g., level of traffic flow q). The latter curve is said to have two states, namely the congested and the hypercongested area (identified as the backward-bending part of the curve). Points where both demand and supply curves intersect each other denote equilibrium points: given both curves, the transportation system as assumed to settle itself at their intersection(s), with a certain travel cost associated with the equilibrium traffic demand.

There are some distinct features noticeable in the relation described by the average cost function. For starters, the curve does not go through the origin, i.e., at low travel demands there is already a *fixed cost* incurred. The depicted cost then typically increases with increasing travel demand, mainly due to the contribution of the value of time associated with the travel time. The most striking feature however, is the fact that the curve contains a *backward-bending* upper branch [Els81]. This peculiar branch has an asymptotic behaviour, i.e.,

$$\lim_{q \to q_{\text{cap}} \to 0} C(q) = +\infty, \tag{3.2}$$

where we have denoted the path taken by the limit, i.e., passing through the capacity flow q_{cap} towards the upper branch, which in fact corresponds to an increase towards the jam density k_{jam}. Also note the presence of an inflection point (for concave $q_{\text{e}}(k)$ fundamental diagrams), which can be located analytically by differentiating the functional relation twice, and solving it with a right hand side equal to zero.

> In contrast to the nomenclature adopted by the traffic engineering community and in this dissertation, economists typically refer to the lower branch of Figure 3.5 as the *congested* state, and to the upper branch as the *hypercongested* state. Their line of reasoning being the conviction that in a certain sense, congestion also occurs when the speed drops below the free-flow speed on the free-flow branch [Ver98; Lin00; Sma03].

With respect to the relevance of this hypercongested state, there has been some debate in literature. Among most economists there seems to be a consensus, in the sense that the hypercongested branch is actually a transient phenomenon [Yan98a; Lin00]. Walters thought of the branch as just a collection of inefficient equilibria, but it was shown by Verhoef that all equilibria obtained on the hypercongested branch are inherently unstable [Ver98; Sma03]. Another argument, that discards the use of the branch, goes as explained by Yang and Huang [Yan98a]: many traditional economical models of transportation assume a static (stationary) model of congestion, similarly as in classical static traffic assignment described in Section 3.1.2.3. Under this premise, the relations as described by the fundamental diagrams, should be considered for complete links, and not only — as is usual in traffic flow theory — at local points in space and time. Therefore, a link may contain two different states: a free-flow state and a congested state. Hence, the average cost function should only describe the properties that are satisfied on the whole link, and as a result this excludes the global hypercongested regime.

Several ad hoc solutions exist for dealing with this problem, which is a consequence of using cost functions based on stationary equilibria: some of these solutions typically entail the use of vertical segments near the capacity flow in Figure 3.5, resulting in finite queueing delays on heavily congested links [Ver98; Yan98a; Sma03; Ver05b]. Another much used solution that ignores the backward-bending branch, is to directly specify the average cost function based on a link's observed capacity, instead of deriving it through the fundamental diagram of space-mean speed versus flow [Lin00]. Note that in most cases, the capacity considered here has a lower value as opposed to its maximum value. A similar example that specifies the relation between travel demand and travel time (e.g., VOT), is the BPR travel impedance function as described in Section 3.1.2.3, in which there is also a difference between the practical capacity and the maximum capacity of a link as defined by a fundamental diagram. Notwithstanding these proposed specific solutions, the mainstream tendency nowadays seems to imply the use of traffic flow models that explicitly describe the dynamics of congestion, either by using queueing models, or more elaborate models based on fluid dynamics or detailed simulations of individual vehicles [Yan98a].

3.1.4.2 Towards road pricing policies

In an economical treatment of transportation, road users in general only take into account their own *private costs*, such as (ii) vehicle ownership and use, (iii) taxes, and (iv) costs related to the travel time. Note that because, as mentioned earlier, we are working with *marginal cost functions*, the cost (i) related to the transportation infrastructure is not taken into account (as this is only a one-time initial cost).

To this end, we consider the average cost function from two different points of view: on the one hand, we have the *private costs* borne by an individual traveller, and on the other hand, we have the *external costs* that the traveller bears to the rest of society. In accordance with economic literature, we call the former associated cost function the *marginal private cost function* (MPCF), and the latter the *marginal social cost function* (MSCF). The extra costs to society brought on by individual travellers, are called *negative externalities*.

In Figure 3.6, we have depicted the resulting equilibria that arise from the intersections of the travel demand function with both marginal private and social cost functions (note that we disregard the upper backward-bending branch of the average cost function as was shown in Figure 3.5). In an unmanaged society, i.e., where no measures are taken to change individual travellers' behaviour, the resulting equilibrium will be found at q_{ue}, which is in fact a *user equilibrium* corresponding to a cost as dictated by the marginal private cost function (MPCF) [Arn94b]. As travellers act selfishly, not considering the costs inflicted upon other travellers (e.g., more road users imply more congestion *for everybody*), this pricing method is termed *average cost pricing*. At this equilibrium, the unpaid external cost to society equals the difference between the MSCF and MPCF curves at a demand level of q_{ue} [Rou06].

As early as in 1920, Arthur Cecil Pigou noted that road users do not take into account the costs they inflict upon other travellers. In order to rectify this situation, he proposed to levy governmental taxes on road use. Pigou actually discussed his idea in a broader economic setting, by making a distinction between the private and the social costs. Charging of a suitable governmental tax could change the balance so the negative externalities would be included, resulting in a new equilibrium [Pig20]. This process is called *internalising the external costs*.

Some years later in 1924, Frank Hyneman Knight further explored Pigou's ideas[9]: Knight fully acknowledged the fact that congestion justified the levying of taxes. In contrast to Pigou however, Knight raised some criticism in the sense that not governmental taxes were necessary, but instead private ownership of the roads would take care of tax levying and consequently resulting in a reduction of congestion [Kni24; Pah06].

In 1927, Frank Plumpton Ramsey cast this methodology — called *marginal cost pricing* — in the light of *social welfare economics*. This latter type employs techniques

[9] Note that Knight apparently was clued in his research by an error made on Pigou's behalf in his study of a two-route road network [Boy04a]. Even more intriguing, is the fact that this type of problem was already considered as far back as 1841, with the work of the German economist Johann Georg Kohl [Koh41].

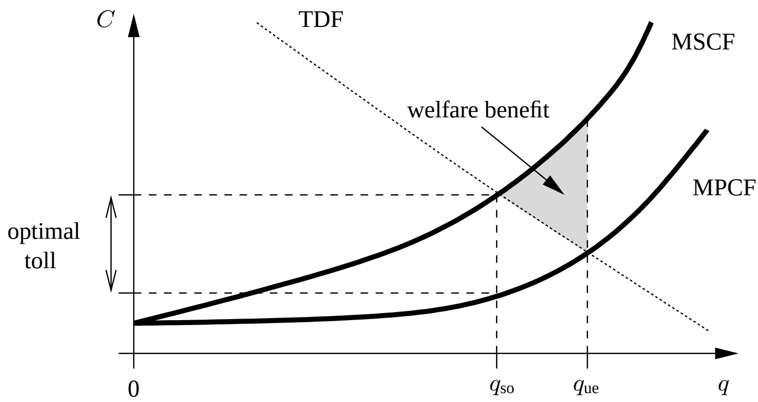


Figure 3.6: An economical equilibrium analysis based on a travel demand function (TDF) represented by the dotted curve, and marginal private and social cost functions (MPCF and MSCF) represented by the thick solid curves. The user equilibrium is located at a demand of q_{ue}, whereas the system optimum is located at a lower demand of q_{so}. The welfare benefit (indicated as the grey triangular region) can be gained by levying a congestion toll equal to the difference between the marginal social and private cost function defined at the system optimum demand level q_{so}.

from a branch that is called micro-economics, which is an economical treatment of society based on the behaviour of individual producers and consumers. Welfare economics embraces two important concepts:

Efficiency: a measure for assessing how much benefit society gains from a certain policy rule. It can be considered with the *(strict) Pareto criterion* (invented by Vilfredo Pareto), which states that efficiency improves if a policy rule implies an increase of welfare for at least one individual, but no other individual of society is worse off. Nicholas Kaldor and John Hicks restated Pareto's criterion, but this time from the point of view of those who gain and those who lose, respectively. Their *Kaldor-Hicks criterion* states that society gains welfare, but not everybody receives personal gain, i.e., there will be winners and losers. The crucial assumption on the Kaldor-Hicks criterion is that the winners could fully compensate the losers, in theory; whether or not this happens at all, is not the issue.

Equity: if society benefits from a certain policy rule, then its efficiency can be measured using, e.g., the Pareto criterion as stated earlier. However, nothing is said about the size of the benefit each individuals of society receives. This is were the concept of equity enters the picture: it refers to a fair distribution of the total benefits over all individuals in society (note that in this case, there typically is a strong correlation with the *income distribution*).

In this context, Ramsey thus stated a policy rule, implying a *maximisation of the so-*

cial welfare [Ram27]. In the field of transportation, this can be done by marginal cost pricing, also called *road pricing*, *congestion tolls*, ... The nature of the measure is that it signifies a demand-side strategy, with the goal of inducing a change in travellers' behaviour. Road pricing typically entails a shift from on-peak to off-peak periods, switching mode (e.g., from private to public transportation), car pooling, route change, ... Considering again Figure 3.6, we can see that if users were to consider the marginal social cost function, instead of only their marginal private cost function, this would shift the resulting equilibrium from q_{ue} to q_{so}, which is a *social optimum*. As said at the beginning of this section, travellers do not take into account the negative externalities they cause to the rest of society, and as such, they can be charged with an *optimal toll* that is defined as the difference between both marginal social and private cost functions. Levying the correct congestion toll, would remove the original market failure, resulting in a net social welfare benefit that is visualised as the grey triangular region in Figure 3.6. Note that in an ideal world, congestion tolls exactly match the caused negative externalities. In practice however, this can not be accomplished, resulting in so-called *second-best pricing* schemes. Practical real-life examples of this are tolling the beltway around a city upon entering it (e.g., London's recent congestion charge), using step-tolls, tolling at fixed time periods instead of based on traffic conditions, ... [Lin01]

Reconsidering the work of Wardrop with respect to the criteria (W1) and (W2) highlighted in Section 3.1.2.2, Beckmann, McGuire, and Winsten found that the system optimum q_{so} can be reached if the standard cost (i.e., journey time) is replaced with a *generalised cost*, which is just the marginal social cost function as described earlier [Bec55]. Consequently, the total travel time in the system can be minimised (from an engineering perspective), by levying a so-called efficiency toll, which corresponds to Ramsey's optimal toll.

One of the most notable extensions in the economic treatment of transportation and congestion tolls, is the seminal work of the late Nobel prize winner William Vickrey [Vic69]. As already stated in Section 3.1.2.5, correct modelling of, e.g., queue spill back, is of fundamental importance when assessing the effectiveness of road pricing schemes for example. Vickrey's *bottleneck model* is one step in this direction: it is based on the behaviour of morning commuters, whereby the model takes into account the departure times of all travellers. As everybody's desire is to arrive at work at the same time, some will arrive earlier, others later. Besides the traditional travel time costs, travellers therefore also experience so-called *schedule delay costs*. Levying suitable tolls that depend on the travellers' arrival times, allows to reach a system optimum [Arn98; Lin00]. It is important to realise here that the levied toll should vary over time and space, in order to correspond to the governing traffic conditions.

To most people in society road pricing is a highly unpopular measure, as well as a controversial political issue, whereby public acceptance is everything [Mar98; Hår01b; Hau05a; Hau05b]. Alternatives to road pricing can include upgrading existing roads and/or public transportation services, strict control-oriented regulation by means of *advanced traffic management systems* (ATMS), issuing elaborate parking systems, fuel taxes, et cetera [Arn94a]. In spirit of second-best pricing schemes, it was a wise thing in the UK to connect London's congestion charge to the simultaneous improvement of public transportation [Sma05]. Similarly, the cordon toll system in the city of Oslo, Norway, quickly found acceptance among the population [Hår01b]. Nevertheless, road pricing is considered an *unfair* policy measure to most people: households (and firms) with higher incomes, can more easily afford to pay the charge, hence they will keep the luxury of travelling at their own discretion, whilst others might not be able to pay the required toll. As a consequence, an inconsiderate internalisation of the external costs, does not lead to an equitable Pareto optimal scenario. Despite this resistance, there does seem to be a general consensus among members of society that congestion caused by traffic induced by recreational activities, is not tolerated during peak periods; congestion tolling for these travellers is deemed appropriate.

Despite the advances in the methodology underlying road pricing, there is still one major gap that has yet to be filled in, i.e., the equity of the principle, or otherwise stated: where do the gained social welfare benefits (i.e., tax revenues) go in the redistribution ? As Small states, road pricing is more acceptable to the broad public, if it is presented as a complete financial package [Sma92]. As welfare economists debate on whether or not the revenues should go back to the transportation sector or rather elsewhere, Small asserts that inclusion of the former is mandatory for achieving substantial support from both the political side and the investors. Complementary, in order to satisfy the global population, it is advisory to use the collected charges in order to diminish, e.g., labour taxes, as they are perceived as being too high [Har01a; Gof04; Hau05a; Hau05b].

In the end, we should note that both economists and traffic engineers are essentially talking about the same subject, although by approaching it from different angles. In the field of economics, road pricing policies are introduced based on average cost functions, allowing an optimisation of the social welfare. This effectively corresponds to the engineers' idea of static traffic assignment, based on a system optimum using travel impedance functions (see, e.g., Section 3.1.2.3). The validity of using these average cost functions (with or without their backward-bending parts as explained in Section 3.1.4.1), has instigated several debates in road pricing literature, most notably between Else and Nash [Els82; Nas82], Evans and Hills [Eva92; Eva93; Hil93], and Ohta and Verhoef [Oht01a; Oht01b; Ver01].

In continuation, the approach followed by Vickrey's bottleneck model provides a nice, first alternative, using schedule delay costs (see Sections 3.1.2.5 and 3.1.4.2). Although Vickrey's idea introduces a hitherto absent time dependence, it has the disadvantage that spatial extents are neglected through the assumption of point-queues (see Section 3.1.2.5). Lo and Szeto have rigourously shown that hyper-congestion is essentially a spatial phenomenon, and that by neglecting this facet, a road pricing policy might actually worsen traffic conditions [Lo05]. The correct way out of this problem, is by explicitly taking the tempo-spatial characteristics of traffic flows into account. As an engineering analogy, this can be accomplished by introducing dynamic traffic assignment (see Section 3.1.2.4) which uses physical propagation models to describe the buildup and dissolution of congestion (see also some of the models presented in Section 3.1.2.5, e.g., the work of Lago and Daganzo [Lag03b]).

3.2 Traffic flow propagation models

In contrast to the previous section, which dealt with high level transportation planning models, the current section considers traffic flow models that explicitly describe the *physical propagation* of traffic flows. In a sense, these models can be seen as being directly applicable for the physical description of *traffic streams*. There exist several methods for discriminating between the families of models that describe traffic flow propagation, i.e., based on whether they operate in continuous or discrete time (or even event-based), whether they are purely deterministic or stochastic, or depending on the *level of detail* (LOD) that is assumed, ... More detailed explanations can found in the overview of Hoogendoorn and Bovy [Hoo01]. In this dissertation, we present an overview that is based on the latter method of discriminating between the level of detail. We believe that this classification most satisfactorily describes the discrepancies between the different traffic flow models. Thus, depending on the level of aggregation, we can classify the propagation models into the following four categories:

- **macroscopic** (highest level of aggregation, lowest level of detail, based on continuum mechanics, typically entailing fluid-dynamic models),

- **mesoscopic** (high level of aggregation, low level of detail, typically based on a gas-kinetic analogy in which driver behaviour is explicitly considered),
- **microscopic** (low level of aggregation, high level of detail, typically based on models that describe the detailed interactions between vehicles in a traffic stream),
- and **submicroscopic** (lowest level of aggregation, highest level of detail, like microscopic models but extended with detailed descriptions of a vehicles' inner workings).

Note that some people regard macroscopic models more from the angle of *network models*. In this case, the focus lies on *performance characteristics* such as total travel times (a measure for the quality of service), number of trips, ... [Gar97] To this end, several quantitative models were introduced, such as Zahavi's so-called α relation between traffic flow, road density, and space-mean speed [Zah72], and Prigogine and Herman's *two-fluid theory* of town traffic [Her79].

3.2.1 Macroscopic traffic flow models

In this section, we take a look at the class of traffic flow models that describe traffic streams at an aggregated level. We first introduce the concept behind the models (i.e., the continuum approach), after which we discuss the classical first-order LWR model. Because of its historical importance, we devote several sections to the model's analytical and numerical solutions, as well as to some extensions that have been proposed by researchers. We conclude our discussion of macroscopic models with a description of several higher-order models, and shed some light on the problems associated with both first-order and higher-order models.

3.2.1.1 The continuum approach

Among the physics disciplines, exists the field of *continuum mechanics* that is concerned with the behaviour of solids and fluids (both liquids and gasses). Considering the class of *fluid dynamics*, it has spawned a rich variety of branches such as aerodynamics, hydrodynamics, hydraulics, ...

Underlying these scientific fields, is the *continuity assumption* (also called the *continuum hypothesis*) that they all have in common. In a nutshell, this assumption states that fluids are to be treated as continuous media (in contrast to, e.g., molecular gasses, which consist of distinct particles). Stated more rigourously, the macroscopic spatial (i.e., the length) and temporal scales are considerably larger than the largest molecular corresponding scales [Cra04]. As a consequence, all quantities can be treated as being continuous (in the infinitesimal limit). The decision to use either a liquid-like or a gas-like treatment, is based on the *Knudsen number* of the fluid: a low value (i.e., smaller than unity) indicates a fluid-dynamic treatment, whereas a high value is indicative of a more granular medium. In this section, we consider the former approach. In the latter case, we enter the realm of statistical mechanics that deals with, e.g., kinetic gasses, requiring the use of the Boltzmann equation (as will be explained in Section 3.2.2 on mesoscopic traffic flow models).

Historically, the fluid-dynamic approach found its roots in the seminal work of Claude Louis Navier (1822), Adhémar de Saint-Venant (1843), and George Gabriel Stokes (1845) [Gir03]. This gave rise to what we know as the *Navier-Stokes equations* (NSE), formulated as a set of *non-linear partial differential equations* (PDEs). For our overview, the most relevant equation is actually the local *conservation law*, stating that the net flux is accompanied by an increase or decrease of material (i.e., fluid). In general, we can discern two subtypes: *compressible* or *incompressible* fluids, and *viscous* or *inviscid* fluids. Incompressibility assumes a constant density (and a high *Mach number*), whereas inviscid fluids have a zero viscosity (with a corresponding high *Reynolds number*) and are typically represented as the *Euler equations*.

Note that the NSE are still not fully understood. The fact of the matter is that the *Clay Mathematics Institute* has devised a list of *Millennium Problems*[10], among which a deeper fundamental understanding of the NSE holds a reward of one million dollar. Because the original Navier-Stokes equations are too complex to solve, scientists developed solutions to specific subproblems, e.g., Euler's version of inviscid fluids. As an example, we give the *Burgers equation*, as derived by Johannes Martinus Burgers [Bur48], for a one-dimensional fluid in the form of a *hyperbolic conservation law*:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \tag{3.3}$$

in which the $u \in \mathbb{R}$ typically represents the velocity, and ν is the viscosity coefficient. For inviscid fluids, $\nu = 0$, such that equation (3.3) corresponds to a *first-order* partial differential equation. This type of *hyperbolic PDE* is very important, as its solution can develop discontinuities, or more clearly stated, it can contain *shock waves* which are of course directly relevant to the modelling of traffic flows. The inviscid Burgers PDE can be solved using the standard *method of characteristics*, as will be explained in further detail in the next three sections.

3.2.1.2 The first-order LWR model

Continuing the previous train of thought, we can consider traffic as an inviscid but compressible fluid. From this assumption, it follows that densities, mean speeds, and flows are defined as continuous variables, in each point in time and space, hence leading to the names of *continuum models*, *fluid-dynamic models*, or *macroscopic models*.

The first aspect of such a fluid-dynamic description of traffic flow, consists of a *scalar conservation law* ('scalar' because it is a first-order PDE). A typical derivation can be found in [Gar97] and [Jün02]: the derivation is based on considering a road segment

[10] `http://www.claymath.org/millennium`

with a finite length on which no vehicles appear or disappear other than the ones that enter and exit it. After taking the infinitesimal limit (i.e., the continuum hypothesis), this will result in an equation that expresses the interplay between continuous densities and flows on a local scale. Another way of deriving the conservation law, is based on the use of a differentiable cumulative count function $\widetilde{N}(t,x)$ of Section 2.3.2.2 [New93a]:

$$\begin{aligned}
k(t,x) = -\partial\widetilde{N}(t,x)/\partial x \quad &\text{and} \quad q(t,x) = \partial\widetilde{N}(t,x)/\partial t, \\
&\Downarrow \\
\frac{\partial k(t,x)}{\partial t} = -\frac{\partial^2\widetilde{N}(t,x)}{\partial t\,\partial x} \quad &\text{and} \quad \frac{\partial q(t,x)}{\partial x} = \frac{\partial^2\widetilde{N}(t,x)}{\partial t\,\partial x}, \\
&\Downarrow \\
\frac{\partial k(t,x)}{\partial t} \quad &+ \quad \frac{\partial q(t,x)}{\partial x} = 0,
\end{aligned} \tag{3.4}$$

with the density k and flow q dynamically (i.e., time varying) defined over a single spatial dimension. Lighthill and Whitham were among the first to develop such a traffic flow model [Lig55] (note that in the same year, Newell had constructed a theory of traffic flow at low densities [New96]). Crucial to their approach, was the fundamental hypothesis as explained in Section 2.5.2.1, essentially stating that flow is a function of density, i.e., the $q_\text{e}(k(t,x))$ equilibrium relationship in Section 2.5.2.2. Essentially to their theory, Lighthill and Whitham assumed that the fundamental hypothesis holds at all traffic densities, not just for light-density traffic but also for congested traffic conditions. Using this trick with the fundamental diagram, relates the two dependent variables in equation (3.4) to each other, thereby making it possible to solve the partial differential equation.

One year later, in 1956, Richards independently derived the same fluid-dynamic model [Ric56], albeit in a slightly different form. The key difference, is that Richards focusses on the derivation of shock waves with respect to the density, whereas Lighthill and Whitham consider this more from the perspective of disruptions of the flow [Pip64]. Another difference between both derivations, is that Richards fixed the equilibrium relation, whereas Lighthill and Whitham did not restrict themselves to an a priori definition; in Richards' paper, we can find the equation $V = a(b - D)$, with V the space-mean speed, D the density, and a and b fitting parameters [Ric56]. Note that all three authors share the following same comment: because of the continuity assumption, the theory only holds for a large number of vehicles, hence the description of *"long crowded roads"* in Lighthill and Whitham's original article.

> Because of the nearly simultaneous and independent development of the theory, the model has become known as the *LWR model*, after the initials of its inventors who receive the credit. In some texts, the model is also referred to as the *hydrodynamic model*, or the *kinematic wave model* (KWM), attributed to the fact that the model's solution is based on characteristics, which are called kinematic waves (e.g., shock waves).

3.2.1.3 Analytical solutions of the LWR model

Reconsidering equation (3.4), taking into account the fundamental diagram, the conservation law is now expressed as:

$$k_{\mathrm{t}} + q_{\mathrm{e}}(k)_{\mathrm{x}} = 0, \tag{3.5}$$

in which we introduced the standard differential calculus notation for PDEs[11]. Recognising the fundamental relation of traffic flow theory (see Section 2.3.4.2 for more details), equation (3.5) then becomes:

$$k_{\mathrm{t}} + (k\,\overline{v}_{\mathrm{s_e}}(k))_{\mathrm{x}} = k_{\mathrm{t}} + \left(\overline{v}_{\mathrm{s_e}}(k) + k\frac{d\overline{v}_{\mathrm{s_e}}(k)}{dk}\right) k_{\mathrm{x}} = 0. \tag{3.6}$$

The above hyperbolic PDE, can be translated into the Burgers equation (3.3), using a suitable transformation to a dimensionless form as explained in the rigourous mathematical treatment provided by Jüngel [Jün02]. The conservation law (3.5) can also be cast in a non-linear wave equation, using the chain rule for differentiation [Gar97]:

$$k_{\mathrm{t}} + \frac{dq_{\mathrm{e}}(k)}{dk} k_{\mathrm{x}} = 0. \tag{3.7}$$

Analytically solving the previous equation using the method of characteristics, results in shock waves that travel with speeds equal to:

$$w = \frac{dq_{\mathrm{e}}(k)}{dk}, \tag{3.8}$$

i.e., the tangent to the $q_{\mathrm{e}}(k)$ fundamental diagram. This tangent corresponds to the in Section 2.5.2.2 mentioned *kinematic wave speed* w. As a consequence, solutions, being the characteristics, of equation (3.7) have the following form:

$$k(t,x) = k(x - wt), \tag{3.9}$$

with the observation that the density is constant along such a characteristic. Note that in order to obtain shock waves that are only decelerating, the used $q_{\mathrm{e}}(k)$ fundamental diagram should be concave (a property that is often neglected) [Del95a]:

$$\frac{d^2 q_{\mathrm{e}}(k)}{dk^2} \leq 0. \tag{3.10}$$

[11] Note that we set the PDE's right-hand side equal to zero. A non-zero term here can be considered as a source term that models a local entry or exit such as an on- or off-ramp [Bag05].

Starting from an initial condition, the problem of finding the solution to the conservation PDE, is also called an *initial value problem* (IVP), whereby the solution describes how the density evolves with increasing time. The problem is called a *generalised Riemann problem* (GRP) when we consider an infinitely long road with given constant initial densities up- and downstream of a discontinuity.

Because the method of characteristics can result in non-unique solutions, a trick is used to select the correct, i.e., physically relevant, one. Recall from equation (3.3) that the right-hand side of the Burgers PDE contained a viscosity term ν. The general principle that is adopted for selection of the correct solution, is based on the *Oleinik entropy condition,* which regards the conservation law as a diffuse equation. In this context, the viscosity coefficient is multiplied with a small diffusion constant ϵ. In the *vanishing viscosity limit* $\epsilon \longrightarrow 0$, the method returns a unique, smooth, and physically relevant solution instead of infinitely many (weak) solutions [LeV92; Nag05]. For more details with respect to the application of traffic flows, we refer to the excellent treatment given by Jüngel [Jün02].

> Ansorge, Bui et al., Velan and Florian later reinterpreted this entropy condition, stating that it is equivalent to a *driver's ride impulse* [Ans90; Bui92; Vel02]. Drivers going from free-flow to congested traffic encounter a sharp shock wave, whereas drivers going in the reverse direction essentially pass through all intermediate points on the fundamental diagrams, i.e., the solution generates a fan of waves. It is for this latter case that the 'ride impulse' is relevant: it denotes the fact that stopped drivers prefer to start riding again, resulting in a fan of waves.

A more intuitive explanation can also be given based on the *anticipation* that drivers adopt when they approach a shock wave: their equilibrium speed $\overline{v}_{\text{s}_\text{e}}(k)$ is also a function of the change in density, e.g.:

$$\overline{v}_{\text{s}_\text{e}}(k) \doteq \overline{v}_{\text{s}_\text{e}}(k) - \frac{\nu}{k}\frac{\partial k}{\partial x}. \tag{3.11}$$

Substituting this new equilibrium relation in equation (3.6), results in a right-hand side equal to $\nu\frac{\partial^2 k}{\partial x^2}$. Applying the same methodology based on the vanishing viscosity limit of the entropy solution, results in the same unique solution. Because the shock waves are in fact mathematical discontinuities, and as such, infinitesimally small, they are typically 'smeared out' by numerical schemes. In fact, this is just the equivalent of introducing an artificial viscosity (as explained earlier), which allows diffusion (i.e., the combined effect of dissipation and dispersion) of the shock waves. Note that this diffusion is a consequence of the numerical solution, and not necessarily corresponding to the real diffusion processes in a viscous fluid. This numerical smoothing helps to retain numerical stability of the final solution.

Whenever in the solution of the conservation equation, two of its characteristics intersect, the density takes on two different values (each one belonging to a single characteristic). As this mathematical quirk is physically impossible, the entropy solution states that both characteristics terminate and breed a *shock wave*; as such, these shock waves form boundaries that discontinuously separate densities, flows, and space-mean speeds [Gar97]. The speed of such a shock wave is related to the following ratio [Pip64]:

$$w_{\text{shock}} = \frac{\Delta q}{\Delta k}, \tag{3.12}$$

with $\Delta q = q_{\text{up}} - q_{\text{down}}$ and $\Delta k = k_{\text{up}} - k_{\text{down}}$ the relative difference in flows, respectively densities, up- and downstream of the shock wave.

Note that going from a low to a high density regime typically results in a shock wave, whereas the reverse transition is accompanied by an emanation of a *fan of characteristics* (also called *expansion*, *acceleration*, or *rarefaction waves*). In shock wave theory, the densities on either side of a shock are well defined (i.e., unique solutions exist); along the shock wave however, the density jumps discontinuously from one value to another. In this latter respect, equation (3.12) is said to satisfy the *Rankine-Hugoniot jump condition*.

The previous remarks with respect to the entropy condition, are closely related to the concavity of the $q_{\text{e}}(k)$ fundamental diagram, as defined by equation (3.10): for concave $q_{\text{e}}(k)$ fundamental diagrams, all shock waves are *compression waves* going from lower to higher densities. However, for $q_{\text{e}}(k)$ fundamental diagrams that contain convex regions, application of the entropy condition can return the wrong solution [LeV01]. Although the mathematics of using these kinds of fundamental diagrams has been worked out, see for example the work of Li [Li03], a unified physical interpretation is still lacking [Mae04g; Nag05]: instead of only deceleration shock waves and acceleration fans, we now also have acceleration shock waves and deceleration fans. Finally, it is important to realise that for non-smooth $q_{\text{e}}(k)$ fundamental diagrams, the entropy condition is not applicable and no fans occur because the correct unique solution is automatically obtained [Vel02].

In Figure 3.7, we have depicted a classic example that is often used when illustrating the tempo-spatial evolution of a traffic flow at a traffic light (left part), based on the LWR first-order macroscopic traffic flow model with a triangular $q_{\text{e}}(k)$ fundamental diagram (right part). The application of the traffic flow model is visible in the time-space diagram to the left. A traffic light is located at position x_{light}; it is initially green, and at t_{red} it turns red until t_{green} when it switches back to green. The initial conditions at the road segment are located at point ① on the fundamental diagram. Because all characteristics of the solution are tangential to the fundamental diagram, the results can be elegantly visualised when using a triangular diagram: except for the fan of rarefaction waves (we approximate the non-differentiable tip of the triangle with a smooth one, such that we can show the fan ④ for all didactical intents and purposes), only two kinematic wave speeds are possible. When the traffic light turns red, a queue of stopped vehicles develops. Inside this queue, the jam density state k_{j} holds, corresponding to point ② on the fundamental diagram. The upstream boundary of the queue is demarcated by the shock wave ③ that is formed by the intersections of the characteristics ① and ②. Downstream of the jam, there are no vehicles: because we are working with a triangular fundamental diagram, the characteristics are parallel to the vehicle trajectories (their speeds are equal to the slopes of points on the free-flow branch). The initial regime at state ① and the 'empty' regime downstream of the queue are separated from each other by a *contact discontinuity* or *slip*. When the traffic light turns green again, the queue starts to dissipate, whereby the solution of characteristics becomes a fan of rarefaction waves ④, taking on all speeds between states ② and ① on the fundamental diagram[12]. A final important aspect that can be seen from Figure 3.7, is the fact that in the LWR model the outflow from a jam, i.e., going from a high to a low density regime, always proceeds via the capacity-flow regime at q_{cap}: so *there is no capacity drop in the LWR model* because the outflow is always optimal.

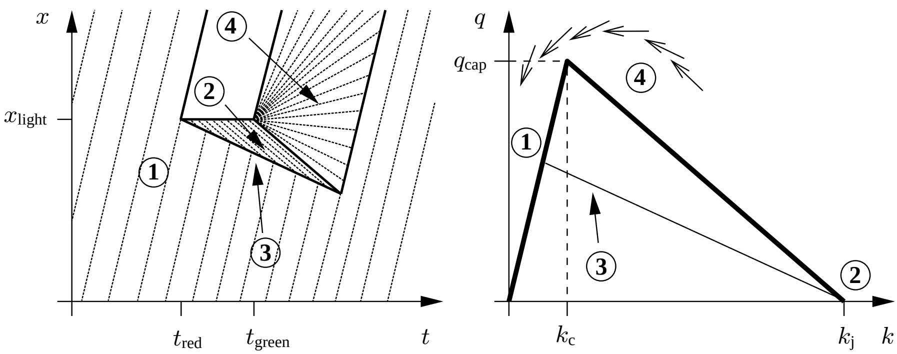


Figure 3.7: Example of an analytical solution based on the LWR first-order macroscopic traffic flow model with a triangular $q_{\text{e}}(k)$ fundamental diagram. *Left:* a time-space diagram with a traffic light located at position x_{light}. It is green, except during the period from t_{red} until t_{green}. The solution is visually sketched by means of the characteristics that evolve during the tempo-spatial evolution of the traffic flow. *Right:* a triangular fundamental diagram, with the initial conditions at state ①. When the traffic light is red, a queue develops in which the jam density state at point ② holds. Its upstream boundary is demarcated by the shock wave ③. When the queue starts to dissipate, the solution of characteristics generally becomes a fan of rarefaction waves ④.

To conclude our summary of analytical derivations, we point the reader to the significant work of Newell, who in 1993 cast the LWR theory in an elegant form. The key ideas he introduced were on the one hand the use of cumulative curves for deriving the conservation law, and on the other hand the use of a triangular $q_{\text{e}}(k)$ fundamental diagram [New93a]. Due to Newell's work, traffic flow analysis in this respect gets very simplified, as it is now possible to give an exact *graphical* solution to the LWR model for both free-flow and congested conditions [New93b]. To complete his theory, Newell also provided us with a means to include multi-destination flows, i.e., specifications of which off-ramp vehicles will use to exit the motorway [New93c]. Note that

[12] Note that at non-differentiable points, e.g., $(k_{\text{c}}, q_{\text{cap}})$, the fundamental diagram should be made continuously differentiable, such that the derivative $dq_{\text{e}}(k)/dk$ can take on all possible values between the left and right derivative.

for the LWR model with a parabolic $q_e(k)$ fundamental diagram and piece-wise linear and piece-wise constant space and time boundaries, respectively, Wong and Wong recently devised an exact analytical solution scheme. Their method is based on the efficient tracking and fitting of generated and dispersed shock waves within a time-space diagram [Won02].

3.2.1.4 Numerical solutions of the LWR model

Besides the previous *analytic derivation* of a solution to the conservation law expressed as a PDE, it is also possible to treat the problem *numerically*. By trying to find a numerical solution to the PDEs, we enter the field of *computational fluid dynamics* (CFD). In a typical setup, the 'fluid domain' is first discretised into adjacent cells (called a one-dimensional *mesh*) of size ΔX (note that all cells need not to be equal in size), after which an *iterative scheme* is used to update the cells' states (e.g., the density k in each fluid cell) at discrete time steps $m\ \Delta T$ with $m \in \mathbb{N}_0$. Typically, this entails *finite difference schemes* (or in a broader context, *finite element methods* or *finite volume methods*), which replace the continuous partial derivative with a *difference operator*, thereby transforming the conservation equation into a *finite difference equation* (FDE). Examples of these difference operators are the *forward difference operator* $\Delta f(x) = f(x+1) - f(x)$ and the *backward difference operator* $\nabla f(x) = f(x) - f(x-1)$, which is not to be confused with the gradient vector of $f(x)$. Examples of finite difference schemes are the *central scheme*, the *Lax-Friedrichs scheme*, the *downwind scheme*, the *upwind scheme*, the *MacCormack scheme*, the *Lax-Wendroff scheme*, the *Steger-Warming Flux Splitter scheme*, the *Rieman-based Harten-van Leer-Lax and Einfeldt scheme*, ... For a more complete overview of these schemes, we refer the reader to the work of Helbing and Treiber [Hel99d], Jüngel [Jün02], and Ngoduy et al. [Ngo03]. A practical software implementation of a *moving-mesh finite-volume solver* for the previously mentioned hyperbolic PDEs, can be found in van Dam's TraFlowPACK software [Dam02]. LeVeque also developed a numerical solver, called CLAWPACK, that is designed to compute numerical solutions to hyperbolic partial differential equations using a wave propagation approach [LeV03]. A central precaution for all these schemes, is the so-called *Courant-Friedrichs-Lewy (CFL) condition* which guarantees numerical stability of the algorithms; for traffic flows, it has the physical interpretation that no vehicles are allowed to 'skip' cells between consecutive time steps (i.e., all physical information that has an influence on the system's behaviour should be included):

$$\Delta T \leq \frac{\Delta X}{\overline{v}_{\text{ff}}}. \tag{3.13}$$

Just over a decade ago, Daganzo constructed a numerical scheme based on finite difference equations. It is known as the *cell transmission model* (CTM), which solves the LWR model using a trapezoidal $q_e(k)$ fundamental diagram [Dag94]. At the heart of his model lies a discretisation of the road into finite cells of width ΔX, each containing a certain number of vehicles (i.e., an average cell density). When time advances, these vehicles are transmitted from upstream to downstream cells, taking into account the capacity constraints imposed by the downstream cells. The CTM converges to the LWR model in the limit when $\Delta X \rightarrow 0$. In 1995, Daganzo extended the model to include network traffic, i.e., two-way merges and diverges, thereby allowing for the correct modelling of dynamic queue spill backs [Dag95a]. He also cast the model in the context of *Godunov FDE methods*[13], allowing for arbitrary $q_e(k)$ fundamental diagrams. The exchange of vehicles between neighbouring cells is then governed by so-called *sending* and *receiving functions* [Dag95b]. Lebacque derived a similar numerical scheme that performed the same functions as Daganzo's CTM, at approximately the same time (the debate on whomever was first is still not resolved) [Leb93; Leb96]. In his derivation, he employed the terms *demand* and *supply functions* to denote the exchange of vehicles between cells. He also provided the means to handle general (i.e., multi-way) merges and diverges. Both the original cell transmission model and an implementation of the Godunov scheme for the LWR model with an arbitrary $q_e(k)$ fundamental diagram, were provided by Daganzo et al. in the form of a software package called NETCELL [Cay97]. Note that, as mentioned earlier, numerical methods tend to smear out the shock waves; this diffusion is therefore a consequence of the solution methodology and not of the LWR model itself [Log03a].

Daganzo also developed another methodology for numerically solving the LWR equations, based on a *variational formulation*. Rather than extending the existing concept of a conservation equation coupled with a vanishing viscosity limit, he derived a solution based on the principles behind cumulative curves. The initial value problem becomes well-posed, and the methodology is able to handle complex boundary conditions. In short, the problem is transformed into finding shortest paths in a network of arcs that comprise the kinematic waves; as a surplus, the method is computationally more efficient than traditional solutions based on conservation laws [Dag03a; Dag05].

Traditional cell-based numerical methods are fairly computationally intensive, because they have to discretise the road entirely (even in regions where there is no variation in density), resulting in a solution that is composed of linear shock waves and continuous fans (i.e., the rarefaction waves). In order to derive a solution that is computationally more efficient, Henn proposed to replace the continuous fans of rarefaction waves with a discrete set of angular sectors (i.e., the density now varies with discrete steps). The efficiency now stems from the fact that, instead of a whole array of cells, only list structures need to be maintained [Hen03].

Only recently, a combination of Daganzo's CTM with a triangular $q_e(k)$ fundamental diagram and Newell's cumulative curves was constructed by Yperman et al., resulting in the *link transmission model* (LTM). Because whole links can be treated at once, the LTM's computational efficiency is much higher than that of classical numerical solution schemes for the LWR model, whilst retaining the same accuracy [Ype05b].

With respect to the applicability of the LWR model to real-life traffic flows, we refer the reader to two studies: the first was done by Lin and Ahanotu in the course of the

[13] Sergei K. Godunov's numerical solution of PDEs is considered as a breakthrough in computational fluid dynamics: it provides a unique solution based on a stable Riemann problem [God59; Leb96; Dag99a].

California Partners for Advanced Transit and Highways (PATH) programme (formerly known as the *Program on Advanced Technology for the Highway*). In their work, they performed a validation for the CTM with respect to the formation and dissipation of queues, concluding that the most important first-order characteristics (correlations in measurements of free-flow traffic at successive detector stations, and the speed of the backward propagating wave under congested conditions) perform reasonable well when comparing them to field data [Lin95].

A second, more thorough and critical study was done by Nagel and Nelson. In it, they scrutinise the LWR model, both with concave $q_e(k)$ fundamental diagrams and those with convex regions. Their main conclusion states that it remains difficult to judge the model's capabilities on a fair basis, largely due to the fact that there do not exist many real-world data sets which also contain a geometrical description of the local infrastructural road layout. This latter ingredient is a requirement for assessing whether or not an observed traffic breakdown is either spontaneously induced or due to the presence of an active bottleneck (because the LWR model constitutes a strictly deterministic model) [Nag05].

3.2.1.5 Flavours of the LWR model

Considering this elegant first-order traffic flow model, its main advantages are that it is simple, and in a sense reproduces the most important features of traffic flows (i.e., shock waves and rarefaction waves). However, because of its restriction to a first-order partial differential equation, certain other effects, such as stop-and-go traffic waves, capacity drop and hysteresis, traffic flow instabilities, finite acceleration capabilities, … can not be represented [Lig55]. In many cases, these 'deficiencies' can be tackled by switching to higher-order models, as will be elaborated upon in Section 3.2.1.6. Interestingly, in their original paper, Lighthill and Whitham recognised the fact that drivers tend to anticipate on downstream conditions, changing their speed gradually when crossing shock waves. This in fact necessitates a diffusion term in the conservation equation that captures a density gradient.

Instead of using a higher-order model, traffic flow engineers can also resort to more sophisticated approaches, such as extensions of the first-order model. To conclude this section, let us give a concise overview of some of the model flavours that have been proposed as straightforward extensions to the seminal LWR model.

An interesting set of extensions launched, was created by Daganzo, dealing with two classes of vehicles, of which one class can use all lanes of a motorway, whereas the other class is restricted to a right-hand subset of these lanes. When the capacity of the latter vehicles in regular lanes is exceeded, a queue will develop in those lanes, but the former vehicles will still be able to use the other lanes; this is called a *2-pipe* regime. Similarly, if the capacity of the yet freely flowing vehicles is exceeded, all lanes enter a queued state, which is called a *1-pipe* regime. In short, interactions between vehicles in this and the following models are nearly always considered from a user equilibrium perspective [Dag97a]. Daganzo et al. applied the theory to a case where there is a set of special lanes on which only priority vehicles can drive. The theory was also suited to describe congestion on a motorway diverge, such that the motorway itself can still be in the free-flow regime [Dag97c]. For the special case of queue spill back at a motorway's off-ramp, Newell also provided a graphical solution that is based on the use of cumulative curves [New99].

Continuing the previous train of thought, Daganzo provided a logical extension: he again considered different lanes, but now introduced two different types of drivers: aggressive ones (called *rabbits*) and timid ones (called *slugs*). Daganzo himself states that a correct traffic flow theory should involve both human psychology and lane-changing aspects, leading him to such a *behavioural* description [Dag02a]. The theory was also used to explain the phenomenon of a capacity funnel (see Section 2.5.5 for more details): according to the theory, once a capacity drop occurs, the recovery to the capacity-flow regime can not occur spontaneously, thereby requiring an exogeneous mechanism. Daganzo provides an explanation, called the *pumping phenomenon*: drivers temporarily accept shorter time headways downstream of an on-ramp, leading to a 'pumped state' of high-density and high-speed traffic, or in other words, a capacity-flow regime [Dag02b]. Chung and Cassidy later provided a validation of the theory, by applying it to describe merge bottlenecks on multi-lane motorways in Toronto (Canada) and Berkeley (California). In their study, they introduced the concept of *semi-congestion*, denoting a regime in which on vehicle class enters a state with a reduced mean speed, whereas the other vehicle class can still travel unimpeded. Their findings indicated an agreement between both shock wave speeds empirically observed and predicted by the model [Chu02].

An interesting case to which the LWR theory can be applied, is the problem of *moving bottlenecks* as stated by Gazis and Herman [Gaz92]. Examples of such bottlenecks are slower trucks on the right shoulder lanes, which can impede upstream traffic. Newell was among the first to try to give a satisfactory consistent treatment of this type of bottlenecks. The trick he used was to translate the problem into a moving coordinate system that is travelling at the bottleneck's velocity. This resulted in a description of a stationary bottleneck, after which the classical LWR theory can be applied [New98]. Although the theory is sound, there exist some serious drawbacks, mainly due to its underlying assumptions. For example, the moving bottlenecks are assumed to be long convoys, and other drivers' behaviours are not affected by the bottlenecks' speeds; even more serious is the fact that the theory is not valid for very light traffic conditions, and that several strange effects are predicted by the theory (e.g., a bottleneck with increasing speed can result in a lower upstream capacity). To this end, Muñoz and Daganzo applied the previously mentioned behavioural model with rabbits and slugs to the problem of a moving bottleneck. Their theory performs satisfactorily and agrees well with empirically observed motorway features. However, because of the fact that it relies on the LWR model, it is not entirely valid for bottlenecks that travel at high speeds under light traffic conditions. In this latter case, they state that driver differences are much more important than the dynamics dictated by the kinematic model [Muñ02a].

Another theory that deals with the problem of moving bottlenecks, is the one proposed by Daganzo and Laval: they treat moving bottlenecks as a sequence of consecutive *fixed obstructions* that have the same capacity restraining effects. Despite the fact that the previous theory of Muñoz and Daganzo has a good performance, it does not easily lend itself to discretisation schemes that allow numerical solutions. In contrast to this, the hybrid theory (fixed obstructions coupled with the LWR model dynamics) of Daganzo and Laval holds high promise as they have shown that it can be discretised in a numerically stable fashion [Dag03b]. As a continuation of this work, Lavel furthermore investigated the power of these fixed obstructions, allowing him to capture lane-changes as random events modelled by moving bottlenecks in a LWR 1-pipe regime. It is suggested that (disruptive) lane changes form the main cause for instabilities in a traffic stream. This leads the 'Berkeley school' to the statement that incorporating lane-change capabilities into multi-lane macroscopic models seems a prerequisite for observing effects such as capacity drops, kinematic waves of fast vehicles, ... [Lav04; Lav06] In this respect, Jin provides a theory that explicitly takes into account to effects of lane changes [Jin05]. The starting point in this model, is the presence of certain road areas in which traffic streams mix. The underlying assumption here is that all lane changes lead to the same traffic conditions in each lane: the crucial element in the model is that vehicles performing lane changes are temporarily counted twice in the density total. This new 'effective density', is then used to transform the $q_e(k)$ fundamental diagram, leading to a reversed lambda shape. However, because the current version of the theory employs a small artificial constant to introduce the lane changes, we question its practical applicability when it comes to calibration and validation.

To conclude this overview of first-order models, we highlight another successful attempt at increasing the capabilities of the classical LWR model. An important extension was made by Logghe, who derived a *multi-class formulation*[14] that allows for the correct modelling of heterogeneous traffic streams (e.g., preserving the FIFO property for interacting classes). Classes are distinguished by their maximum speed, vehicle length, and capacity (all intended for a triangular $q_{\text{e}}(k)$ fundamental diagram). A central ingredient to his theory, is the interactions between different user classes that reside on a road: in this respect, each class acts selfishly, with slower vehicles taking on the role of moving bottlenecks. Besides being able to construct analytical and graphical solutions, Logghe also provided a stable numerical scheme, as well as a complete network version with road inhomogeneities, and two-way merges and diverges [Log03a; Log03b].

> A finally important aspect that is mainly related to lane changes, is the *anisotropy property* of a traffic stream. This property basically states that drivers are not influenced by the presence upstream vehicles. In a sense, most models describing the acceleration behaviour of a vehicle, only take into account the state of the vehicle directly in front. For most macroscopic traffic flow models, this anisotropy constitutes a necessary ingredient. However, in his original paper, Richards very subtly points out that the fact of whether or not drivers only react to the conditions ahead, remains an open question [Ric56]. In contrast to this, Newell states that a driver is only influenced by downstream conditions, leading to a natural cause-and-effect relation, making the problem mathematically well-posed [New93b]. Recently, Zhang stated that the anisotropy property is likely to be violated in multi-lane traffic flows. His explanation is closely tied to the concavity character of a $q_{\text{e}}(k)$ fundamental diagram (non-concave regions can lead to characteristics that travel faster than the space-mean speed of the traffic stream). He also provides an intuitive reasoning based on Daganzo's rabbits and slugs, whereby *tailgating* vehicles induce slower downstream vehicles to 'make way'. Note that for single-lane traffic flows, the anisotropy property is expected to hold because of the FIFO property (vehicles can not pass each other), although there are exceptions in the case of some higher-order macroscopic traffic flow models [Zha03a]. Finally, Logghe refined this notion for heterogeneous traffic flows, whereby now the anisotropy condition remains valid for each vehicle class separately (i.e., vehicles are not influenced by *similar* upstream vehicles) [Log03a].

3.2.1.6 Higher-order models

The development of higher-order macroscopic models came as a response to the apparent shortcomings of the first-order LWR model. Harold Payne was among the first

[14] At approximately the same time, Chanut and Buisson constructed a first-order model that incorporates vehicles with different lengths and free-flow speeds [Cha03]. Their model can be considered as a trimmed-down version of Logghe's multi-class formulation.

in 1971 to develop such a higher-order model [Pay71]. In those days, ramp metering[15] control strategies were basically an all-empirical occasion. Payne recognised the necessity to include dynamic models in the control of on-ramps; the celebrated LWR model however, was found to perform unsatisfactorily with respect to the modelling of real-life traffic flows. One of these shortcomings, was the model's inability to generate stop-and-go waves. Zhang later traced this to be a consequence of the model's persistent reliance on a single equilibrium curve (i.e., the fundamental diagrams) [Zha03b]. In the LWR model, drivers are assumed to adapt their vehicle speed *instantaneously* according to the fundamental diagram when crossing a shock wave, a phenomenon termed the *no-memory effect* (i.e., they encounter infinite accelerations and decelerations [Zha98]). One option that leads to a solution of the previously mentioned problems, is to introduce different fundamental diagrams for vehicles driving under different traffic conditions; this avenue was explored by Newell [New63b] and Zhang [Zha99] (see also Section 2.5.3 for a connection with the capacity drop and hysteresis phenomena). Another, more popular type of solution was proposed by Payne (as well as by Whitham some years later [Whi74]): they suggested to add an equation to the LWR conservation law (3.6) and its fundamental diagram[16]. This new *dynamic speed equation* was derived from the classical car-following theories of Gazis et al. [Gaz59] (see also Section 3.2.3.1 for more details). An important aspect is this derivation, is the fact that the car-following model includes a *reaction time*, resulting in a *momentum equation* that relates the space-mean speed of a vehicle stream to its density. As a result, vehicles no longer instantaneously change their speed when crossing a shock wave. Payne's second-order macroscopic traffic flow model is now described by the following pair of PDEs, i.e., a conservation law and a momentum equation:

$$k_{\mathrm{t}} + (k\,\overline{v}_{\mathrm{s}})_{\mathrm{x}} = 0, \tag{3.14}$$

$$d\overline{v}_{\mathrm{s}} = \overline{v}_{\mathrm{s_t}} + \underbrace{\overline{v}_{\mathrm{s}}\overline{v}_{\mathrm{s_x}}}_{\text{convection}} = \underbrace{\frac{\overline{v}_{\mathrm{s_e}}(k) - \overline{v}_{\mathrm{s}}}{\tau}}_{\text{relaxation}} - \underbrace{\frac{c^2(k)}{k}k_{\mathrm{x}}}_{\text{anticipation}}, \tag{3.15}$$

with $\overline{v}_{\mathrm{s_t}}$ and $\overline{v}_{\mathrm{s_x}}$ denoting the partial derivatives of the space-mean speed with respect to time and space, respectively, $\overline{v}_{\mathrm{s_e}}$ the traditional fundamental diagram, and τ the *reaction time*. The function $c(k)$ corresponds to the model-dependent *sound speed of traffic* (i.e., the typical speed of a backward propagating kinematic shock wave); examples of $c(k)$ are [Zha03c]:

[15] Ramp metering is an ATMS whereby a traffic light is placed at an on-ramp, such that traffic enters the highway from the on-ramp by drops. We refer the reader to the work of Bellemans [Bel03] and Hegyi [Heg04] for an overview and some recent advancements.

[16] Note that Lighthill and Whitham originally proposed to extend the conservation law in their model with relaxation and diffusion terms, but the idea did not receive much thought at the time [Lig55].

$$-\sqrt{-\frac{1}{2\tau}\frac{d\overline{v}_{s_e}(k)}{dk}} \qquad \text{(Payne)} \tag{3.16}$$

$$-\sqrt{\frac{\nu_W}{\tau}} \qquad \text{(Whitham)} \tag{3.17}$$

$$k\,\frac{d\overline{v}_{s_e}(k)}{dk} \qquad \text{(Zhang)} \tag{3.18}$$

with ν_W being a parameter in equation (3.17).

In equation (3.15), the left hand side corresponds to the derivative of the speed, i.e., the acceleration of vehicles. As can be seen from the formulation, Payne identified three different aspects for the momentum equation: a *convection* term describing how the space-mean speed changes due to the arrival and departure of vehicles at the time-space location (t, x), a *relaxation* term describing how vehicles adapt their speed to the conditions dictated by the fundamental diagram, but with respect to a certain reaction time (as opposed to the instantaneous adaption in the LWR model), and finally an *anticipation* term describing how vehicles react to downstream traffic conditions.

In continuation of the above derivation, many other higher-order models have been based on the Payne-Whitham (PW) second-order traffic flow model. An example is the work of Phillips, who changed the reaction time τ in the relaxation term of equation (3.15) from a constant to a value that is dependent on the current density [Phi79]. Another example is due to Kühne, who artificially introduced a viscosity term into equation (3.15), in order to smooth the shock waves [Küh84]. The physical role that viscosity plays in a vehicular traffic stream is however not entirely understood: according to Zhang, the viscosity reflects the resistance of drivers against sharp changes in speeds [Zha03b]. For a rather complete overview of extensions to the PW model, we refer the reader to the work of Helbing [Hel01b].

3.2.1.7 Critiques on higher-order models

Higher-order models have been successfully applied in various computer simulations of traffic flows, e.g., the original FREFLO implementation by Payne [Pay78], the work of Kwon and Machalopoulos who developed KRONOS which is an FDE solver for a motorway corridor [Kwo95], the METANET model of Messmer and Papageorgiou [Mes90], the *Macroscopic Dynamic Assignment Model* (MaDAM) which is a similar Payne-type model and is used in the OmniTRANS software suite as mentioned in Section 3.1.2.4 [Ver03a], … Despite their success, it was Daganzo who in 1995 published their final requiem, which stood out as an obituary for all higher-order models [Dag95d]. From a theoretical perspective, there were some serious *physical flaws* that littered these second-order models. Most notably was the fact that there exist two families of characteristics (called *Mach lines*) in the Payne-Whitham type models. On the one hand, there are characteristics that imply a diffusion-like behaviour, which under certain circumstances can lead to negative speeds at the end of a queue, i.e., vehicles travelling backwards. On the other hand, there are characteristics that have the property of travelling faster than the propagation of traffic flow. This latter gas-like behaviour means that vehicles can get influenced by upstream conditions (because information is sent along the characteristics), which is a clear violation of the anisotropy property for single-lane traffic as explained in the previous section. From a physical point of view, the relaxation term in equation (3.15) may even introduce a 'suction process' because slower vehicles can get sucked along by leading faster ones [Heg01].

Several years after these critiques, Papageorgiou responded directly to the comments stated in Daganzo's article [Pap98]. In his response, Papageorgiou put a lot of emphasis at the incapabilities of first-order traffic flow models for use in a traffic control strategy (e.g., ramp metering). He very briefly reacts to the anisotropy violation, by mentioning that in multi-lane traffic flows, the space-mean speeds of the different lanes are not all the same, leading to characteristics that are *allowed* to travel faster than the space-mean speed of all lanes combined. With respect to negative speeds (and hence, negative flows), he proposes to simply include an a posteriori check that allows to set the negative flows equal to zero. One year later, in 1999, Heidemann reconsidered these higher-order models, but this time from the perspective of *mathematical flaws*. His main argument was the fact that the models led to an internal inconsistency, because they ignored some aspects related to the conservation law [Hei99]. However, after careful scrutiny, Zhang later refuted Heidemann's claims: the inconsistencies that plague the models are a result of the insistence on the universality of a conservation law and the imposing of arbitrary solutions. As a consequence, the Payne-Whitham type of models are mathematically consistent theories, although they may suffer from the aforementioned physical quirks [Zha03c].

Note that the dynamic speed equation (3.15), can also be cast in another form that is more closely related to a gas-kinetic analogy. With this in mind, we can rewrite the momentum equation as follows [Hoo01]:

$$d\overline{v}_{\text{s}} = \overline{v}_{\text{s}_\text{t}} + \underbrace{\overline{v}_{\text{s}}\overline{v}_{\text{s}_\text{x}}}_{\text{transport}} = \underbrace{\frac{\overline{v}_{\text{s}_\text{e}}(k) - \overline{v}_{\text{s}}}{\tau}}_{\text{relaxation}} - \underbrace{\frac{P_{\text{x}}}{k}}_{\text{pressure}} + \underbrace{\frac{\nu}{k}\,\overline{v}_{\text{s}_\text{xx}}}_{\text{viscosity}}, \tag{3.19}$$

with now P denoting the *traffic pressure* and ν the *kinematic traffic viscosity* (as introduced by Kühne [Küh84]). The convection term has been relabelled a *transport term*, describing the propagation of the speed profile with the speed of the vehicles. The *pressure term* reflects the change in space-mean speed due to arriving vehicles having different speeds, and the *viscosity term* reflects changes due to the 'friction' between different successive vehicles. The classical Payne model is obtained if we set $P = kc^2(k)$ and $\nu = 0$.

In contrast to Papageorgiou's response which did not provide a definite answer, Aw and Rascle carefully examined the reason why the PW model exhibited the strange phenomena indicated in Daganzo's requiem [Aw00]. The root cause of this behaviour can be traced back to the spatial derivative P_x of the pressure term in equation (3.19).

Their solution suggests to abandon the transport and relaxation terms, and replace the spatial derivative of the pressure P (which is a function of the density k) with a *convective (Lagrangian) derivative*, i.e., $D/Dt = \partial_{\mathrm{t}} + (\overline{v}_{\mathrm{s}} \cdot \nabla) = \partial_{\mathrm{t}} + \overline{v}_{\mathrm{s}}\partial_{\mathrm{x}}$, with $(\overline{v}_{\mathrm{s}} \cdot \nabla)$ called the *advective derivative term* [Pri05]:

$$(\overline{v}_{\mathrm{s}} + P(k))_{\mathrm{t}} + \overline{v}_{\mathrm{s}} \, (\overline{v}_{\mathrm{s}} + P(k))_{\mathrm{x}} = 0. \tag{3.20}$$

This new formulation allows to remedy all Daganzo's stated problems [Aw00]. Because of the somewhat limited character of their derivation of equation (3.20), Rascle add a relaxation term to the equation's right-hand side, and developed a numerically stable discretisation scheme, as well as showing convergence to the classical LWR model when the relaxation tends towards zero [Ras02].

To end our overview of higher-order models, we illustrate two other types. The first model is actually a *third-order model* created by Helbing. It is based on the two PDEs of the Payne-Whitham type models, but is extended with a third equation that describes the change in the *variance of the speed*, denoted by Θ [Hel96]. Helbing derived his equations using a gas-kinetic analogy, resulting in the following Navier-Stokes-like equation (it is typically encountered in the pressure term for P):

$$\Theta_{\mathrm{t}} + \overline{v}_{\mathrm{s}}\Theta_{\mathrm{x}} = \underbrace{\frac{2(\Theta_{\mathrm{e}}(k) - \Theta)}{\tau}}_{\text{relaxation}} + \underbrace{\frac{2}{k}\,\overline{v}_{\mathrm{s_x}}(\nu\,\overline{v}_{\mathrm{s_x}} - P)}_{\text{pressure}} + \underbrace{\frac{\nu}{k}\,\overline{v}_{\mathrm{s_{xx}}} + \frac{\kappa}{k}\,\Theta_{\mathrm{xx}}}_{\text{viscosity}}, \tag{3.21}$$

with now the equilibrium relation $\overline{v}_{\mathrm{s_e}}(k, \Theta)$ of equation (3.19) also depending on the speed variance Θ. In addition to the viscosity ν, the dynamic speed variance equation (3.21) also contains an equilibrium relation $\Theta_{\mathrm{e}}(k, \overline{v}_{\mathrm{s}})$ for the variance of the speed, and κ which is a *kinetic coefficient* that is related to the reaction time τ, the density k, and the speed variance Θ. For $\nu = \kappa = 0$, Helbing's model reduces to an inviscid Euler type model as explained in Section 3.2.1.1 [Kla96]. Whereas in the LWR model there is only one family of characteristics, and in the PW model there are two families, the Helbing model generates three different families of characteristics; this implies that small perturbations in the traffic flow propagate both with the traffic flow itself, as well as in upstream and downstream direction relative to this flow [Hoo01].

The second model we illustrate, is the *non-equilibrium model* of Zhang. Because of the relaxation terms in the Payne-Whitham equations, drivers initially tend to 'overshoot' the equilibrium speed as dictated by the $\overline{v}_{\mathrm{s_e}}(k)$ fundamental diagram. It takes a certain amount of time for them to adapt to their speed to the new traffic conditions (i.e., a change in density is accompanied by a *smooth* change in space-mean speed), after which they converge on the diagram. This latter aspect gives rise to the empirically observed scatter in the (k,q) phase space, leading Zhang to the terminology of 'non-equilibrium models' because of the deviation from the one-dimensional equilibrium fundamental diagram [Zha98].

In his model, Zhang considers equilibrium traffic to be a state in which $d\overline{v}_{\mathrm{s}}/dt = \partial k/\partial x = 0$. In similar spirit of Payne's theory, Zhang constructs his model using an equilibrium relation between density and space-mean speed (i.e., the fundamental diagram), a reaction time that allows relaxation, and an anticipation term that adjusts the space-mean speed to downstream traffic conditions. This results in a macroscopic model that contains equation (3.14) as the conservation law, as well as the following momentum equation:

$$d\overline{v}_{\mathrm{s}} = \overline{v}_{\mathrm{s}_{\mathrm{t}}} + \overline{v}_{\mathrm{s}}\overline{v}_{\mathrm{s}_{\mathrm{x}}} = \frac{\overline{v}_{\mathrm{s}_{\mathrm{e}}}(k) - \overline{v}_{\mathrm{s}}}{\tau} - k\left(\frac{d\overline{v}_{\mathrm{s}_{\mathrm{e}}}(k)}{dk}\right)^2 k_{\mathrm{x}}, \tag{3.22}$$

with the last anticipation term showing the dependence on the spatial change of the density. Zhang also complements the theory with a finite difference scheme that allows to solve the equations in a numerically stable fashion, based on an extension of the Godunov scheme that satisfies the entropy condition referred to in Section 3.2.1.3 [Zha01a].

Just as with the improved PW model of Aw and Rascle, this model alleviates Daganzo's stated problem of wrong-way travel, even though there are also two families of characteristics, travelling slower, respectively faster, than the space-mean speed of traffic. An important fact here is that for the slower characteristics, the associated shock waves and fans correspond perfectly to those of the first-order LWR model. However, the shock waves and fans associated with the faster family of characteristics can still violate the anisotropy property of traffic (although they decay exponentially), but in the end, Zhang questions its universal validity, stating that traffic might occasionally violate this principle due to the heterogeneity of a traffic stream [Zha00; Zha03a]. The violation of anisotropy, i.e., drivers get influenced by upstream traffic, is sometimes referred to as gas-like behaviour, because in contrast to fluid-dynamics, gas particles are not anisotropic. In an attempt to remove this faulty behaviour, Zhang developed yet another non-equilibrium model that removed the gas-like behaviour, thereby respecting the anisotropy property. Moreover, both families of characteristics in his model satisfy the condition of travelling at a lower speed than the space-mean speed of the traffic stream, but still keeping the one-to-one correspondence between the slower characteristics and those of the first-order LWR model. At present, it is however unclear if this new model can generate stop-and-go waves, although there are indications that it can because of the non-equilibrium transitions that can occur [Zha02].

Despite the significant progress that has been made on the front of higher-order macroscopic traffic flow models, the Berkeley school firmly holds its faith in first-order models and their extensions. Its main reason is because of the numerical solution schemes that are well developed and understood. This is not the case for higher-order models, as these contain other characteristics that complicate the finite difference schemes (because information is now carried both up- and downstream, and because their numerical schemes initially were flawed [Zha01a; Lav04]). Related to this critique, is the fact that in contrast to the first-order model, no analytical solutions exist for the higher-order models. Also note that a couple of years ago, Lebacque and Lesort provided a nice discussion of the then-existing macroscopic models [Leb99].

Another reason for sticking with first-order models, is because the school believes that first-order characteristics are sufficient for the description of traffic flows [Cas01; Win01]. Using a triangular $q_{\text{e}}(k)$ fundamental diagram that captures the most important traffic flow characteristics (i.e., the free-flow speed $\overline{v}_{\text{ff}}$, the capacity q_{cap}, the jam density k_{j}, and the backward kinematic wave speed w), results in a further elegance of the models.

3.2.2 Mesoscopic traffic flow models

The previous section dealt with macroscopic models that described traffic streams at an aggregated level, derived from a fluid-dynamic analogy. This section describes how traffic can be modelled at this aggregate level, but with special consideration for microscopic characteristics (e.g., driver behaviour). Because of the ambiguity that surrounds mesoscopic models, we first elucidate what is meant by the term mesoscopic (i.e., it is something between a macroscopic and a microscopic approach). In the sections thereafter, we zoom in on a derivation of mesoscopic models based on a gas-kinetic analogy. For an outstanding overview of gas-kinetic models, we refer the reader to the work of Tampère [Tam04a].

3.2.2.1 The different meanings of 'mesoscopic'

Considering the amount of literature that has been generated during the last few decades, it seems to us that there exists no unanimous consensus as to what exactly constitutes mesoscopic traffic flow models. In general, there are three popular approaches when it comes to mesoscopic models [Hoo01]:

- **Cluster models**

  When considering vehicles driving on a road, a popular method is to group nearby vehicles together with respect to one of their traffic flow characteristics, e.g., their space-mean speed. Instead of having to perform detailed updates of all vehicles' speeds and positions, the cluster approach allows to treat these vehicles as a set of groups (called *clusters*, *cells*, *packets*, or *macroparticles*); these groups are then propagated downstream without the need for explicit lane-changing manoeuvres (leading to the coalescing and splitting of colliding and separating groups).

Examples of this kind of models, are the *CONtinuous TRaffic Assignment Model* (CONTRAM) of Leonard et al. [Leo89; Tay03], the work of Ben-Akiva et al., called *Dynamic network assignment for the Management of Information to Travellers* (DynaMIT), which is based on a cell transmission model with a cell of a link containing a set of vehicles with identical speeds [BA96; BA98; Sun02], the *Mesoscopic Traffic Simulator* (MesoTS) of Yang, which allows fast predictions of future traffic states [Yan97], ...

- **Headway distribution models**

  This rather unknown and somewhat outdated class of models, places the emphasis on the probability distributions of time headways of successive vehicles (this aggregation makes them mesoscopic). Two popular examples are Buckley's semi-Poisson model [Buc68], and Branston's generalised queueing model [Bra76]. As clarified in the summary of Hoogendoorn and Bovy, the original versions of these headway distribution models assume homogeneous traffic flows and they are inadequate at describing the proper dynamics of traffic flows [Hoo01].

- **Gas-kinetic models**

  The third and most important characterisation of mesoscopic models comes from a gas-kinetic analogy. Because macroscopic models aim towards obeying the fundamental diagram (either instantaneously as in the first-order LWR model or through a relaxation process as in higher-order models), the focus there lies on the generation and dissipation of shock and rarefaction waves. As a consequence, more complex and non-linear dynamics such as the different traffic regimes, encountered in Sections 2.5.1 and 2.5.4, can not be reproduced. To remedy this, gas-kinetic models implicitly bridge the gap between microscopic driver behaviour and the aggregated macroscopic modelling approach.

In the next sections, we will first give an overview of the original gas-kinetic model as derived by Prigogine and Herman, after which we discuss some of the recent successful modifications that allow for heterogeneity in the traffic stream (i.e., multi-class modelling), as well as the inclusion of more specific driver behavioural characteristics.

3.2.2.2 Mesoscopic models considered from a gas-kinetic perspective

As opposed to the macroscopic traffic flows models that are derived from a conservation equation based on the Navier-Stokes equations, mesoscopic models can be derived from a gas-kinetic analogy. From individual driving behaviour (termed a microscopic approach), a macroscopic model is derived. The earliest model can be traced back to the work of the late Nobel laureate Ilya Prigogine, in cooperation with Frank

Andrews and Robert Herman [Pri60; Pri71]. A central component in their theory, is the concept of a phase-space density (PSD):

$$\widetilde{k}(t,x,\overline{v}_{\mathrm{s}}) = k(t,x)\;P(t,x,\overline{v}_{\mathrm{s}}), \tag{3.23}$$

in which $P(t,x,\overline{v}_{\mathrm{s}})$ denotes the *distribution* of the vehicles with space-mean speed $\overline{v}_{\mathrm{s}}$ at location x and time t; the concept of this distribution originated in Boltzmann's theory of gas dynamics. For the above density function, a kinetic conservation equation can be derived, looking as follows [Hel01b]:

$$\frac{d\widetilde{k}}{dt} = \widetilde{k}_{\mathrm{t}} + \overline{v}_{\mathrm{s}}\;\widetilde{k}_{\mathrm{x}} = \left(\widetilde{k}_{\mathrm{t}}\right)_{\mathrm{acc}} + \left(\widetilde{k}_{\mathrm{t}}\right)_{\mathrm{int}}, \tag{3.24}$$

with now the two terms on the right hand side denoting the *accelerations* of and *interactions* between the vehicles; they are also called *gains* and *losses*, *relaxation* and *slowing down*, or *continuous* and *discrete* terms, respectively [Kla96; Tam04a]. Equation (3.24) is called the *Prigogine-Herman kinetic model* and it actually describes three processes:

1. Similar to the macroscopic conservation equation, the term $\overline{v}_{\mathrm{s}}\;\widetilde{k}_{\mathrm{x}}$ describes a convective behaviour: arriving and departing vehicles cause a change in the distribution $\widetilde{k}$ of vehicle speeds.

2. The first term on the equation's right hand side, $\left(\widetilde{k}_{\mathrm{t}}\right)_{\mathrm{acc}}$, describes the acceleration behaviour of vehicles, which is assumed to be a *density-dependent relaxation* process of the speed distribution P of equation (3.23) towards some pre-specified target speed distribution P_0 (typically based on an equilibrium speed).

3. The second term on the equation's right hand side, $\left(\widetilde{k}_{\mathrm{t}}\right)_{\mathrm{int}}$, describes the interactions between vehicles, as fast vehicles either must slow down or overtake slower ones (hence implying inherently multi-lane traffic). The decision on when to either slow down or to overtake (which is assumed to be a *discrete event*), is governed by the probabilities $(1-\pi)$ and π, respectively. The interaction term is called a *collision equation*, in analogy with the physics of the Boltzmann equation (where the collision term describes the scattering of the gas molecules). Because there occur joint distributions in this latter equation (i.e., the probability of a faster vehicle encountering a slower one), a common assumption called *vehicular chaos* is used, which states that vehicles' speeds are uncorrelated, hence allowing to split the joint distribution.

More than a decade later, Paveri-Fontana criticised the assumption of vehicular chaos in the interaction term [Pav75]. He subsequently proposed an improved gas-kinetic model, in which he extended the phase-space density of equation (3.23) with a dependence on the desired speed v_{des}, i.e., $\widetilde{k}(t, x, \overline{v}_{\text{s}}, v_{\text{des}})$; in Prigogine's original derivation, this desired speed was incorrectly considered to be a property of the road, instead of being a driver-related property [Hel01b].

> An interesting property of the gas-kinetic modelling approach instigated by the seminal work of Prigogine, is that for densities beyond a certain critical density, Nelson and Sopasakis found that the model solutions split into two distinct families. The current hypothesis surrounding this phenomenon states that this corresponds to the widely observed data scatter in the empirically obtained (k,q) fundamental diagrams [Nel98].

3.2.2.3 Improvements to the mesoscopic modelling approach

Significant contributions to the gas-kinetic mesoscopic model have been sporadic; after the work of Paveri-Fontana, Nelson was among the first to tackle the computational complexity associated with the four-dimensional phase-space density $\widetilde{k}(t, x, \overline{v}_{\text{s}}, v_{\text{des}})$ [Nel95]. In his derivation, he reformulated the relaxation and interaction terms both as discrete events, based on a bimodal distribution of the vehicles' speeds (i.e., corresponding to stopped and moving vehicles). In contrast to the classical model which uses a relaxation process in the acceleration term, Nelson furthermore based his derivation on a microscopic behavioural model [Hel01b; Hoo01; Tam04a].

Building on the work of Nelson (which is, as he describes, just a first initial step towards constructing a suitable kinetic model), Wegener and Klar derived a kinetic model in similar spirit, based on a microscopic description of individual driver behaviour with respect to accelerations and lane changes. Attractive to their work, is the fact that they also pay attention to the numerical solutions of their model, with respect to the description of homogeneous traffic flows [Weg96].

Noting that the correct reproducing of traffic flow behaviour at moderate to higher densities still troubled the existing mesoscopic models, Helbing et al. explored an interesting avenue. Not only did they capture the effect that vehicles require a certain finite space (leading to an Enskog- instead of a Boltzmann-equation), they also generalised the interaction term of equation (3.24). This last method allowed them to dismiss the traditional assumption of vehicular chaos, i.e., they were now able to treat correlations between vehicles' speeds (which have a substantial influence at higher densities). The trick to obtain this behaviour, was to assume that drivers react to the downstream traffic conditions. This leads to the inclusion of non-local interaction (braking) term, and hence their model is referred to as the *non-local gas-kinetic traffic flow model* [Hel98a; Hel02a]. Interestingly, this non-locality can generate effects that are similar to the ones induced by viscosity/diffusion terms in macroscopic traffic flows models, causing smooth behaviour at density jumps [Hel01b]. The power of their model is also demonstrated as it is able to reproduce all traffic regimes listed in Sections 2.5.4 and 2.5.5.

Central to some of the recently proposed models, is the step process that transforms one model class into another. Starting from microscopic driver behavioural principles (e.g., accelerating, braking, ...), a mesoscopic model is deduced. This mesoscopic model can then be translated into an equivalent macroscopic one by applying the *method of moments*. This allows to obtain PDEs that describe the dynamic evolution of the density k, space-mean speed $\overline{v}_{\text{s}}$, and its variance Θ (an exception to this methodology is the previously mentioned model of Wegener and Klar that obtains dynamic solutions directly [Weg96]). As an example, Helbing et al. also devised a numerical scheme for their previously discussed model. It was implemented in a simulation package called MASTER [Hel01a].

Important progress was made by the work of Hoogendoorn et al., who extended the gas-kinetic traffic flow models with *multiple user classes*, in the sense that different classes of drivers have different desired speeds. In order to achieve this, they replaced the traditional phase-space density with a *multi-class phase-space density* (MUC-PSD). The kinetic conservation equation thus describes the tempo-spatial evolution of this MUC-PSD (i.e., the interactions between different user classes), after which an equivalent system of macroscopic model equations is derived. The generalisation power of their model is exemplified as the previously mentioned model of Helbing et al., which is just a special case, having only one class [Hoo99; Hoo00]. The developed multiclass gas-kinetic model is currently being integrated in a macroscopic simulation model for complete road networks, called HELENA, which allows prediction of future traffic states, and hence to assess the effectiveness of policy measures [Hoo02b; Hoo02a].

Recently, Waldeer derived a kinetic model that is based on the description of a driver's acceleration behaviour (as opposed to his *observed* speed behaviour). This novel approach attempts to alleviate the unrealistic jumps in speeds that are typically encountered in kinetic models. To this end, Waldeer extends the phase-space density even further, including a vehicle's acceleration in addition to its position, speed, and desired speed (leading to an even more complex system). Because now the acceleration is updated discretely, the speed will change continuously as a result [Wal04b]. Furthermore, Waldeer provided a numerical scheme for solving his model, by employing a Monte Carlo technique that is frequently used in non-equilibrium gas-kinetic theory [Wal04a].

To end this overview of gas-kinetic models, we mention the important work of Tampère, who significantly extended the previous modelling approaches [Tam04a; Tam04b]. In his work, he used the *generalised phase-space density* (as derived by Hoogendoorn [Hoo99]), which incorporates a dependency on the *traffic state* S (e.g., encompassing vehicles' speeds and their desired speeds). As it is an increasingly recognised fact that a complete traffic flow model should contain elements which describe the human behaviour (see for example the comments made by Daganzo [Dag02a; Dag02b]), Tampère proposes to include a driver's *activation level*. His *human-kinetic model* (HKM) is, just like that of Helbing et al. and Hoogendoorn, able to reproduce all known traffic regimes listed in Sections 2.5.4 and 2.5.5. Because of the dependency of the PSD on a behavioural parameter (i.e., the activation level that describes a driver's awareness of the governing traffic conditions), the model is well-suited to evaluate the applicability of *advanced driver assistance systems* (ADAS). As another illustrating example, the phenomenon of a capacity funnel (see Section 2.5.5 for a description) can be realistically explained and reproduced [Tam03]. However, despite the progress related to incorporating human behaviour into mathematical models for traffic flows, Tampère argues that most of the work can currently not be validated because there is no appropriate data yet available.

3.2.3 Microscopic traffic flow models

Having discussed both mesoscopic and macroscopic traffic flows models, we now arrive at the other end of the spectrum where the microscopic models reside. Whereas the former describe traffic operations on an aggregate scale, the latter kind is based on the explicit consideration of the *interactions between individual vehicles* within a traffic stream. The models typically employ characteristics such as vehicle lengths, speeds, accelerations, and time and space headways, vehicle and engine capabilities, as explained in Section 2.2, as well as some rudimentary human characteristics that describe the driving behaviour.

The material in this section is organised as follows: we first introduce the classical car-following (and lane-changing) models as well as some of their modern successors, after which we discuss the optimal velocity model, then introduce the more human behaviourally psycho-physiological spacing models, which are subsequently followed by a brief description of traffic cellular automata models (this latter family of models will be elaborated upon in Chapter 4). After some words on models based on queueing theory, the section concludes with a concise overview of some of the (commercially) available microscopic traffic flow simulators, as well as some of the issues that are related to the calibration and validation of microscopic traffic flow models.

More detailed information with respect to microscopic models (more specifically, car-following models), can be found in the book of May [May90], the overview of Rothery [Gar97], the work of Ahmed [Ahm99], and the overview of Brackstone and McDonald [Bra00].

3.2.3.1 Classical car-following and lane-changing models

Probably the most widely known class of microscopic traffic flow models is the so-called family of *car-following* or *follow-the-leader* models. One of the oldest 'models' in this case, is the one due to Reuschel [Reu50], Pipes [Pip53], and Forbes et al. [For58]. It is probably best known as the *"two-second rule"* taught in driving schools everywhere[17] . An earlier example of this line of reasoning is the work of Herrey and Herrey, who specified a *safe driving distance* that also included the distance needed to come to a full stop [Her45].

[17] Note that, in his article, Pipes actually stated his safe-distance rule as keeping at least a space gap equal to a vehicle length for every 15 km/h of speed you are travelling at [Pip53; Hoo01].

It still remains astonishing that the seemingly daunting and complex task that encompasses driving a vehicle, can be executed with such relative ease and little exercise, as is testified by the many millions of kilometres that are driven each year. In spite of this remark, the first mathematical car-following models that have been developed, were based on a description of the interaction between two neighbouring vehicles in a traffic stream, i.e., a follower and its leader. In this section, we historically sketch the development of car-following theories, as they evolved from conclusions about early experiments into more sophisticated models.

The above mentioned model was originally formulated as the following *ordinary differential equation* (ODE) for single-lane traffic:

$$\frac{dv_i(t)}{dt} = \frac{v_{i+1}(t) - v_i(t)}{T_\mathrm{r}}, \tag{3.25}$$

with $v_i(t)$ and $v_{i+1}(t)$ the speeds of the following, respectively leading, vehicle at time t, and T_r a relaxation parameter. For the above case, the underlying assumption/justification is that vehicle i (the follower) tries to achieve the speed $v_{i+1}(t)$ of vehicle $i + 1$ (its leader), whilst taking a certain relaxation time T_r into account.

As equation (3.25) describes a stable system, Chandler et al. were among the first to include an explicit *reaction time* τ into the model (e.g., $\tau =$1.5 s), leading to destabilisation of vehicle platoons [Cha58]. This reaction encompasses both a *perception-reaction time* (PRT), i.e., the driver sees an event occurring (for example the brake lights of the leading vehicle), as well as a *movement time* (MT), i.e., the driver needs to take action by applying pressure to the vehicle's brake pedal [Gar97]. Introducing this behaviour, resulted in what is called a *stimulus-response model*, whereby the right-hand side of equation (3.25) describes the stimulus and the left-hand side the response (the response is frequently identified as the acceleration, i.e., the actions a driver takes by pushing the acceleration or brake pedal). The relaxation parameter is then reciprocally reformulated as the sensitivity to the stimulus, i.e., $\lambda = T_\mathrm{r}^{-1}$, resulting in the following expression:

$$\begin{aligned} \text{response} &= \text{sensitivity} \times \text{stimulus} \\ \frac{dv_i(t+\tau)}{dt} &= \lambda\,(v_{i+1}(t) - v_i(t)). \end{aligned} \tag{3.26}$$

Additional to this theoretical work, there were also some early controlled car-following experiments, e.g., the ones done by Kometani and Sasaki, who add a non-zero acceleration term to the right-hand side of the stimulus-response relation, in order to describe collision-free driving based on a safety distance [Kom58; Kom61].

Equation (3.26) is called a *delayed differential equation* (DDE), which, in this case, is known to behave in an unstable manner, even resulting in collisions under certain initial conditions. Gazis et al. remedied this situation by making the stimulus λ dependent on the distance, i.e., the space gap g_{s_i} between both vehicles [Gaz59]:

$$\frac{dv_i(t+\tau)}{dt} = \lambda \, \frac{v_{i+1}(t) - v_i(t)}{x_{i+1}(t) - x_i(t)}. \tag{3.27}$$

Further advancements to this car-following model were made by Edie, who introduced the current speed of the following vehicle [Edi61]. Gazis et al, forming the club of people working at General Motors' research laboratories, generalised the above set of models into what is called the *General Motors non-linear model* or the *Gazis-Herman-Rothery (GHR) model* [Gaz61]:

$$\frac{dv_i(t+\tau)}{dt} = \lambda \, v_i^m(t) \frac{v_{i+1}(t) - v_i(t)}{(x_{i+1}(t) - x_i(t))^l}, \tag{3.28}$$

with now λ, l, and m model parameters (in the early days, the model was also called the *L&M model* [Gaz02]). For a good overview of the different combinations of parameters attributed to the resulting models, we refer the reader to the book of May [May90], and the work of Ahmed [Ahm99].

A recent extension to the classical car-following theory, is the work of Treiber and Helbing, who developed the *intelligent driver model* (IDM). Its governing equation is the following [Tre99; Tre00; Tre01]:

$$\frac{dv_i}{dt} = a_{\text{max}} \left[\underbrace{1 - \left(\frac{v_i}{v_{\text{des}}} \right)^{\delta}}_{\text{acceleration}} - \underbrace{\left(\frac{g_s^*(v_i, \Delta v_i)}{g_{s_i}} \right)^2}_{\text{deceleration}} \right], \tag{3.29}$$

with a_{max} the maximum acceleration, v_{des} the vehicles' desired speed, and Δv_i the speed difference with the leading vehicle (we have dropped the dependencies on time t for the sake of visual clarity). The first terms within the brackets denote the tendency of a vehicle to accelerate on a free road, whereas the last term is used to allow braking in order to avoid a collision (the effective desired space gap $g_s^*(v_i, \Delta v_i)$ is based on the vehicle's speed, its relative speed with respect to its leader, a comfortable maximum deceleration, a desired time headway, and a jam space gap). The finer qualities of the IDM are that it elegantly generalises most existing car-following models, and that it has an explicit link with the non-local gas-kinetic mesoscopic model discussed in Section 3.2.2.3 [Tre00]. It is furthermore quite capable of generating all known traffic regimes, encountered in Sections 2.5.1 and 2.5.4 [Tre03]. Based on the IDM, Treiber et al. also constructed the *human driver model* (HDM), which includes a finite reaction time, estimation errors, temporal and spatial anticipation, and adaptation to the global traffic situation [Tre05a].

Similar to the work of Kometani and Sasaki, Gipps proposed a car-following model based on a safe braking distance, leading to collision-free dynamics [Gip81]. The model is interesting because no differential equations are involved (i.e., the speeds are computed directly from one discrete time step to another), and because it can capture underestimation and overreactions of drivers, which can lead to traffic flow instabilities. In similar spirit of Gipps' work, Krauß developed a model that is based on assumptions about general properties of traffic flows, as well as typical acceleration and deceleration capabilities of vehicles. Fundamental to his approach, is that all vehicles strive for collision-free driving, resulting in a model that has the ability to generalise most known car-following models [Kra97a; Kra98].

Another example of a recently proposed car-following model, is the 'simple' model of Newell, who formulates his theory in terms of vehicle trajectories whereby the trajectory of a following vehicle is essentially the same as that of its leader[18]. Remarkable properties are that the model has no driver reaction time, and that it corresponds to the first-order macroscopic LWR traffic flow model with a triangular $q_e(k)$ fundamental diagram (see Section 3.2.1.3) [New02b]. The model furthermore also agrees quite well with empirical observations made at a signallised intersection, which support the model and consequently also the first-order macroscopic LWR model [Ahn04].

As a final example, we briefly illustrate Zhang's car-following theory which is based on a multi-phase vehicular traffic flow. This means that the model is able to reproduce both the capacity drop and hysteresis phenomena (see Section 2.5.3), because his theory is based on the asymmetry between acceleration and deceleration characteristics of vehicles. The model also holds a generalisation strength, as it is possible to derive all other classical car-following models [Zha05].

With respect to the stability of the car-following models, there exist two criteria, i.e., *local* and *asymptotic stability* (also called *string stability*). The former describes how initial disturbances in the behaviour of a leading vehicle affect a following vehicle, whereas the latter is used to denote the stability of a *platoon of following vehicles*. By such a stable platoon it is then meant that initial finite disturbances exponentially die out along the platoon. Early experiments by Herman et al. already considered these criteria for both real-life as for the developed mathematical car-following models [Her59].

As an example, we graphically illustrate in Figure 3.8 the asymptotic stability of a platoon of some 10 identical vehicles. We have used the simple car-following model of Gazis et al. of equation (3.27) to describe how a following vehicle changes his acceleration, based on the speed difference and space gap with its direct leader in the platoon (the sensitivity λ was set to 5000 m/s^2 with a reaction time $\tau = 1$ s). The left part of the figure shows all the vehicles' positions, whereas the middle and right parts show the speeds and accelerations of the 2$^{\text{nd}}$, the 5$^{\text{th}}$, and the 10$^{\text{th}}$ vehicle respectively. We can see that all vehicles are initially at rest (homogeneously spaced), after which the leading vehicle applies an acceleration of 1 m/s^2, decelerates with -1 m/s^2, and then comes to a full stop. As can be seen, the first 4 following vehicles mimic the leader's behaviour rather well, but from the 5$^{\text{th}}$ following vehicle on, an instability starts to form (note that all following vehicles suffer from oscillations in their accel-

[18] A similar model was proposed earlier by Helly [Hel61; Ger64; New02b].

eration behaviour). This instability grows and leads to very large accelerations for the last vehicle, which even momentarily reaches a negative speed of some -150 km/h; this is clearly unrealistic (the vehicles shouldn't be driving backwards on the road), indicating that the specified car-following model is unsuitable to capture the realistic behaviour of drivers under these circumstances.

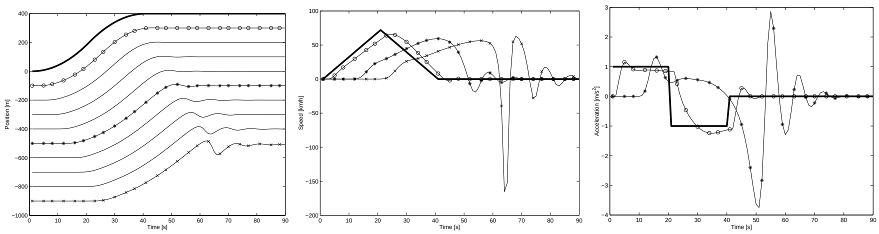


Figure 3.8: An example based on Gazis et al.'s car-following model of equation (3.27), indicating an asymptotically unstable platoon of 10 vehicles. *Left:* the time-space trajectories of all ten vehicles (the leading vehicle is shown with a thick solid line). We can see an instability occurring at the 6th ($\star$) vehicle, growing severely such that the last vehicle even has to drive backwards at a high speed of -150 km/h. *Middle:* the speeds of the first, the 2nd ($\circ$), the 6th ($\star$), and the last vehicle ($\times$). Note the oscillations and negative values for the speeds of the vehicles at the end of the platoon. *Right:* the accelerations of the first, the 2nd ($\circ$), and the 6th ($\star$) vehicle (example based on [Imm98b]).

In continuation of this small excerpt on stability, we refer the reader to the work of Zhang and Jarrett who analytically and numerically derive the general stability conditions (in function of the reaction time and the sensitivity to the stimulus) for the previously mentioned classical car-following models [Zha97], the work of Holland who derives general stability conditions and validates them with empirical data containing non-identical drivers (i.e., aggressive and timid ones); central to Holland's work is the source for instability with respect to a breakdown of a traffic flow. He relates this event to a so-called anticipation time that describes the duration for a wave containing an instability to travel to the current driver [Hol98]. Finally, we mention that stability analysis is of paramount importance for, e.g., automated vehicle technologies ('smart cars') such as *intelligent* or *adaptive cruise control* (ICC/ACC), as in for example the platooning experiments in the PATH project where a platoon of vehicles autonomously drives close to each other at high speeds.

To conclude this section, we shed some light on the typical mechanisms behind lane-changing models. With respect to microscopic models for multi-lane traffic, it is a frequent approximation to only take lateral movements *between* neighbouring lanes into account (as opposed to the *within-lane* lateral dynamics of a vehicle). In such cases, a vehicle changes a lane based on an incentive: these lane changes can then be classified as being *discretionary* (e.g., to overtake a slower vehicle), or *mandatory* (e.g., to take an off-ramp). When a vehicle (i.e., driver) has decided to perform a lane change, a check is made on whether or not it is physically possible to merge in to the adjacent lane (note this lane changing process also describes vehicles turning at street intersections). This latter process is called the *gap acceptance* behaviour: if there is no such possibility (as it is frequently the case in dense traffic), a driver may initiate at *forced merging*, in which case the following vehicle in the target lane might have to *yield*. This interaction between forced merging and yielding can be frequently observed at on-ramps where heavy duty vehicles enter the motorway. Although it seems intuitive that there is an asymmetry between the frontal and backward space gaps in the target lane (i.e., the former is usually smaller than the latter due to the human behaviour associated with forced merging and yielding), there is in our opinion nevertheless not enough empirical data available to calibrate the microscopic models that describe lane-changing (see for example the work of Ahmed [Ahm99]). One way to obtain a correct behaviour is to use a kind of a black box approach, in which for example the downstream capacity of a motorway section is used as a measure for calibrating the interactions (i.e., lane changes) between vehicles in a traffic stream. Note that as technology advances, new detailed data sets are constructed. An example is the work of Hoogendoorn et al. who use a *remote sensing* technique to capture vehicle trajectories based on aerial filming of driving behaviour under congested conditions [Hoo03].

3.2.3.2 Optimal velocity models

Closely related to the previously discussed classical car-following models, are the so-called *optimal velocity models* (OVM) of Newell and Bando et al. Whereas the previous car-following models mostly describe the behaviour of a vehicle that is following a leader, the OVMs modify the acceleration mechanism, such that a vehicle's desired speed is selected on the basis of its space headway, instead of only considering the speed of the leading vehicle [Hel01b]. Newell was the first to suggest such an approach, using an equilibrium relation for the desired speed as a function of its space headway (e.g., the $\overline{v}_{\text{s}}(\overline{h}_{\text{s}})$ fundamental diagram of Section 2.5.2.2) [New63a].

Bando et al. later improved this model, resulting in the following equation that describes a vehicle's acceleration behaviour [Ban95]:

$$\frac{dv_i(t)}{dt} = \alpha \left(V(h_{\text{s}_i}(t)) - v_i(t) \right), \tag{3.30}$$

in which $V()$ is called the *optimal velocity function* (OVF). The difference between this desired speed, associated with the driver's current space headway, and the vehicle's current speed, is corrected with an acceleration $\alpha V()$, with now α a coefficient expressing the sensitivity of a driver. This sensitivity corresponds to the inverse of the relaxation time, which is the time needed to reach the speed dictated by the OVF. Specification of the optimal velocity function (typically a sigmoid function such as $\tanh$) is done such that it is zero for $h_{\text{s}_i} \to 0$, and bounded to v_{max} for $h_{\text{s}_i} \to +\infty$; this latter condition means that the model is able to describe the acceleration of vehicles without the explicit need for a leader as in the previous car-following models.

Interestingly, the OVM requires, in contrast to the classical car-following models, no need for a reaction time in order to obtain spontaneous clustering of vehicles [Kra98]. Unfortunately, the model is not always free of collisions, and can result in unrealistically large accelerations [Nag03a].

3.2.3.3 Psycho-physiological spacing models

Instead of using continuous changes in space gaps and relative speeds, it was already recognised in the early sixties that drivers respond to certain *perception thresholds* [Bra00]. For example, a leading vehicle that is looming in front of a follower, will be perceived as having approximately the same small size for a large duration, but once the space gap has shrunk to a certain size, the size of the looming vehicle will suddenly seem a lot bigger (i.e., like crossing a threshold), inducing the following vehicle to either slow down or overtake.

The underlying thresholds with respect to speeds, speed differences, and space gaps, were cast into a model by the work of Wiedemann et al. [Wie74]. In this respect, the models are called *psycho-physiological spacing models*, and although they seem quite successful in explaining the traffic dynamics from a behavioural point of view (even lane-change dynamics can be included based on suitable perception thresholds), calibration of the models has nevertheless been a difficult issue [Bra00].

3.2.3.4 Traffic cellular automata models

In the field of traffic flow modelling, microscopic traffic simulation has always been regarded as a time consuming, complex process involving detailed models that describe the behaviour of individual vehicles. Approximately a decade ago, however, new microscopic models were being developed, based on the *cellular automata* programming paradigm from *statistical physics*. The main advantage was an *efficient and fast performance* when used in computer simulations, due to their rather low accuracy on a microscopic scale. These so-called *traffic cellular automata* (TCA) are dynamical systems that are discrete in nature, in the sense that time advances with discrete steps and space is coarse-grained (e.g., the road is discretised into cells of 7.5 metres wide, each cell being empty or containing a vehicle). This coarse-graininess is fundamentally different from the usual microscopic models, which adopt a semi-continuous space, formed by the usage of IEEE floating-point numbers. TCA models are very flexible and powerful, in that they are also able to capture all previously mentioned basic phenomena that occur in traffic flows [Bar99; Cho00; Mah05]. In a larger setting, these models describe *self-driven, many-particle systems, operating far from equilibrium*. And in contrast to strictly gaseous analogies, the particles in these systems are intelligent and able to learn from past experience, thereby opening the door to the incorporation of behavioural and psychological aspects [Cho99a; Wol99; Hel01b; Mah05].

The cellular automata approach proved to be quite useful, not only in the field of vehicular traffic flow modelling, but also in other fields such as pedestrian behaviour, escape and panic dynamics, the spreading of forest fires, population growth and migrations, cloud formation, material properties (corrosion, cracks, creases, peeling et cetera), ant colonies and pheromone trails, ... [Hel99c; Kar97; Nag92a; Gob01; Nis03a; Cho05]. It is now feasible to simulate large systems containing many 'intelligent particles', such that is it possible to observe their interactions, collective behaviour, self-organisation, ... [Imm98a; Zuy99; Hel99c; Hel01b; Hel04; Nag02a; Nag02b; Cho04]

Because the subsequent parts in the remainder of this dissertation are mainly focussed on the use of cellular automata models for traffic flows, we refer the reader to Chapter 4 for a more complete overview of the existing TCA models.

3.2.3.5 Models based on queueing theory

In this final section dealing with types of microscopic traffic flow models, we briefly summarise some of the models that are based on the paradigm of queueing theory. Early applications of queueing theory to the field of transportation engineering are mostly related to descriptions of the behaviour signallised and unsignallised intersections, overtaking on two-lane roads with opposing traffic, ... [Cle64]. Later, Vickrey introduced the bottleneck model which actually is a point-queue model, as explained in Sections 3.1.2.5 and 3.1.4.2 [Vic69]. Another more theoretically oriented application can be traced back to the work of Newell, who gives a nice summary of the mathematical details related to the *practical* application of the methodology. Newell was one of the few people who directly questioned the usefulness of cleverly devising a lot of methods and solutions, whereby corresponding problems remained absent [New82]. In his later work, Newell reintroduced the concept of arrival and departure functions (i.e., the cumulative curves as described in Section 2.3.2.2), giving an analytical but still highly intuitive method for solving traffic flow problems, and drawing parallels with the well-known and studied first-order macroscopic LWR model, thereby linking both model classes [New93a; New93b; New93c].

During the mid-nineties, Heidemann developed several queueing-based traffic flow models, of which the most powerful version deals with non-stationary conditions and is able to model the capacity drop and hysteresis phenomena, as well as providing an explanation for the wide scatter observed in empirical fundamental diagrams [Hei01].

Central to the approach in this field, is the partitioning of a road into equal pieces of width $1/k_{\text{jam}}$. Each of these pieces is then considered as a *service station* operating with a *service rate* $\mu = k_{\text{jam}} \cdot \overline{v}_{\text{ff}}$. Equivalently, vehicles arrive at each service station with an *arrival rate* $\lambda = k \cdot \overline{v}_{\text{ff}}$, with the assumption that k is the prevailing density and that traffic can flow unimpeded in the free-flow traffic regime. When vehicles enter the motorway, they can get stuck inside the queues, thereby reducing the space-mean speed in the system. Different *queueing policies* can be specified in the form of service and arrival distributions. In queueing theory, the *Kendall notation* is adopted, whereby a system is described as $A/S/m$ with A the arrival distribution, S the service distribution, and m the number of servers (i.e., service stations). Typical forms are the $M/M/1$ queues that have an exponentially distributed arrival time, exponentially distributed service time, and one server (with an infinite buffer).

Recently, Van Woensel extended the existing queueing models for traffic flows, leading to, e.g., analytical derivations of fundamental diagrams based on $G/G/m$ queues that have general distributions for the arrival and service rates with multiple servers [Woe03]. The methodology also includes queues with finite buffers, and has been applied to the estimation of emissions, although we question the validity of this latter approach (which is essentially based on a one-dimensional fundamental diagram) as we believe dynamic models are necessary, e.g., to capture transients in traffic flows [Van00].

Queue-based models were also used to describe large-scale traffic systems, e.g., complete countries, as was mentioned in Section 3.1.2.5 [Cet03]. In that section, we already mentioned that queues with finite buffer capacities are to be preferred in order to correctly model queue spill back. However, with respect to a proper description of traffic flow phenomena, some of the problems can not be so easily solved, e.g., the speed of a backward propagating kinematic shock wave. Take for example vehicles queued behind each other at a traffic light: once the light turns green, the first-order macroscopic LWR model correctly shows the dispersal of this queue. In a queue-based model however, once a vehicle exits the front of the queue, all vehicles simultaneously and instantly move up one place, thus the kinematic wave propagates backwards at an infinite speed !

To conclude this short summary on queueing models, we mention the work of Júlvez and Boel, who present a similar approach, based on the use of *Petri nets*[19]. Their work allows them to construct complete urban networks, based on the joining together of short sections; they employ continuous Petri nets for the propagation of traffic flows, and discrete Petri nets for the description of the traffic lights [Bas04; Júl05].

3.2.3.6 Microscopic traffic flow simulators

In continuation of the previous sections that gave an overview of the different types of existing microscopic traffic flow models, this section introduces some of the computer implementations that have been built around these models. In most cases, the computer simulators incorporate the car-following and lane-changing processes as submodels, as opposed to strategic and operational modules that work at a higher-level layer (i.e., route choice, . . .).

Whereas most microscopic traffic simulators allow to build a road network, specify travel demands (e.g., by means of OD tables), there was quite some effort spent over the last decade, in order to achieve a qualitative visualisation (e.g., complete virtual environments with trees, buildings, pedestrians, bicycles, . . . An example of such a virtual environment is shown in Figure 3.9, which is based on VISSIM's visualisation

[19] Petri nets (invented in the sixties by Carl Adam Petri) are a formalism for describing discrete systems [Pet62]; they consist of directed graphs of 'transitions' and 'places', with arcs forming the connections between them. Places can contain 'tokens', which can be 'consumed' when a transition 'fires'.

module. Note that in our opinion, the usefulness of these virtual scenes should not be underestimated, as in some cases a project's approval might hinge on a good visual representation of the results. It is one thing for policy makers to judge the effects of replacing a signallised intersection with a roundabout, based on a report of the observed downstream flows of each intersection arm, but it gives a whole other feeling when they are able to *see* how the traffic streams will interact ! Even in the early sixties, it was recognised that a visual representation of the underlying traffic flow process, was an undeniable fact for promoting its acceptance among traffic engineers [Ger64]. With respect to this latest comment, Lieberman even states that *"There is a need to view vehicle animation displays, to gain an understanding of how the system is behaving, in order to explain why the resulting statistics were produced"* [Gar97].

Figure 3.9: A screenshot of the VISSIM microscopic traffic flow simulator, showing a detailed virtual environment containing trees, buildings, pedestrians, … (image reproduced after [PTV05]).

Quite a large amount of microscopic traffic flow models have been developed, in most cases starting from a research tool, and — by the law of profit — naturally evolving into full-blown commercial packages, including, e.g., dynamic traffic assignment and other transportation planning features. Note the sad observation that this commercialisation inherently tends to obscure the underlying models. In such cases, privacy concerns, company policies, and project contracts and agreements prohibit a total disclosure of the mathematical details involved. Some of these computer models are listed here. For starters, the *Generic Environment for TRaffic Analysis and Modeling* (GETRAM) couples the multi-modal traffic assignment model EMME/2 to the *Advanced Interactive Microscopic Simulator for Urban and Non-Urban Networks* (AIMSUN2) model [Bar02a]. Next, the *Parallel microscopic traffic simulator* (Paramics), initially developed at the Edinburgh Parallel Computing Centre, but afterwards bought by Quadstone [Cam94; Lim00]. Subsequently, Yang developed a *MIcroscopic Traffic flow SIMulator* (MITSIM) [Yan97], and Maerivoet constructed a *MIcroscopic TRAffic flow SIMulator* (Mitrasim *2000*) which was mostly based on and influenced by MITSIM's and Paramics' dynamic behaviour [Mae01b]. Two further examples are the Open Source Software (OSS) package called *Simulation of Urban MObility* (SUMO), developed at the Deutsches Zentrum für Luft- und Raumfahrt [Kra04], and the *Verkehr in Stadten SIMulation* (VISSIM) programme developed by the German PTV group [PTV05]. In addition, there is the INTEGRATION software package developed by Van Aerde et al. This latter simulator deserves a special mention: it is microscopic in nature, but the speeds of the vehicles that are propagated through the network, are based on a macroscopic $\overline{v}_{s_e}(\overline{h}_s)$ fundamental diagram for each link [Aer96]. Finally, we mention the *TRansportation ANalysis and SIMulation System* (TRANSIMS) project [Nag98c], ...

An extensive overview of all existing microscopic traffic flow simulators until 1998 is provided by the *Simulation Modelling Applied to Road Transport European Scheme Tests*, or better known as the SMARTEST report [Alg98].

> When using one of these microscopic simulators, it is important to understand the assumptions and limitations inherent to the implemented models, in order to judge the results objectively. Indeed, as with any model, the question on whether some observed behaviour arises due to the implemented model, or as a result of the imposed boundary conditions, should always be asked, understood, and answered.

3.2.4 Submicroscopic traffic flow models

As the level of modelling detail is increased, we enter the realm of submicroscopic models. Traditional microscopic models describe vehicles as single operating units, putting emphasis on the interactions between different (successive) vehicles. In addition to this, submicroscopic models push the boundaries even further, giving detailed descriptions of a vehicle's inner workings. This typically entails modelling of the *physical characteristics* such as engine performance, detailed gearbox operations, acceleration, braking, and steering manoeuvres, ... Complementary to the functioning of a vehicle's physical components, submicroscopic models can also describe a *human driver's decision taking process* in much more detail than is usually done. Some examples of submicroscopic models are:

- van Arem's *Microscopic model for Simulation of Intelligent Cruise Control* MIXIC: it contains a driver model (for deciding on and executing of lane changes, car-following behaviour, and the application of intelligent (or adaptive) cruise control – ICC/ACC) and a vehicle model (dealing with the engine, the transmission, road friction, aerodynamic, rolling, and slope resistance) [Are97].

- In similar spirit, Minderhoud has developed the *Simulation model of Motorways with Next generation vehicles* (Simone); this model focusses on intelligent driver support systems, such as ICC/ACC, platoon driving, centralised control of vehicles, et cetera. In contrast to most other (sub)microscopic models, Simone explicitly allows for rear-end collisions to occur under certain parameter combinations. As there is a close coupling between driver behaviour related parameters and those of the simulation, these collision dynamics enable the modeller to find realistic values (or ranges) for these parameters [Min99].

- Ludmann's *Program for the dEvelopment of Longitudinal micrOscopic traffic Processes in a Systemrelevant environment* (PELOPS), is akin to the previous two models. It is however more technologically oriented with respect to the car-following behaviour of vehicles, aiming at merging both a driver's perceptions and decisions, the car's handling, and the surrounding traffic conditions. At the core of the model, there are four modules that respectively describe vehicle routing in a road network, human decision taking (i.e., car-following, tactical decisions with respect to lane-changing, . . .), vehicle handling (i.e., a driver's physical acts of steering, accelerating and braking, . . .), and finally a module that describes physical vehicle characteristics (traction on elevations, engine capabilities, exhaust gas modelling, . . .) [Lud98; Ehm00].

To conclude this section, we like to mention an often scientifically-neglected area of research, namely the popular field of simulation in the *computer gaming industry*. Over the last couple of decades, numerous arcade-style *racing simulations* have been developed, allowing a *player* to be completely immersed in a three-dimensional virtual world in which racing at high speeds is paramount. Examples of these kinds of programmes are the highly addictive world of *Formula 1* racing, street racing in city environments, off-road rally races, . . . The underlying submicroscopic models in these games, have over the course of several years been evolved to incorporate all sorts of physical effects. Friction characteristics (e.g., pavement versus asphalt), road elevation, wet conditions, air drag and wind resistance (including effects such as slip streaming and downforce), car weight depending on fuel consumption, tyre wear, . . . have had influences on what we commonly refer to as *car handling*, i.e., realistic behaviour with respect to car acceleration, braking, and steering. Thanks to the increasing computational power of desktop computers, graphics cards, as well as dedicated gaming consoles (e.g., Microsoft's *Xbox*, Sony's *PlayStation*, Nintendo's *GameCube*, . . .), the path to a whole plethora of extra realistic effects has been paved: skidding, under- and oversteering, sun glare, overly realistic collision dynamics (in our opinion, this is where the arcade sensation plays a major role), . . .

3.2.5 The debate between microscopic and macroscopic models

Deciding which class of models, i.e., microscopic, or macroscopic (and we also include the mesoscopic models), is the correct one to formulate traffic flow problems, has been a debate among traffic engineers ever since the late fifties. Although the debate was not as intense as say, the one between first- and higher-order macroscopic traffic flow models (see Section 3.2.1.7 for more details), it nevertheless sparkled some interesting issues. As is nearly always the case, the true answer to the above question depends on the kind of problem one is interested in solving [Gaz02].

In the beginning years of traffic flow engineering, a bridge was formed between the microscopic General Motors car-following model of equation (3.28), and the Greenberg macroscopic model [Gre59; Gaz61]. This proved to be quite a significant breakthrough, as it was now possible to obtain all known steady-state macroscopic fundamental diagrams, by integrating the car-following equation with suitably chosen parameter values [Gaz02]. A recent example of this kind of linking, was done by Treiber and Helbing, who provided a *micro-macro link* between their non-local gas-kinetic mesoscopic model (see Section 3.2.2.3) and the intelligent driver model (see Section 3.2.3.1) [Hel98a; Hel02a].

Besides this explicit translation of microscopic into macroscopic (mesoscopic) models and vice versa, it is also possible to develop *hybrid models* that couple macroscopically modelled road links to microscopically modelled ones. Examples include the work of Magne et al., who develop a hybrid simulator that couples a METANET-like second-order macroscopic traffic flow model with the *SImulation TRAfic* (SITRA-B+) microscopic traffic flow model. Special attention is given to the interfaces between macroscopically and microscopically modelled road segments; each macroscopic time iteration in the simulator, is accompanied by a number of microscopic iterations [Mag00]. Lerner et al. also describe such a system, in which they employ a 'disaggregator' that combines macroscopic measurements with microscopic historical information in order to obtain correct vehicle time series [Ler00]. In similar spirit, the work of Bourrel and Henn links macroscopic representations of traffic flows to microscopic ones, using interfaces that describe the transitions between them. As an application of their methodology, they describe the translation between the first-order macroscopic LWR model and a vehicle representation of this model (based on trajectories) [Bou02]. Another avenue was pursued by the Wilco [Bur04b] and Wilco et al. [Bur05], who developed an integration framework between the MITSIMLab microscopic model and the Mezzo mesoscopic model. By building upon a mesoscopic approach, the strength of their work lies in the fact that no aggregation and disaggregation of flows needs to be performed.

3.2.6 Calibration and validation issues

As is the case for all simulation models, it is a necessity to perform calibration and validation when applying them to real-world case studies. In this section, we take a look at some of the issues related to these phases, considering them for both mesoscopic/macroscopic and microscopic models. We conclude with some general remarks that are related to a correct calibration methodology, holding for any type of model.

3.2.6.1 The case for mesoscopic and macroscopic models

As a general rule of thumb, these models are the easiest to calibrate. Due to their structure, they have a feasible amount of parameters that need to be tuned. In many cases, an explicit automatic optimisation of the parameter set is possible within a reasonable computation time. As the models typically exhibit a high-level behaviour on an aggregated scale, the main inputs for the calibration phase consist of flows and mean speeds over large sections (e.g., to estimate a $\overline{v}_{s_e}(k)$ fundamental diagram). These latter data can be derived from measurements stemming from sensors such as single and double inductive loop detectors, camera's, ...

Some examples of calibration and validation issues that go beyond the traditional OD-estimation procedures, are the early work of Cremer and Papageorgiou who used a parameter identification technique [Cre81], the work of Bellemans [Bel03] and Hegyi [Heg04] who used the second-order macroscopic METANET model (see also Section 3.2.1.7) in combination with a model predictive control (MPC) setup. Based on traffic sensor data, Balakrishna provided a framework for jointly calibrating the OD-matrices and route choice within the DynaMIT planning system [Bal02]. An interesting example of the calibration of the first-order macroscopic LWR model (see also Section 3.2.1.2) is the work of Logghe who combined both tempo-spatial plots and oblique cumulative curves (see also Section 2.3.2.2) [Log03a].

3.2.6.2 The case for microscopic models

Due to the sometimes large amount of parameters typically involved in microscopic traffic flow models, their computational complexity is often a significant disadvantage when compared to meso- or macroscopic models (although there are some exceptions, e.g., the traffic cellular automata models of Section 3.2.3.4). From the point of view of model calibration (matching the model's output to real-world observations) and validation (testing if the model's output matches with different data sets), this poses an interesting conundrum, as in many cases not all parameters are equally influential on the results (thus requiring some sensitivity analyses). In this sense, microscopic models contain a real danger of purporting to convey a sort of fake accuracy. Different parameter combinations can lead to the same phenomenological effects, leaving us pondering as to what exactly is causing the observed behaviour [Tam04a]. As there is no clear road map on how to calibrate microscopic traffic flow models, we here give a small sample of some of the numerous attempts that have been made.

There is the work of Jayakrishan and Sahraoui who distinguish between calibration in the conceptual (i.e., at the level of the underlying mathematical model) and operational (i.e., within the global context of the study) phases; they apply their operational methodology to both PARAMICS (micro) and DYNASMART (macro), using the *California Freeway Performance Measurement System* (PeMS) database from the PATH project to feed and couple both models [Jay00].

Based on a publicly available data set of a one-lay road corridor of six kilometres long (the data contained detailed cumulative curves), Brockfeld et al. systematically tested the predicted travel times of some ten well-known microscopic traffic flow models. As a result of a non-linear optimisation process to calibrate the models, they found that the intelligent driver model and the cell-transmission model perform the best (i.e., below an error rate of 17%), due to the fact that these models require the least amount of parameters (there were even some models such as the Gipps-based ones that had *hidden* parameters). Their final conclusion is noteworthy, as they state that *"creating a new model is often done, however calibrating this model to reality is a formidable task, which explains why there currently are more models than results about them"* [Bro03].

Related to the previous study, Hourdakis et al. present an automated systematic calibration methodology based on an optimisation process, applied to the AIMSUN2 simulator. The data used for the calibration procedure stem from a twenty kilometres long motorway in Minneapolis, Minnesota. The process first involves a calibration of the global model parameters (i.e., to get the macroscopic flows and speeds correct), after which the local parameters are dealt with (i.e., ramp metering setups, et cetera). In their results, Hourdakis et al. state an obtained average correlation coefficient of 0.961 for manual calibration (the results for the automated calibration are similar), which is quite high (they mostly explain this due to the data's high level of detail, as well as the quality of the simulator software) [Hou03].

Recently, Chu et al. extended the systematic, multi-stage calibration approach for the PARAMICS simulator. Based on data of a highly congested six kilometres long corridor network in the city of Irvine, Orange County, California, they first calibrate the driving behaviour models, then the route choice model, after which estimation and fine-tuning of the OD tables is done. Despite the good reproduction of travel times, their calibration methodology was done manually, and an automated optimisation procedure remains future work [Chu04a].

Other examples of calibration of microscopic traffic flow models, include the work of Rakha et al., who describe the data collection challenges for simulating a large-scale road network with the INTEGRATION simulator while providing a calibration framework [Rak98], the work of Dowling et al., who give an extensive account on the application of commercially available simulation tools to typically encountered traffic engineering problems [Dow02], the work of Mahanti which is primarily based on the correct representation of OD tables [Mah04], and the work of Panwai and Dia, who compared the car-following models in AIMSUN2, PARAMICS, and VISSIM using radar speed data from the Robert Bosch GmbH Research Group [Pan05].

And finally, we note the interesting work with respect to issues related to the general calibration of traffic flow models of Rakha et al. who provide an explicit framework in which they distinguish between calibration, validation, and model verification [Rak96], Hellinga who poses some key questions involved in the calibration process [Hel98c], and Burghout who gives an estimation on the number of runs for stochastic traffic simulation models [Bur04a].

3.2.6.3 Some general remarks

To end this section, we state some important principles that are — in our opinion — related to a correct calibration methodology. First and foremost, we believe that all traffic flow models (whether they are macro-, meso- or microscopic in nature), should be able to accurately reproduce and predict the encountered delays, queue lengths, and other macroscopic *first-order characteristics* (i.e., the kinematic wave speed, a correct and realistic road capacity, ...). One way to test this is the use of cumulative curves, as they provide an elegant way to automatically perform a good calibration. It is for example possible to consider the difference between observed and simulated curves, and then use a Kolmogorov-Smirnov goodness-of-fit statistical test to decide on whether the difference is statistically significant, or if it is just a Brownian motion with a zero mean. Only when these first-order effects can be correctly reproduced, the next step can be to consider *second-order effects* such as waves of stop-and-go traffic, oscillations, ...

Furthermore, it is important to take into account the *spatial nature* of the study area, i.e., a detailed description of the road infrastructure, with bottleneck locations as well as up- and downstream boundary conditions. With respect to the model that is created within the computer, it is paramount to know how the model behaves on both the link as well as the node level. Because the models are most of the time working with fairly homogeneous road links (e.g., constant elevations, no road curvature, ...), it might be necessary to allow for small deviations from (or fixes to) reality (e.g., inserting extra intermediate nodes in the network in order to artificially obtain bottlenecks).

3.3 Conclusions

The material elaborated upon in this chapter, spanned a broad range going from transportation planning models that operate on a high level, to traffic flow models that explicitly describe the physical propagation of traffic flows.

As explained in the introduction, we feel there is a frequent confusion among traffic engineers and policy makers when it comes to transportation planning models and the role that traffic flow models play therein. To this day, many transportation planning bureaus continue to use static tools for evaluating policy decisions, whereas the need for dynamic models is getting more and more pronounced [Mae04b]. Still more troublesome is the fact that in the present day, incorrect studies (e.g., wrong assumptions, an inadequate modelling approach, ...) may lead to unsound policy decisions. Indeed, as Brinkman states, "*Thirty years ago scholars first presented convincing evidence that local officials use biased travel demand forecasts to justify decisions based on unstated considerations.*" [Bri03]

Even after more than sixty years of traffic flow modelling, the debate on what is the correct modelling approach remains highly active. On the transportation planning side, many agencies still primarily focus on the traditional four step model (4SM), because it is the best intuitively understood approach. In contrast to this, activity-based modelling (ABM) is gaining momentum, although it remains a rather obscure discipline to many people. At the basis of this scrutiny towards the ABM, lies the absence of a generally accepted framework such as the one of the 4SM. It is tempting to translate the ABM approach to the 4SM, by which, e.g., the ABM's synthetic population generation (including activity generation, household choices and scheduling) corresponds to the 4SM's production and attraction, distribution, and modal split (or to discrete choice theory in a broader setting), thereby generating (time dependent) OD tables. Similarly, the ABM's agent simulation can be seen as an implementation of the 4SM's traffic assignment. However, it remains difficult to gain insight into this kind of direct translation and the resulting travel behaviour, although the ABM's scientific field is continuously in a state of flux thanks to the increasing computational power.

On the traffic flow modelling side, the debate on whether or not to use macro-/meso- or microscopic models still continues to spawn many intriguing discussions. Despite the respective criticisms, it is widely agreed upon that modelling driver behaviour entails complex human-human, human-vehicle, and vehicle-vehicle interactions. These call for interdisciplinary research, drawing from fields such as mathematics, physics, and engineering, as well as sociology and psychology (see, e.g., the overview of Helbing and Nagel [Hel04]).

In the next chapter, we go into more detail on the class of traffic cellular automata, which are, as explained in Section 3.2.3.4, a special case of computationally efficient microscopic traffic flow models.

Part II

Cellular Automata Models of Road Traffic

Chapter 4

Traffic cellular automata

As hinted at in the previous chapter, we now focus our attention towards computationally efficient microscopic traffic flow models. Traffic cellular automata (TCA) models fit this description nicely. True to the spirit of statistical mechanics, all the TCA models discussed in this dissertation do not have a realistic microscopic description of traffic flows as their primary intent[1]*, but are rather aimed at obtaining a correct macroscopic behaviour through their crude microscopic description. As such, they are able to positively capture the first- and second-order macroscopic effects of traffic streams.*

In this chapter, we provide a detailed description of the methodology of cellular automata applied to traffic flows[2]*. We first discuss their background and physical setup, followed by an account of the mathematical notations we adopt. The remaining majority of this chapter extensively discusses the behavioural aspects of several state-of-the-art TCA models encountered in literature (our overview distinguishes between single-cell and multi-cell models). The chapter concludes with a concise overview of TCA models in a multi-lane setting, and TCA models used to describe two-dimensional traffic (e.g., a grid for city traffic). We end with a description of several common analytical approximations to single-cell TCA models.*

> *Note that aside from our phenomenological discussion of different TCA models, we refer the reader to the work of Chowdhury et al. [Cho00], Santen [San99], and Knospe et al. [Kno04] for more theoretically- and quantitatively-oriented overviews. Mahnke et al. also provide a well-defined background for probabilistic traffic flow models [Mah05].*

[1] Such an approach would involve more human-oriented aspects such as those found in socio-economic, behavioural, and psychological sciences. Due to large lack of knowledge about the manner in which human beings operate in a traffic system, traffic engineers currently stick with this higher-level scientific approach.

[2] Note that this chapter was also published as a stand-alone review in *Physics Reports* [Mae05].

4.1 Background and physical setup for road traffic

In this section, we give a brief overview of the historic origins of cellular automata, as they were conceived around 1950. We subsequently describe which main ingredients constitute a cellular automaton: the physical environment, the cells' states, their neighbourhoods, and finally a local transition rule. We then move on to a general description on how cellular automata are applied to vehicular road traffic, discussing their physical environment and the accompanying rule set that describes the vehicles' physical propagation.

4.1.1 Historic origins of cellular automata

The mathematical concepts of cellular automata (CA) models can be traced back as far as 1948, when Johann Louis von Neumann introduced them to study (living) biological systems [Neu48]. Central to von Neumann's work, was the notion of *self-reproduction* and theoretical machines (called *kinematons*) that could accomplish this. As his work progressed, von Neumann started to cooperate with Stanislaw Marcin Ulam, who introduced him to the concept of *cellular spaces*. These described the physical structure of a cellular automaton, i.e., a grid of cells which can be either 'on' or 'off' [Wol83; Del98]. Interestingly, Alan Mathison Turing proposed in 1952 a model that illustrated reaction-diffusion in the context of *morphogenesis* (e.g., to explain the patterns of spots on giraffes, of stripes on zebras, ...). His model can be seen as a type of continuous CA, in which the cells have a direct analogy with a simplified biological organism [Tur52].

In the seventies, CA models found their way to one of the most popular applications called 'simulation games', of which John Horton Conway's *"Game of Life"* [Gar70] is probably the most famous. The game found its widespread fame due to Martin Gardner who, at that time, devoted a Scientific American column, called *"Mathematical Games"*, to it. Life, as it is called for short, is traditionally 'played' on an infinitely large grid of cells. Each cell can either be 'alive' or 'dead'. The game evolves by considering a cell's all surrounding neighbours, deciding whether or not the cell should live or die, leading to phenomena called *'birth'* (a dead cell becomes alive when there are exactly three neighbouring cells alive), *'survival'* (a live cell with two or three live neighbours stays alive), and *'overcrowding'* or *'loneliness'* (in all other cases a cell dies or remains dead). An example of a Life game board can be seen in Figure 4.1. Typical of Life, is the spawning of a whole plethora of patterns or shapes, having illustrious names such as gliders, guns, space ships, puffers, beehives, oscillators, ... The Game of Life is now all about how these shapes evolve, and whether or not they die out or live indefinitely (either by remaining stationary or moving around).

The widespread popularisation of CA models was achieved in the eighties through the work of Stephen Wolfram. Based on empirical experiments using computers, he gave an extensive classification of CA models as mathematical models for self-organising statistical systems [Wol83; Wol02]. Wolfram's work culminated in his mammoth

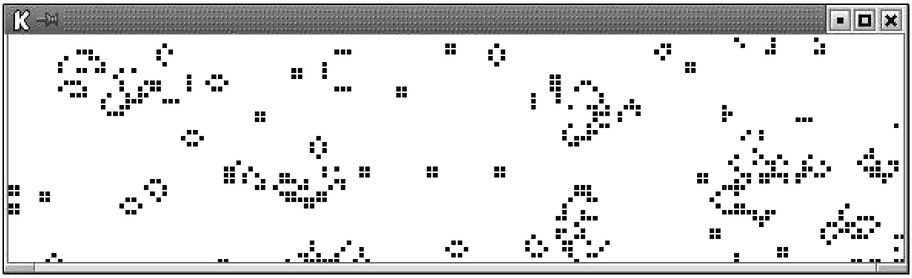


Figure 4.1: An example of the Game of Life, with a rectangular grid of cells. Live cells are coloured black, whereas dead cells remain white. The image shows a snapshot during the game's course, illustrating many different shapes to either die out, or live indefinitely by remaining stationary or moving around (image adapted from [Geo02]).

monograph, called *A New Kind of Science* [Wol02]. In this book, Wolfram related cellular automata to all disciplines of science (e.g., sociology, biology, physics, mathematics, ...). Despite the broad range of science areas touched upon, Wolfram's book has received its share of criticism. As an example of this, we mention the comments of Gray, who points out that Wolfram's results suffer from a rigourous mathematical test. As a consequence, the physical examples in his book are deemed either uncheckable or unconvincing. Gray's final critique is that *"... he [Wolfram] has helped to popularise a relatively little-known mathematical area (CA theory), and he has unwittingly provided several highly instructive examples of the pitfalls of trying to dispense with mathematical rigour"* [Gra03]. However, with respect to their computational power, CA models can emulate universal Turing machines within the theories of computation and complexity. Recently, Chua took Wolfram's empirical observations one step further, proving that some of the CA models are capable of Turing universal computations. He furthermore introduced the paradigm of *cellular neural networks* (CNN), which provide a very efficient method for performing massive parallel computations, and are a generalisation of cellular automata [Chu04b].

Finally, an important step in this direction, is Bill Gosper's proof that the Game of Life is computationally universal, i.e., it can mimic arbitrary algorithms [Gos74]. Notably, one of the most profound testimonies related to this concept, is the work of Konrad Zuse and Edward Fredkin at the end of the sixties. Their Zuse-Fredkin thesis states that *"The Universe is a cellular automaton"*, and is based on the assumption that the Universe's physical laws are discrete in nature [Zus67; Zus69; Fre90]. This latter statement was also conveyed by Wolfram in his famous CA compendium [Wol02].

4.1.2 Ingredients of a cellular automaton

From a theoretical point of view, four main ingredients play an important role in cellular automata models [Gut96; Del98; Sar00]:

(1) The physical environment
This defines the *universe* on which the CA is computed. This underlying structure consists of a *discrete lattice of cells* with a rectangular, hexagonal, or other topology (see Figure 4.2 for some examples). Typically, these cells are all equal in size; the lattice itself can be finite or infinite in size, and its dimensionality can be 1 (a linear string of cells called an *elementary cellular automaton* or ECA), 2 (a grid), or even higher dimensional. In most cases, a common — but often neglected — assumption, is that the CA's lattice is embedded in a *Euclidean space*.

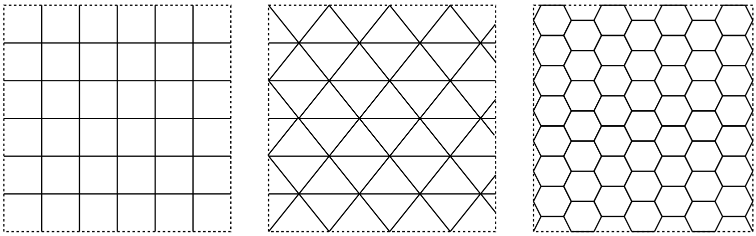


Figure 4.2: Some examples of different Euclidean lattice topologies for a cellular automaton in two dimensions. *Left:* rectangular. *Middle:* triangular/isometric. *Right:* hexagonal.

(2) The cells' states
Each cell can be in a certain state, where typically an integer represents the number of distinct states a cell can be in, e.g., a binary state. Note that a cell's state is not restricted to such an integer domain (e.g., $\mathbb{Z}_2$), as a continuous range of values is also possible (e.g., $\mathbb{R}^+$), in which case we are dealing with *coupled map lattices* (CML) [Cru87; Kan90]. We call the states of all cells collectively a CA's *global configuration*[3].

(3) The cells' neighbourhoods
For each cell, we define a neighbourhood that locally determines the evolution of the cell. The size of neighbourhood is the same for each cell in the lattice. In the simplest case, i.e., a 1D lattice, the neighbourhood consists of the cell itself plus its adjacent cells. In a 2D rectangular lattice, there are several possibilities, e.g., with a radius of 1 there are, besides the cell itself, the four north, east, south, and west adjacent cells (*von Neumann neighbourhood)*, or the previous five cells as well as the four north-east, south-east, south-west, and north-west diagonal cells (*Moore neighbourhood)*; see Figure 4.3 for an example of both types of neighbourhoods. Note that as the dimensionality of the lattice increases, the

[3]This convention asserts that states are local and refer to cells, while a configuration is global and refers to the whole lattice.

number of direct neighbours of a cell increases exponentially.

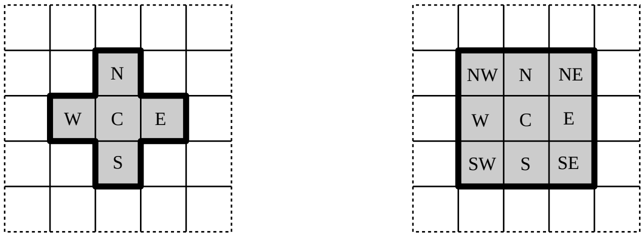


Figure 4.3: Two commonly used two-dimensional CA neighbourhoods with a radius of 1: the von Neumann neighbourhood (left) consisting of the central cell itself plus 4 adjacent cells, and the Moore neighbourhood (right) where there are 8 adjacent cells. Note that for one-dimensional CAs, both types of neighbourhoods are the same.

(4) A local transition rule

This rule (also denoted as a function δ) acts upon a cell and its direct neighbourhood, such that the cell's state changes from one *discrete time step* t to another $t + 1$ (i.e., the system's iterations[4]), as is depicted in Figure 4.4. The CA evolves in time and space as the rule is subsequently applied to all the cells *in parallel*. Typically, the same rule is used for all the cells (if the converse is true, then the term *hybrid* CA is used). When there are no stochastic components present in this rule, we call the model a *deterministic* CA, as opposed to a *stochastic* (also called *probabilistic*) CA.

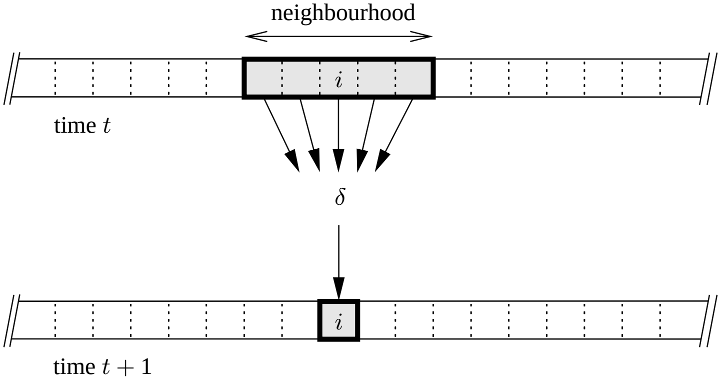


Figure 4.4: An illustration of the local transition rule δ, acting upon a cell i and its direct neighbourhood in the lattice of a one-dimensional cellular automaton. The cell's state changes from one discrete time step t to another $t + 1$.

[4]Note that in case the new state of a cell is based on more than one previous time step, the resulting cellular automaton is said to be of *higher order*, as opposed to most classical *first-order* CAs.

> As the local transition rule is applied to all the cells in the CA's lattice, the global configuration of the CA changes. This is also called the CA's *global map*, which transforms one global configuration into another. This corresponds to the notion of *computing a function* in automata theory, see also Section 4.2.1. Sometimes, the CA's evolution can be reversed by computing past states out of future states. By evolving the CA backwards in time in this manner, the CA's *inverse global map* is computed. If this is possible, the CA is called *reversible*, but if there are states for which no precursive state exists, these states are called *Garden of Eden* (GoE) states and the CA is said to be *irreversible*.
>
> Finally, when the local transition rule is applied to all cells, its global map is computed. In the context of the theory of dynamical systems, this phenomenon of *local simple interactions* that lead to a *global complex behaviour* (i.e., the spontaneous development of order in a system due to *internal* interactions), is termed *self-organisation* or *emergence*.

Whereas the previous paragraphs discussed the classical approach to CA models, the following sections will exclusively focus on vehicular traffic flows, leading to traffic cellular automata (TCA) models: Section 4.1.3 discusses the physical environment on which these TCA models are based, and Section 4.1.4 deals with their accompanying rule set that determines the vehicular motion.

4.1.3 Road layout and the physical environment

When applying the cellular automaton analogy to vehicular road traffic flows, the physical environment of the system represents the road on which the vehicles are driving. In a classical single-lane setup for traffic cellular automata, this layout consists of a one-dimensional lattice that is composed of individual cells (our description here thus focuses on unidirectional, single-lane traffic). Each cell can either be empty, or is occupied by *exactly* one vehicle; we use the term *single-cell models* to describe these systems. Another possibility is to allow a vehicle to span several consecutive cells, resulting in what we call *multi-cell models*. Because vehicles move from one cell to another, TCA models are also called *particle-hopping models* [Nag96].

An example of the tempo-spatial dynamics of such a system is depicted in Figure 4.5, where two consecutive vehicles i and j are driving on a 1D lattice. A typical discretisation scheme assumes $\Delta T = 1$ s and $\Delta X = 7.5$ m, corresponding to speed increments of $\Delta V = \Delta X / \Delta T = 27$ km/h. The spatial discretisation corresponds to the average length a conventional vehicle occupies in a closely packed jam (and as such, its width is neglected), whereas the temporal discretisation is based on a typical driver's reaction time and we implicitly assume that a driver does not react to events between two consecutive time steps [Nag92b].

With respect to the layout of the system, we can distinguish two main cases: closed versus open systems. They correspond to periodic (or cyclic) versus open boundary

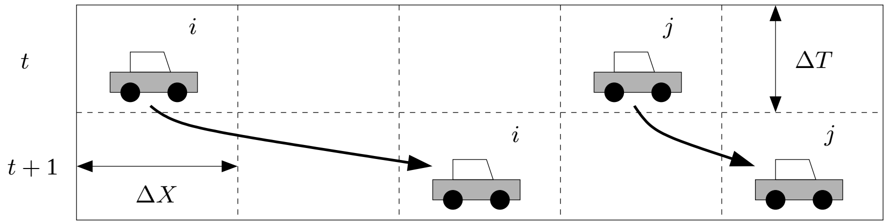


Figure 4.5: Schematic diagram of the operation of a single-lane traffic cellular automaton (TCA); here, the time axis is oriented downwards, the space axis extends to the right. The TCA's configuration is shown for two consecutive time steps t and $t+1$, during which two vehicles i and j propagate through the lattice.

conditions. The former is usually implemented as a closed ring of cells, sometimes called the *Indianapolis scenario*, while the latter considers an open road. This last type of system, is also called the *bottleneck scenario*. The name is derived from the fact that this situation can be seen as the outflow from a jam, where vehicles are placed at the left boundary whenever there is a vacant spot. Note that, in closed systems, the number of vehicles is always conserved, leading to the description of *number conserving cellular automata* (NCCA) [Mor03].

4.1.4 Vehicle movements and the rule set

The propagation of the individual vehicles in a traffic stream, is described by means of a rule set that reflects the car-following and lane-changing behaviour of a traffic cellular automaton evolving in time and space. The TCA's local transition rule actually comprises this set of rules. They are consecutively applied to all vehicles in parallel (called a *parallel update*). So in a classic setup, the system's state is changed through *synchronous position updates* of all the vehicles: for each vehicle, the new speed is computed, after which its position is updated according to this speed and a possible lane-change manoeuvre. Note that there are other ways to perform this update procedure, e.g., a random sequential update (see Section 4.3.2.4). Because time is discretised in units of ΔT seconds, an *implicit reaction time* is assumed in TCA models. It is furthermore assumed that a driver does not react to events between consecutive time steps.

Similar to our discussion of anisotropy at the end of Section 3.2.1.5, we assume that for single-lane traffic, vehicles act as *anisotropic particles*: they respond only to frontal stimuli. So typically, the car-following part of a rule set only considers the direct frontal neighbourhood of the vehicle to which the rules are applied. The radius of this neighbourhood should be taken large enough such that vehicles are able to drive collision-free. Typically, this radius is equal to the maximum speed a vehicle can achieve, expressed in cells per time step.

From a microscopic point of view, the process of a vehicle following its predecessor is typically expressed using a *stimulus-response relation* (see Section 3.2.3.1 for more details). Typically, this response is the speed or the acceleration of a vehicle; in TCA models, a vehicle's stimulus is mainly composed of its speed and the distance to its leader, with the response directly being a new (adjusted) speed of the vehicle. In a strict sense, this only leads to the avoidance of accidents. Some models however, incorporate more detailed stimuli, such as anticipation terms[5]. When these effects are taken into account together with a safety distance, strong accelerations and abrupt braking can be avoided. Hence, as the speed variance is decreased, this results in a more stable traffic stream [Kno02b; Eis03; Lár04].

To conclude this section, we note that a TCA model can also be derived from a Gipps car-following model. As explained in Section 3.2.3.1, all speeds in the Gipps model are directly computed from one discrete time step to another. If now the spatial dimension is also discretised (a procedure called *coarse graining*), then this will result in a TCA model.

4.2 Mathematical notation

In this section, we give an overview of the mathematical notation adopted throughout this dissertation. The focus will be on the variables in TCA models, the measurements that can be done on a TCA model's lattice, and their conversion to real-world units. We first take a look at the notation that is commonly used in automata theory, from which cellular automata sprung.

4.2.1 Classical notation based on automata theory

Let us first briefly present the notation for cellular automata models, adopted in spirit of *automata theory*. As mentioned in Section 4.1, a CA model represents a discrete dynamic system, consisting of four ingredients:

$$\text{CA} = (\mathcal{L}, \Sigma, \mathcal{N}, \delta), \tag{4.1}$$

where the physical environment is represented by the discrete lattice $\mathcal{L}$ and the set of possible states denoted by Σ. Each i^{th} cell of the lattice, has at time step t a state $\sigma_i(t) \in \Sigma$. Furthermore, the associated neighbourhood with this cell is represented by $\mathcal{N}_i(t)$, i.e., a (partially) ordered set of cells. Finally, the local transition rule is represented as:

$$\delta : \Sigma^{|\mathcal{N}|} \longrightarrow \Sigma : \bigcup_{j \in \mathcal{N}_i(t)} \sigma_j(t) \longmapsto \sigma_i(t+1). \tag{4.2}$$

[5] These forms of 'anticipation' only take leaders' reactions into account, *without predicting* them.

Equation (4.2) shows that the state of the i^{th} cell at the next time step $t+1$ is computed by δ based on the states of all the cells in its neighbourhood at the current time step t. In the previous equation, $|\mathcal{N}|$ represents the number of cells in this neighbourhood, which is taken to be invariant with respect to time and space. Note that the local transition rule is commonly given by a *rule table*, where the output state is listed for each possible input configuration of the neighbourhood. Given the sizes of Σ and $\mathcal{N}$, the total number of possible rules equals:

$$|\Sigma^{\Sigma^{\mathcal{N}}}|, \tag{4.3}$$

where each of the $|\Sigma^{\mathcal{N}}|$ possible configurations of a cell's neighbourhood is mapped to the number of possible states a cell can be in.

Considering the ordered set of all the states of all cells collectively at time step t, a CA's global configuration is obtained as:

$$\mathcal{C}(t) = \bigcup_{j \in \mathcal{L}} \sigma_j(t), \tag{4.4}$$

with $\mathcal{C}(t) \in \Sigma^{\mathcal{L}}$ where the latter refers to the set of all possible global configurations a CA can be in (also called its *phase space*). Sometimes, such a global configuration $\mathcal{C}(t)$ is also represented by its characteristic polynomial (i.e., generating function) [Wol84b]:

$$\mathcal{C}(t) = \sum_{j=0}^{|\mathcal{L}|} \sigma_j(t)\, x^j. \tag{4.5}$$

If we now apply the local transition rule to all the cells in the CA's lattice, the next configuration of the CA can be computed by its induced global map:

$$G : \Sigma^{\mathcal{L}} \longrightarrow \Sigma^{\mathcal{L}} : \mathcal{C}(t) \longmapsto \mathcal{C}(t+1). \tag{4.6}$$

Note that if the CA is reversible, the inverse global map G^{-1} can be computed. As the CA evolves in time and space, the global map is iterated from a certain initial configuration $\mathcal{C}(0)$ at $t = 0$, leading to the following sequence of configurations:

$$\mathcal{C}(0) \rightarrow G(\mathcal{C}(0)) \rightarrow G^2(\mathcal{C}(0)) \rightarrow G^3(\mathcal{C}(0)) \rightarrow \cdots \tag{4.7}$$

The above sequence is called the *trajectory* of the initial configuration $\mathcal{C}(0)$ under the global map G, and we denote it by:

$$\mathcal{T}_{\mathcal{C}(0)|G} = \{G^n(\mathcal{C}(0)) \mid n \in \mathbb{N}\}. \tag{4.8}$$

When this trajectory is periodic or chaotic, we use the terminology *forward orbit* and denote it by $\mathcal{O}^{+}_{\mathcal{C}(0)|G}$. Similarly, the *backward orbit* (i.e., the reverse trajectory) is denoted by $\mathcal{O}^{-}_{\mathcal{C}(t)|G^{-1}}$, where we specify a certain global configuration $\mathcal{C}(t)$ at time step t under the inverse global map G^{-1}.

4.2.1.1 Classification of CA rules

Computing the global map G is rather difficult, as it may require many or even an infinite amount of iterations in order to obtain the trajectories. In practice, the system's lattice size should be taken infinitely large, but even only considering 1000 cells of a binary elementary cellular automaton (ECA) would increase the size of the search space of global configurations to $2^{1000} \approx 10^{300}$.

A more intuitive methodology, is to observe a CA's tempo-spatial behaviour, i.e., its evolution on the lattice in the course of time. To this end, Stephen Wolfram empirically studied many configurations of binary ECA rules, with a neighbourhood of three cells. According to equation (4.3), this amounts to $2^{2^3} = 256$ different rules. In 1984, based on this research, Wolfram conjectured four distinct *universality classes* [Wol84a]:

Class I
: These CA evolve after a finite number of iterations to a unique homogeneous state, i.e., a *limit point*.

Class II
: These CA generate regular, periodic patterns, i.e., entering a *limit cycle*.

Class III
: CAs in this class evolve to aperiodic patterns, independent of the initial configuration; their trajectories in the configuration space lie on a *chaotic attractor*.

Class IV
: This class encompasses all the CAs that seem to behave in a *complex* way, with features such as propagating structures, long transients; they are thought to have the capability of universal computation.

Although Wolfram's classification scheme is widely adopted, it still remains a tentative result as he himself states [Wol02]. Note that the type of classification he provides is *phenotypical*, in the sense that it is based on observed behaviour, whereas a *genotypical* classification would be based on the intrinsic structure of the rules in each class.

Despite these observations, classification still remains a difficult task as is evidenced by the ongoing research in dynamical systems. Other attempts at classification of ECA rules include the following. Firstly, Čulik and Yu gave a formalisation of Wolfram's classes [Čul88]. Secondly, Li and Packard studied the structure of the ECA rule space according to a certain distance metric, resulting in five classes [Li90]. Then, Braga et al. identified three classes based on the growth of patterns observed in CA models [Bra95]. Next, Wuensche used a whole arsenal of local measures to automatically create complex rules, thereby classifying the rule space for the CAs' dynamics [Wue99]. Furthermore, Dubacq et al. classified CA models based on their algorithmic complexity by measuring the information content of the local transition rule [Dub01]. And finally, Fatès who used a macroscopic parameter, i.e., the density of 1's, to separate chaotic ECA rules from non-chaotic ones [Fat03].

4.2.1.2 An example of a CA

To end this section, let us give some definitions of a one-dimensional, infinitely large, binary state CA with a neighbourhood of radius 1:

$$\begin{aligned}
\mathcal{L} &= \mathbb{Z}^d \quad \text{(with } d = 1\text{)}, && (4.9)\\
\Sigma &= \mathbb{Z}_2 = \{0, 1\}, && (4.10)\\
\mathcal{N}_i &= \{i-1, i, i+1\}, && (4.11)\\
\delta(i,t) &: \mathbb{Z}_2^3 \longrightarrow \mathbb{Z}_2 && \\
&: \{\sigma_{i-1}(t), \sigma_i(t), \sigma_{i+1}(t)\} \longmapsto \sigma_i(t+1), && (4.12)\\
G(\mathcal{C}(t)) &: \mathbb{Z}_2^{\mathbb{Z}} \longrightarrow \mathbb{Z}_2^{\mathbb{Z}} && \\
&: \mathcal{C}(t) \longmapsto \mathcal{C}(t+1). && (4.13)
\end{aligned}$$

Note that in equation (4.11), we assume that the i^{th} cell's neighbourhood is represented by integer indices (i.e., the cells form a totally ordered set). This alleviates the need for an explicit representation of the cells themselves, as it is now sufficient to work with the cells' indices and states. The transition rule δ in equation (4.12) takes as its arguments a cell's index i and current time step t, but operates on the states of this cell's neighbourhood. The global map in equation (4.13) operates on the global configuration of the CA at time step t.

4.2.2 Basic variables and conventions

Conforming to the setup and notation discussed in the previous sections, we denote a TCA's discrete lattice by $\mathcal{L}$ (for the remainder of this section, we assume a *rectangular lattice*). This lattice physically represents the road on which vehicles will drive in a TCA model. It consists of $L_{\mathcal{L}}$ lanes, each of which has $K_{\mathcal{L}}$ cells, so in total there are $L_{\mathcal{L}} \times K_{\mathcal{L}}$ cells in the lattice ($L_{\mathcal{L}}, K_{\mathcal{L}} \in \mathbb{N}_0$). Each cell can either be empty, or occupied with a single vehicle that spans one or more consecutive cells. An example of a lattice containing several vehicles, can be seen in Figure 4.6.

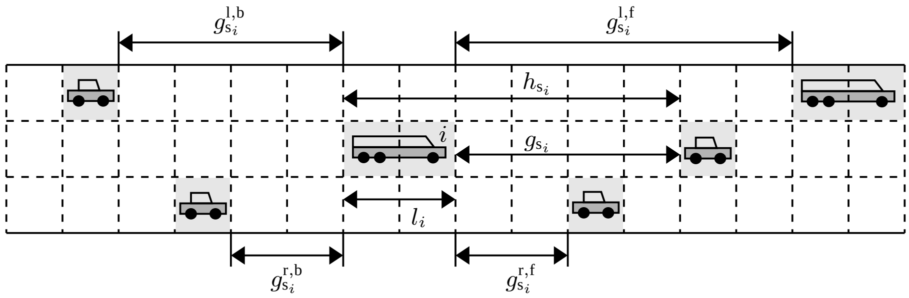


Figure 4.6: A portion of the lattice $\mathcal{L}$ at a certain time step; it has $L_{\mathcal{L}} = 3$ lanes, containing six vehicles. The central vehicle i has a space headway $h_{s_i} = 6$ cells, consisting of a space gap $g_{s_i} = 4$ cells and its length $l_i = 2$ cells. There are four other space gaps to be considered when the neighbouring lanes are taken into account: $g_{s_i}^{l,f}$ (left-front), $g_{s_i}^{l,b}$ (left-back), $g_{s_i}^{r,f}$ (right-front), and $g_{s_i}^{r,b}$ (right-back), equalling 6, 4, 2 and 2 cells, respectively.

Based on the microscopic vehicle characteristics of Section 2.2.2, we propose to use the following set of definitions[6] for *multi-lane* vehicular road traffic flows that are *heterogeneous* (in the sense of having different vehicle lengths):

$$g_{s_i}^{l,f} = x_i^{l,f} - x_i - l_i, \tag{4.14}$$
$$g_{s_i}^{r,f} = x_i^{r,f} - x_i - l_i, \tag{4.15}$$
$$g_{s_i}^{l,b} = x_i - x_i^{l,b} - l_i^{l,b}, \tag{4.16}$$
$$g_{s_i}^{r,b} = x_i - x_i^{r,b} - l_i^{r,b}, \tag{4.17}$$

for which we assume that a vehicle's position is denoted by the cell that contains its rear bumper. For the example in Figure 4.6, the left and right frontal and backward space gaps of the central vehicle i are 6, 4, 2 and 2 cells, respectively (all these space gaps thus represent effective distances, corresponding to the number of empty cells between vehicles). Similar definitions hold for the space headways $h_{s_i}^{l,f}$, $h_{s_i}^{r,f}$, $h_{s_i}^{l,b}$, and $h_{s_i}^{r,b}$, i.e., the vehicle lengths in the right-hand sides of equations (4.14) – (4.17) are dropped. Derivations for the time gaps $g_{t_i}^{l,f}$, $g_{t_i}^{r,f}$, $g_{t_i}^{l,b}$, and $g_{t_i}^{r,b}$, and time headways $h_{t_i}^{l,f}$, $h_{t_i}^{r,f}$, $h_{t_i}^{l,b}$, and $h_{t_i}^{r,b}$ are analogous.

Discriminating between frontal and backward neighbours in the adjacent lanes to the i^{th} vehicle, is done based on their positions, i.e.:

$$\{x_i^{l,b}, x_i^{r,b}\} < x_i \leq \{x_i^{l,f}, x_i^{r,f}\}. \tag{4.18}$$

[6]Note that in the remainder of this dissertation, we have dropped the functional dependencies in favour of visual clarity.

According to equation (4.18), a vehicle that is driving alongside in an adjacent lane to the i^{th} vehicle, will be considered as a backward neighbour as long as its rear bumper is located strictly behind the rear bumper of the i^{th} vehicle (even if this neighbour has a large length that 'sticks out' in front of the i^{th} vehicle).

Under the above set of assumptions, we can now write the conditions for a successful lane change (i.e., a possible gap acceptance) as the following constraints:

$$g_{\mathrm{s}_i}^{\mathrm{l,f}} \geq 0 \quad \wedge \quad g_{\mathrm{s}_i}^{\mathrm{l,b}} \geq 0 \qquad \text{(left lane change)}, \tag{4.19}$$

$$g_{\mathrm{s}_i}^{\mathrm{r,f}} \geq 0 \quad \wedge \quad g_{\mathrm{s}_i}^{\mathrm{r,b}} \geq 0 \qquad \text{(right lane change)}. \tag{4.20}$$

With respect to the domains of all variables, we note that all vehicle lengths, space gaps, and space headways are expressed as integers, or more specifically:

$$\begin{aligned} l_i, h_{\mathrm{s}_i}, h_{\mathrm{s}_i}^{\mathrm{l,b}}, h_{\mathrm{s}_i}^{\mathrm{r,b}} &\in \mathbb{N}_0, \\ g_{\mathrm{s}_i}, h_{\mathrm{s}_i}^{\mathrm{l,f}}, h_{\mathrm{s}_i}^{\mathrm{r,f}} &\in \mathbb{N}, \\ g_{\mathrm{s}_i}^{\mathrm{l,b}}, g_{\mathrm{s}_i}^{\mathrm{r,b}}, g_{\mathrm{s}_i}^{\mathrm{l,f}}, g_{\mathrm{s}_i}^{\mathrm{r,f}} &\in \mathbb{Z}. \end{aligned}$$

In contrast to this, the occupancy times, time headways, and time gaps are not restricted to the domain of integers, i.e.:

$$\begin{aligned} \rho_i, h_{\mathrm{t}_i}, h_{\mathrm{t}_i}^{\mathrm{l,b}}, h_{\mathrm{t}_i}^{\mathrm{r,b}} &\in \mathbb{R}_0^+, \\ g_{\mathrm{t}_i}, h_{\mathrm{t}_i}^{\mathrm{l,f}}, h_{\mathrm{t}_i}^{\mathrm{r,f}} &\in \mathbb{R}^+, \\ g_{\mathrm{t}_i}^{\mathrm{l,b}}, g_{\mathrm{t}_i}^{\mathrm{r,b}}, g_{\mathrm{t}_i}^{\mathrm{l,f}}, g_{\mathrm{t}_i}^{\mathrm{r,f}} &\in \mathbb{R}. \end{aligned}$$

For example, the occupancy time as defined by equation (2.3) in Section 2.2.2, corresponds to the time a vehicle 'spends' in its own cells.

To conclude, each vehicle i in the lattice has an associated speed $v_i \in \mathbb{N}$ (expressed in cells per time step ΔT), which is bounded by a maximum speed $v_{\text{max}} \in \mathbb{N}_0$. For example, if we set $\Delta T = 1.2$ s, $\Delta X = 7.5$ m, and $v_{\text{max}} = 5$ cells/time step, then $v_i \in \{0, \ldots, 5\}$ which corresponds to a maximum of $5 \times \Delta X / \Delta T = 5 \times 7.5$ m/s $\div$ 1.2 s $= 31.25$ m/s $= 112.5$ km/h. As can be seen in this derivation, we only consider positive speeds in our models, i.e., vehicles always move forward.

4.2.3 Performing macroscopic measurements

The previously discussed quantities are all microscopic traffic stream characteristics, corresponding to those of Section 2.2. In this section, we reconsider the macroscopic quantities of Section 2.3, i.e., densities, flows, and mean speeds. As we now have to measure these quantities on a TCA's lattice $\mathcal{L}$, we present three possibilities for obtaining the data points:

- by performing local measurements with an artificial loop detector of finite length (open and closed systems),
- by performing global measurements on the entire lattice (closed system),
- and by performing local measurements with an artificial loop detector of unit length (open and closed systems).

In the following three sections, we give detailed derivations of each of these measurement techniques. Locally measured quantities are indicated by a 'l' subscript, whereas globally measured ones are indicated by an 'g' subscript. A temporal and spatial discretisation of respectively ΔT (in seconds) and ΔX (in metres) is implicitly assumed.

> For all following techniques, we assume an integer measurement period of T_{mp} time steps. Thus, aggregating data into intervals of 60 seconds with $\Delta T = 1.2$ s, requires a measurement period of:
>
> $$T_{\text{mp}} = \left\lceil \frac{60}{1.2} \right\rceil = 50 \text{ time steps.} \tag{4.21}$$
>
> Furthermore, densities are expressed in vehicles per cell, flows in vehicles per time step, and space-mean speeds in cells per time step.

4.2.3.1 Local measurements with a detector of finite length

In this section, we deal with an artificial loop detector of finite length $K_{\text{ld}} \in \mathbb{N}_0$, located in a single lane. Note that typically, $K_{\text{ld}} \geq v_{\text{max}}$, so as to ensure that no vehicles can 'skip' the detector between consecutive time steps. The first step in our approach for performing these measurements, is based on obtaining local measurements of the density and flow for a measurement region R_{s} (see Figure 2.3 of Section 2.3) at a certain time step t, using equations (2.4) and (2.15), respectively. Once these are known, the space-mean speed can be derived using the fundamental relation (2.33):

$$k_{\text{l}}(t) = \frac{N(t)}{K_{\text{ld}}}, \tag{4.22}$$

$$q_{\text{l}}(t) = \frac{1}{K_{\text{ld}}} \sum_{i=1}^{N(t)} v_i(t), \tag{4.23}$$

$$\Downarrow$$

$$\overline{v}_{\text{s}_{\text{l}}}(t) = \frac{q_{\text{l}}(t)}{k_{\text{l}}(t)} = \frac{1}{N(t)} \sum_{i=1}^{N(t)} v_i(t), \tag{4.24}$$

where we assumed $N(t)$ vehicles are present at time t in the loop detector's segment. The density and flow measurements of consecutive time steps are now temporally averaged over subsequent R_s measurement regions, according to equations (2.9) and (2.17), respectively. In similar fashion as before, the space-mean speed is derived using the fundamental relation (2.33):

$$
\begin{aligned}
k_\mathrm{l} &= \frac{1}{T_\mathrm{mp}} \sum_{t=1}^{T_\mathrm{mp}} k_\mathrm{l}(t) = \frac{1}{T_\mathrm{mp}\, K_\mathrm{ld}} \sum_{t=1}^{T_\mathrm{mp}} N(t), && (4.25)\\
q_\mathrm{l} &= \frac{1}{T_\mathrm{mp}} \sum_{t=1}^{T_\mathrm{mp}} q_\mathrm{l}(t) = \frac{1}{T_\mathrm{mp}\, K_\mathrm{ld}} \sum_{t=1}^{T_\mathrm{mp}} \sum_{i=1}^{N(t)} v_i(t), && (4.26)\\
&\Downarrow \\
\overline{v}_{\mathrm{s}_\mathrm{l}} &= \frac{q_\mathrm{l}}{k_\mathrm{l}} = \sum_{t=1}^{T_\mathrm{mp}} \sum_{i=1}^{N(t)} v_i(t) \Bigg/ \sum_{t=1}^{T_\mathrm{mp}} N(t)\ , && (4.27)\\
&= \sum_{t=1}^{T_\mathrm{mp}} N(t) \frac{1}{N(t)} \sum_{i=1}^{N(t)} v_i(t) \Bigg/ \sum_{t=1}^{T_\mathrm{mp}} N(t)\ , \\
&= \sum_{t=1}^{T_\mathrm{mp}} N(t)\, \overline{v}_{\mathrm{s}_\mathrm{l}}(t) \Bigg/ \sum_{t=1}^{T_\mathrm{mp}} N(t)\ . && (4.28)
\end{aligned}
$$

Our derivations for k_l and q_l as outlined above, also correspond to the generalised definitions encountered in Sections 2.3.1 and 2.3.2, where the total time spent, respectively the total distance travelled, was divided by the area of the measurement region (which corresponds to $T_\mathrm{mp} \times K_\mathrm{ld}$). Furthermore, note that the last equation (4.28) essentially is a weighted mean of the local space-mean speeds $\overline{v}_{\mathrm{s}_\mathrm{l}}(t)$ at each time step t, with the number of vehicles $N(t)$ as weights. This actually corresponds to the space-mean speed based on different substreams, as was computed by equation (2.36).

4.2.3.2 Global measurements on the entire lattice

For the global measurements, we consider N vehicles that are driving in a closed single-lane system, i.e., with a length of $K_\mathcal{L}$ cells (the extension to multi-lane traffic is straightforward). As a consequence, the global density k_g remains constant during the entire measurement period. The derivations of the equations for k_g, q_g, and $\overline{v}_{\mathrm{s}_\mathrm{g}}$, are completely equivalent to those of the previous Section 4.2.3.1, but now with $K_\mathrm{ld} = K_\mathcal{L}$:

$$k_{\text{g}} = \frac{N}{K_{\mathcal{L}}}, \tag{4.29}$$

$$q_{\text{g}} = \frac{1}{T_{\text{mp}}\, K_{\mathcal{L}}} \sum_{t=1}^{T_{\text{mp}}} \sum_{i=1}^{N} v_i(t), \tag{4.30}$$

$$\Downarrow$$

$$\overline{v}_{\text{s}_{\text{g}}} = \frac{q_{\text{g}}}{k_{\text{g}}} = \frac{1}{T_{\text{mp}}\, N} \sum_{t=1}^{T_{\text{mp}}} \sum_{i=1}^{N} v_i(t), \tag{4.31}$$

$$= \frac{1}{T_{\text{mp}}\, N} \sum_{t=1}^{T_{\text{mp}}} N \frac{1}{N} \sum_{i=1}^{N} v_i(t),$$

$$= \frac{1}{T_{\text{mp}}} \sum_{t=1}^{T_{\text{mp}}} \overline{v}_{\text{s}_{\text{g}}}(t). \tag{4.32}$$

Note that, for single-cell TCA models, the global density computed with equation (4.29) corresponds to the occupancy ρ as defined by equation (2.26). For multi-cell models, the number of vehicles is in general less than the number of occupied cells.

4.2.3.3 Local measurements with a detector of unit length

The third technique for measuring macroscopic traffic flow characteristics on a TCA's lattice, bears perhaps the closest resemblance to reality: it is based on an artificial loop detector with unit length, i.e., $K_{\text{ld}} = 1$ cell. The loop detector now explicitly counts all the vehicles that pass it at each time step ΔT during the measurement period T_{mp}.

This type of measurement corresponds to the point measurement region R_{t}, as depicted in Figure 2.3 of Section 2.3. Because of this, the appropriate method for computation is different from the one used in the previous two sections: we now first compute the local flow, using equation (2.12), and the local space-mean speed, using equation (2.27), both for single-lane traffic. The local density is then derived according to the fundamental relation (2.33), resulting in the following set of equations:

$$q_{\text{l}} = \frac{N}{T_{\text{mp}}}, \tag{4.33}$$

$$\overline{v}_{\text{s}_{\text{l}}} = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{v_i} \right)^{-1}, \tag{4.34}$$

$$\Downarrow$$

$$k_{\text{l}} = \frac{q_{\text{l}}}{\overline{v}_{\text{s}_{\text{l}}}}, \tag{4.35}$$

in which N now denotes the number of vehicles that have passed the detector during the measurement period T_{mp}. Because the detector physically occupies one cell and because a vehicle has to 'drive by' in order to get counted, this means that stopped vehicles are ignored: *only moving vehicles are counted*. Note that, as opposed to the previous two techniques, the above measurements no longer denote temporal averages. And because of the temporal region R_{t}, we have to take the harmonic average of the vehicles' speeds v_i in order to obtain the local space-mean speed $\overline{v}_{\text{s}_\text{l}}$. As it turns out, our derivation corresponds perfectly to equation (2.7) which computes the local density at a point in space.

4.2.4 Conversion to real-world units

Converting between TCA and real-world units seems straightforward, as we only need to suitably multiply with or divide by the temporal and spatial discretisations ΔT and ΔX, respectively. However, problems arise due to the discrete nature of a TCA model, involving some intricacies with respect to coordinate systems and their associated units. For example, as defined in Section 4.2.2, a vehicle i's space headway h_{s_i} is always an integer, expressing the number of cells. The same holds true for its space gap g_{s_i} and length l_i. The difficulty now lies in the fact that fractions of cells are not representable in our definition of a TCA model. Keeping equation (2.1) in mind, and noting that $h_{\text{s}_i} \in \mathbb{N}_0$, it follows that $g_{\text{s}_i} + l_i > 0$, which means that either $g_{\text{s}_i} \neq 0$ and/or $l_i \neq 0$.

As a solution, we therefore adopt throughout this dissertation the convention that, without loss of generality, a vehicle's length $l_i \geq 1$ cell (which agrees perfectly with our earlier definitions in Section 4.2.2). Consequently, when a vehicle i is residing in a compact jam (i.e., 'bumper-to-bumper' traffic), its space headway $h_{\text{s}_i} = l$ cells and its space gap $g_{\text{s}_i} = 0$ cells. Our convention thus gives a rigourous justification to formulate the TCA's update rules more intuitively using space gaps, because as already stated in Section 4.1.4, the rules in a TCA rule set are typically not expressed in terms of space headways, but rather in terms of speeds and space gaps (i.e., the distance to the leading vehicle).

In a similar fashion, time headways, time gaps, and occupancy times represent multiples of the temporal discretisation ΔT. But note that, as explained before in Section 4.2.2, these are however no longer constrained to integer values.

In the following two sections, we explain how to convert between coordinate systems of TCA models and the real world. All common variables (e.g., h_{s_i}) are expressed in *TCA units*, except for their 'primed' counterparts (e.g., h'_{s_i}), which are expressed in *real-world units*. The conversions will be done with respect to the following conventions:

- **TCA model**
  - h_{s_i}, g_{s_i}, and l_i are dimensionless integers, denoting a number of cells,

  - h_{t_i}, g_{t_i}, and ρ_i are dimensionless real numbers, denoting a fractional multiple of a time step,
  - k_l and k_g are real numbers, expressed in vehicles/cell,
  - q_l and q_g are real numbers, expressed in vehicles/time step,
  - and v_i, $\overline{v}_{s_l}$, and $\overline{v}_{s_g}$ are real numbers, expressed in cells/time step.

- **Real world**
  - ΔX, h'_{s_i}, g'_{s_i}, and l'_i are real numbers, expressed in metres,
  - ΔT, h'_{t_i}, g'_{t_i}, and ρ'_i are real numbers, expressed in seconds,
  - k'_l and k'_g are real numbers, expressed in vehicles/kilometre,
  - q'_l and q'_g are real numbers, expressed in vehicles/hour,
  - and v'_i, $\overline{v}'_{s_l}$, and $\overline{v}'_{s_g}$ are real numbers, expressed in kilometres/hour.

4.2.4.1 From a TCA model to the real world

Under the previously mentioned convention that $l_i \in \mathbb{N}_0$, we can write the conversions of the microscopic characteristics related to the space and time headways and gaps, and the vehicle lengths and occupancy times, in a straightforward manner:

$$\left\{ \begin{array}{lll} h'_{s_i} = h_{s_i} \cdot \Delta X, & g'_{s_i} = g_{s_i} \cdot \Delta X, & l'_i = l_i \cdot \Delta X, \\ h'_{t_i} = h_{t_i} \cdot \Delta T, & g'_{t_i} = g_{t_i} \cdot \Delta T, & \rho'_i = \rho_i \cdot \Delta T. \end{array} \right. \tag{4.36}$$

Related to equations (4.36), there is a small but important detail that is easily overlooked: we can not just convert between g_{s_i}, g'_{s_i}, l_i, and l'_i without making some assumptions. Because we adopted the convention that $l_i \geq 1$ cell, it follows that $l'_i \geq \Delta X$. So it is not possible to take the real length of a vehicle smaller than the spatial discretisation, because we assumed that the spatial units of a TCA model are all integer values.

The conversions for the macroscopic traffic stream characteristics densities, flows, and space-mean speeds, as well as the microscopic vehicle speed, are as follows:

$$\left\{ \begin{array}{lll} k' & = & k \cdot \dfrac{1000}{\Delta X}, \\ q' & = & q \cdot \dfrac{3600}{\Delta T}, \\ \overline{v}'_s & = & \overline{v}_s \cdot 3.6 \cdot \dfrac{\Delta X}{\Delta T}. \end{array} \right. \tag{4.37}$$

To keep the previous equations clear, we have dropped the subscripts denoting global and local measurements.

It is interesting to see what happens at the jam density, i.e., the maximum density when all cells in the lattice are occupied. As all vehicles are standing still bumper-to-bumper, the associated space gap at this density, equals zero. Computing the space headway according to equation (2.1), results in $h_{s_i} = 0 + l_i$. By virtue of equation (2.11), we can cast this space headway into a density, e.g., for a single-cell TCA model: $k_\mathrm{j} = \overline{h_{\mathrm{s_j}}}^{-1} = \overline{l}^{-1} = l_i^{-1} = 1$. Applying the conversion by means of equations (4.37) and assuming a spatial discretisation $\Delta X = 7.5$ m, results in a real-world jam density $k'_\mathrm{j} = 1000 \div 7.5$ m ≈ 133 vehicles/kilometre. Conversely, if we know k'_j, then we can derive k_j (see Section 4.2.4.2) and hence we have a method to pick a ΔX.

If we were to consider multi-cell traffic, e.g., vehicles with different lengths, then the jam density would be inversely proportional to the average vehicle length. A solution here is to assume a common unit for all vehicle lengths, e.g., the passenger car units (PCU) as explained in Sections 2.3.1.2 and 2.5.1.3. Even though the jam density can be defined for each vehicle class separately, it would be more correct to speak of an *average jam density* at this point due to the temporal and spatial variations in traffic flows.

4.2.4.2 From the real world to a TCA model

Based on equations (4.36), we can write the reverse conversion of the microscopic characteristics in the following manner:

$$\begin{cases} h_{\mathrm{s}_i} = \dfrac{h'_{\mathrm{s}_i}}{\Delta X}, \quad g_{\mathrm{s}_i} = \dfrac{g'_{\mathrm{s}_i}}{\Delta X}, \quad l_i = \dfrac{l'_i}{\Delta X}, \\ h_{\mathrm{t}_i} = \dfrac{h'_{\mathrm{t}_i}}{\Delta T}, \quad g_{\mathrm{t}_i} = \dfrac{g'_{\mathrm{t}_i}}{\Delta T}, \quad \rho_i = \dfrac{\rho'_i}{\Delta T}. \end{cases} \tag{4.38}$$

In order to agree with our previously stated convention, i.e., all spatial microscopic characteristics in a TCA model are integers, equations (4.38) implicitly assume that the real-world spatial variables are multiples of the spatial discretisation (e.g., $h'_{\mathrm{s}_i} = m \cdot \Delta X$ with $m \in \mathbb{N}_0$).

Another possible approach to the spatial conversion to TCA model units, is to *approximate* the real-world values as best as possible, whilst keeping our adopted convention. As $l_i \geq 1$ cell, this leads to the following scheme where we use upward rounding (i.e., ceiling):

$$\begin{cases} h_{\mathrm{s}_i} = \left\lceil \dfrac{h'_{\mathrm{s}_i}}{\Delta X} \right\rceil, \quad l_i = \left\lceil \dfrac{l'_i}{\Delta X} \right\rceil, \\ \Longrightarrow g_{\mathrm{s}_i} = h_{\mathrm{s}_i} - l_i. \end{cases} \tag{4.39}$$

For example, if $\Delta X = 7.5$ m, $l'_i = 4.5$ m, and $g'_{s_i} = 5$ m, then $h'_{s_i} = 4.5 + 5 = 9.5$ m, and from equation (4.39) it follows that $h_{s_i} = 2$ cells, $l_i = 1$ cell, and $g_{s_i} = 2 - 1 = 1$ cell. Because equation (4.39) is only an approximation, it more than often occurs that the computed space headway 'exceeds' the real-world space headway.

In similar spirit, the conversion for the macroscopic characteristics can be easily derived from equations (4.37). However, as opposed to equations (4.38) and (4.39), there is no need for an approximation by means of rounding, because these quantities are real numbers, as mentioned in the introduction of Section 4.2.4.

4.3 Single-cell models

Having discussed the mathematical and physical aspects of cellular automata and TCA models in particular, we now focus on single-cell models. As explained before in Section 4.1.3, each cell can either be empty, or is occupied by exactly one vehicle; all vehicles have the same length $l_i = 1$ cell. Traffic is also considered to be homogeneous, so all vehicles' characteristics are assumed to be the same. In the subsequent sections, we take a look at the following TCA models (accompanied by their suggested abbreviations):

- *Deterministic models*
  - Wolfram's rule 184 (CA-184)
  - Deterministic Fukui-Ishibashi TCA (DFI-TCA)
- *Stochastic models*
  - Nagel-Schreckenberg TCA (STCA)
  - STCA with cruise control (STCA-CC)
  - Stochastic Fukui-Ishibashi TCA (SFI-TCA)
  - Totally asymmetric simple exclusion process (TASEP)
  - Emmerich-Rank TCA (ER-TCA)
- *Slow-to-start models*
  - Takayasu-Takayasu TCA (T^2-TCA)
  - Benjamin, Johnson, and Hui TCA (BJH-TCA)
  - Velocity-dependent randomisation TCA (VDR-TCA)
  - Time-oriented TCA (TOCA)
  - TCA models incorporating anticipation
  - Ultra discretisation, slow-to-accelerate, and driver's perspective

For other excellent overviews of TCA models, we refer the reader to the works of Chowdhury et al. [Cho00], Knospe et al. [Kno04], Nagel [Nag96], Nagel et al. [Nag03a], Schadschneider [Sch00; Sch02a], and Schreckenberg et al. [Sch01].

All following TCA models will be empirically studied using simulations that are performed on a *unidirectional, single-lane lattice* with periodic boundary conditions, i.e., a closed loop with $L_{\mathcal{L}} = 1$. The length of this lattice equals $K_{\mathcal{L}} = 10^3$ cells, which is taken large enough in order to reduce most unwanted *finite-size effects*. Our own experiments indicate that larger lattice sizes do not render any significant advantage, aside from the burden of a larger computation time.

The importance of studying closed-loop, single-lane traffic

There is often a criticism expressed as to why it is important to study the behaviour of traffic flows in such a simplified system. After all, can such a basic system capture all the dynamics of real-life traffic flows, or be even representative of them ? The answer to this question is that, in our opinion, the dynamics of these constrained systems play an important, non-negligible role. For example, when considering traffic flows on most unidirectional two-lane European motorways, drivers are by law obliged to drive on the right shoulder lane, unless when performing overtaking manoeuvres. A frequently observed phenomenon is then that under light traffic conditions (e.g., 10 vehicles/km/lane), a slower moving vehicle (e.g., a truck) is located on the right lane, and is acting as a *moving bottleneck*. As a result, all faster vehicles will line up on the left lane (overtaking on the right lane is prohibited by law), thereby causing a *density* or *lane inversion* [Nag98d; New98; Wol99; Ker04]. It is under these circumstances that the stability of the car-following behaviour plays an important role. Similarly, in densely congested traffic, e.g., the synchronised-flow regime, the same stability may govern the fact whether or not a traffic breakdown is likely to be induced (see Section 2.5.5 for a discussion on the nature of this breakdown). Even for multi-lane traffic, we believe its dynamics are essentially those of parallel single lanes when considering densely congested traffic flows. Another argument for the necessity of studying these simplified systems, is the one given by Nagel and Nelson. They state that this is the easiest way to determine whether or not internal effects of a traffic flow model play a role in, e.g., the spontaneous breakdown of traffic, as all external effects (i.e., the boundary conditions) are eliminated [Nag05]. Nevertheless, when applying these models to real-life traffic networks, closed-loop traffic is not very representative, as the behaviour near bottlenecks plays a far more important role [Hel01b].

All measurements on the TCA models' lattices are based on two possible initial conditions: depending on the nature of the study, we will either use *homogeneous initial conditions* (the default), or a *compact superjam* to start with. In the former case, all vehicles are uniformly distributed over the lattice, implying equal space headways. In the latter case, all vehicles are 'bunched up' behind each other, with zero space gaps. When going from one global density to another, an equivalent method would be to *adiabatically* add (or remove) vehicles to an already homogeneous or jammed state. In our experiments however, we always reset the initial conditions, corresponding to the first method. The simulations[7] ran each time for 10^4 time steps, after an initial period of 10^3 time steps was discarded in order to let transients from the initial conditions in the system die out. Global densities, flows, and space-mean speeds are computed by means of equations (4.29) – (4.32), whereas we use a point detector, i.e., equations (4.33) – (4.35), to compute their local variants. In this latter case, the data points were collected with a measurement period $T_{\text{mp}} = 60$ time steps. Based on these results, we can construct (k_{g},$\overline{v}_{\text{s}_{\text{g}}}$), ($k_{\text{g}}$,$q_{\text{g}}$), ($k_{\text{l}}$,$\overline{v}_{\text{s}_{\text{l}}}$), and ($k_{\text{l}}$,$q_{\text{l}}$) diagrams. *To keep a clear formulation, we will however from now on drop the subscripts denoting global and local measurements.*

For a deeper insight into the behaviour of the space-mean speed $\overline{v}_{\text{s}}$, the average space gap $\overline{g}_{\text{s}}$, and the median time gap[8] $\overline{g}_{\text{t}}$, detailed histograms showing their *distributions* are provided. These are interesting because in the existing literature (e.g., [Cho98; Sch00; Hel01b]) these distributions are only considered at several distinct global densities, whereas we show them for *all* densities. Each of our histograms is constructed by varying the global density k between 0 and 1, computing the space-mean speed, the average space gap and the median time gap for each simulation run. A simulation run consists of 5×10^4 time steps (with a transient period of 500 time steps) on systems of 300 cells, varying the density in 150 steps. Note that a larger size of the system's lattice, has no significant effects on the results, except for an increase of the variance [Mae04g].

> Before giving an elaborate discussion of some of the classical TCA models, it is worthwhile to mention the first historical and practical implementations of traffic cellular automata. Cremer and Ludwig conceived an implementation of traffic flows based on *lattice gas automata* (LGA), which are a special case of cellular automata typically employed when simulating viscous fluids [Cre86]. Their seminal work, using individual bits to represent vehicles, was extended by Schütt, who provided a simulation package for heterogeneous traffic, multi-lane motorways, and network and city traffic [Sch91]. Unfortunately, the developed models were quite inefficient when they were used in setting that called for large scale Monte Carlo simulations [Nag95a].

4.3.1 Deterministic models

In this section, we discuss Wolfram's original rule 184, and its generalisation to higher speeds as proposed by Fukui and Ishibashi's deterministic model. We abbreviate these two TCA models as CA-184 and DFI-TCA, respectively.

[7] All simulations were performed by means of our *Traffic Cellular Automata +* software (developed for the Java™ Virtual Machine); more information can be found in Appendix B.

[8] Note that with respect to the time gaps and time headways, we will work in the remainder of this dissertation with the *median* instead of the arithmetic mean. The median gives more robust results when $h_{\text{t}_i}, g_{\text{t}_i} \rightarrow +\infty$, which, according to equation (2.3), occurs when a vehicle i stops.

4.3.1.1 Wolfram's rule 184 (CA-184)

The first deterministic model we consider, is a one-dimensional TCA model with binary states. As $L_{\mathcal{L}} = 1$, this model is called an elementary cellular automaton (ECA), according to the terminology introduced in Section 4.1.2. If we furthermore assume a local neighbourhood of three cells wide (i.e., a radius of 1), then there are $2^{2^3} = 256$ different rules possible, according to equation (4.3). Around 1983, Stephen Wolfram classified all these 256 binary ECAs [Wol83]. One of these is called *rule 184*, who's name is derived from Wolfram's naming scheme.

Wolfram's scheme is based on the representation of how a cell's state evolves in time, depending on its local neighbourhood. In Figure 4.7, we have provided a convenient visualisation for the evolution of the states in a binary ECA. Here, we can see the state $\sigma_i(t)$ of a central cell i at time step t, together with the states $\sigma_{i-1}(t)$ and $\sigma_{i+1}(t)$ of its two direct neighbours $i-1$ and $i+1$, respectively. All three of them constitute the local neighbourhood $\mathcal{N}_i(t)$ of radius 1 (see also our example of a CA in Section 4.2.1.2). Because states are binary, we can indicate them with a colour, i.e., a black square represents a state of 1 (e.g., state $\sigma_{i+1}(t)$ in Figure 4.7), whereas an empty (white) square represents a state of 0. According to the local transition rule $\delta(i,t)$, the local neighbourhood $\mathcal{N}_i(t)$ is then mapped from t to $t+1$ onto a new state $\sigma_i(t+1)$. The graphical representation in Figure 4.7 thus provides us with an illustrative method to indicate the evolution of $\{\sigma_{i-1}(t), \sigma_i(t), \sigma_{i+1}(t)\} \longmapsto \sigma_i(t+1)$.

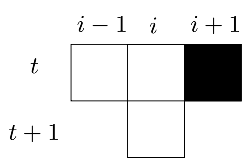


Figure 4.7: An illustrative method for representing the evolution of a cell's state in time, based on its local neighbourhood. We can see the state $\sigma_i(t)$ of a central cell i at time step t, together with the states $\sigma_{i-1}(t)$ and $\sigma_{i+1}(t)$ of its two direct neighbours $i-1$ and $i+1$, respectively. This local neighbourhood is mapped onto a new state $\sigma_i(t+1)$. For binary states, we use a black square to represent a state of 1 (e.g., state $\sigma_{i+1}(t)$), and an empty (white) square for a state of 0. The depicted transition maps the triplet $(001)_2$ onto the state $\sigma_i(t+1) = 0$.

Considering the transition depicted in Figure 4.7, we can see that a complete neighbourhood contains three cells, each of which can be in a 0 (white) or 1 (black) state. So in total, there are $2^3 = 8$ possible configurations for such a local neighbourhood. Wolfram's naming scheme for the binary ECAs is now based on an integer coding of this neighbourhood. Indeed, the local transition rule $\delta(i,t)$ is given by a table lookup containing eight entries, one for each of the possible local neighbourhoods. If we binary sort these eight configurations in the descending order (111), (110), (101), (100), (011), $\ldots$, then we obtain a graphic scheme such as the one in Figure 4.8. As can be seen, for each of the local configurations, a resulting 0 or 1 state is returned for cell i at time step $t+1$. Collecting all resulting states, and writing them in base 2, results in the number $(10111000)_2$. Converting this code to base 10, we obtain the number 184. Wolfram now coded all 256 possible binary ECAs by a unique number in the range from 0 to 255, resulting in 256 rules for these CAs.

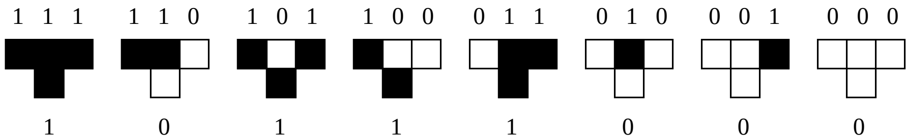


Figure 4.8: A graphical representation of Wolfram's rule 184, which is written as $(10111000)_2$ in base 2. All 8 possible configurations for the local neighbourhood are sorted in descending order, expressing the local transition rule $\delta(i, t)$ as explained by Figure 4.7. For example, the local neighbourhood $(100)_2$ gets mapped onto a state of 1. This has the physical meaning that a particle (black square) moves to the right if its neighbouring cell is empty.

Rule 184 (which we abbreviate as CA-184) is an *asymmetrical* rule because $\delta((110)_2, t) = 0 \neq \delta((011)_2, t) = 1$. It is also called a *quiescent* rule because $\delta((000)_2, t) = 0$ (so all zero-initial conditions remain zero). As an example of the rule's evolution, Figure 4.8 shows that the local neighbourhood $(100)_2$ gets mapped onto a state of 1. If we consider these 1 states as *particles* (i.e., vehicles), and the 0 states as *holes*, then rule 184 dictates that all particles move one cell to the right, on the condition that this right neighbour cell is empty. Equivalently, all holes have the tendency to move to the left for each particle that moves to the right, a phenomenon which is termed the *particle-hole symmetry*.

For a TCA model, we can rewrite the previous actions as a set of rules that are consecutively applied to all vehicles in the lattice, as explained in Section 4.1.4. For the CA-184, we have the following two rules:

R1: *acceleration and braking*

$$v_i(t) \leftarrow \min\{g_{s_i}(t-1), 1\}, \tag{4.40}$$

R2: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t). \tag{4.41}$$

Rule R1, equation (4.40), sets the speed of the i^{th} vehicle, for the current updated configuration of the system; it states that a vehicle always strives to drive at a speed of 1 cell/time step, unless its impeded by its direct leader, in which case $g_{s_i}(t-1) = 0$ and the vehicle consequently stops in order to avoid a collision. The second rule R2, equation (4.41), is not actually a 'real' rule; it just allows the vehicles to advance in the system.

In Figure 4.9, we have applied these rules to a lattice consisting of 300 cells (closed loop), showing the evolution over a period of 580 time steps. The time and space axes are oriented from left to right, and bottom to top, respectively. In the left part, we show a free-flow regime with a global density $k = 0.2$ vehicles/cell, in the right part we have a congested regime with $k = 0.75$ vehicles/cell. Each vehicle is represented as a single coloured dot; as time advances, vehicles move to the upper right corner, whereas congestion waves move to the lower right corner, i.e., backwards in space. From both parts of Figure 4.9, we can see that the CA-184 TCA model constitutes a fully deterministic system that continuously repeats itself. A characteristic of the encountered congestion waves is that they have an eternal life time in the system.

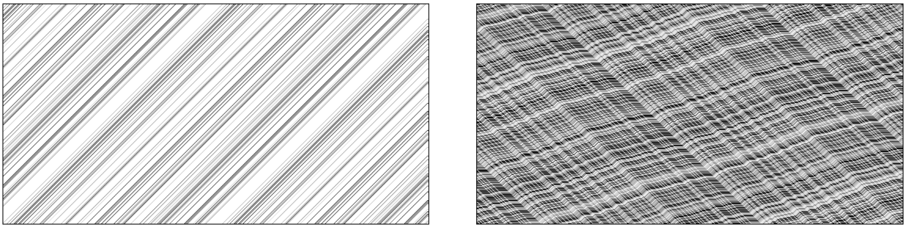


Figure 4.9: Typical time-space diagrams of the CA-184 TCA model. The shown closed-loop lattices each contain 300 cells, with a visible period of 580 time steps (each vehicle is represented as a single coloured dot). *Left:* vehicles driving a free-flow regime with a global density $k = 0.2$ vehicles/cell. *Right:* vehicles driving in a congested regime with $k = 0.75$ vehicles/cell. The congestion waves can be seen as propagating in the opposite direction of traffic; they have an eternal life time in the system. Both time-space diagrams show a fully deterministic system that continuously repeats itself.

In Figure 4.10, we have plotted both the $(k,\overline{v}_s)$ and (k,q) diagrams. As can be seen from the left part, the global space-mean speed remains constant at $\overline{v}_s = 1$ cell/time step, until the critical density $k_c = 0.5$ is reached, at which point $\overline{v}_s$ will start to diminish towards zero where the critical density $k_j = 1$ is reached. Similarly, the global flow first increases and then decreases linearly with the density, below and respectively above, the critical density. Here, the capacity flow $q_{cap} = 0.5$ vehicles/time step is reached. The transition from the free-flowing to the congested regime is characterised by a first-order phase transition. As is evidenced by the *isosceles triangular shape* of the CA-184's resulting (k,q) fundamental diagram, there are only two possible kinematic wave speeds, i.e., +1 and -1 cell/time step. Both speeds are also clearly visible in the left, respectively right, time-space diagrams of Figure 4.9. More analytical details on these values will be provided in the following Section 4.3.1.2.

4.3.1.2 Deterministic Fukui-Ishibashi TCA (DFI-TCA)

In 1996, Fukui and Ishibashi constructed a generalisation of the prototypical CA-184 TCA model [Fuk96]. Although their model is essentially a stochastic one (see Section 4.3.2.3), we will first discuss its deterministic version. Fukui and Ishibashi's idea was two-fold: on the one hand, the maximum speed was increased from 1 to v_{max} cells/time step, and on the other hand, vehicles would accelerate *instantaneously* to the highest

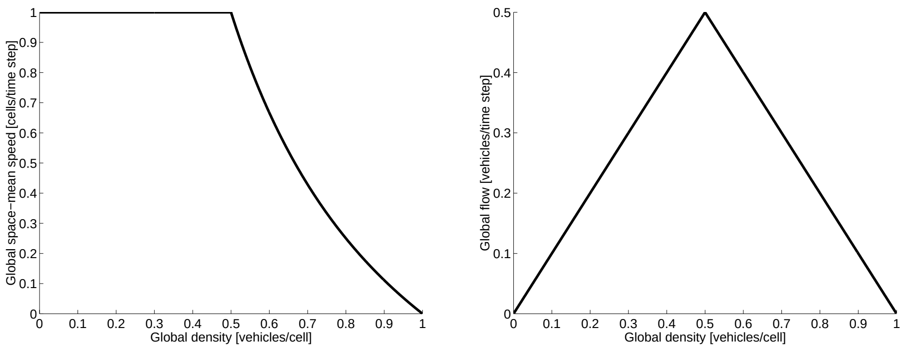


Figure 4.10: *Left:* the (k,$\overline{v}_s$) diagram for the CA-184, based on global measurements on the lattice. The global space-mean speed remains constant at $\overline{v}_s = 1$ cell/time step, until the critical density $k_c = 0.5$ is reached, at which point $\overline{v}_s$ will start to diminish towards zero. *Right:* the CA-184's (k,q) diagram, with its characteristic isosceles triangular shape. The transition between the free-flowing and the congested regime is of a first-order nature.

possible speed. Corresponding to the definitions of the rule set of a TCA model, the CA-184's rule R1, equation (4.40), changes as follows:

R1: *acceleration and braking*

$$v_i(t) \leftarrow \min\{g_{s_i}(t-1), v_{\max}\}. \tag{4.42}$$

Just as before, a vehicle will now avoid a collision by taking into account the size of its space gap. To this end, it will apply an instantaneous deceleration: for example, a fast-moving vehicle might have to come to a complete stop when nearing the end of a jam, thereby *abruptly* dropping its speed from $v_{\max}$ to 0 in one time step.

Due to the strictly deterministic behaviour of the system, the time-space diagrams of the DFI-TCA do not differ much from those of the CA-184. The only difference is the speed of the vehicles in the free-flow regime, leading to steeper trajectories. It is however interesting to study the (k,$\overline{v}_s$) and (k,q) diagrams in Figure 4.11. Here we can see that increasing the maximum speed $v_{\max}$ creates — as expected — a steeper free-flow branch in the (k,q) diagram. Interestingly, the slope of the congested branch does not change, logically implying that the kinematic wave speed for jams remains constant, i.e., -1 cell/time step. This can be confirmed with an analytical kinematic wave analysis, as explained by Nagel [Nag03a].

Based on the behaviour of the vehicles near the critical density, we can analytically compute the capacity flow as follows: in the free-flow regime, all vehicles move with a constant speed of $v_{\max}$ cells/time step. When the critical density is reached, all vehicles drive collision-free at this maximum speed, which implies that $g_{s_i} = v_{\max}$ cells. According to equation (2.1) the space headway $h_{s_i} = v_{\max} + 1$ (because $l_i = 1$

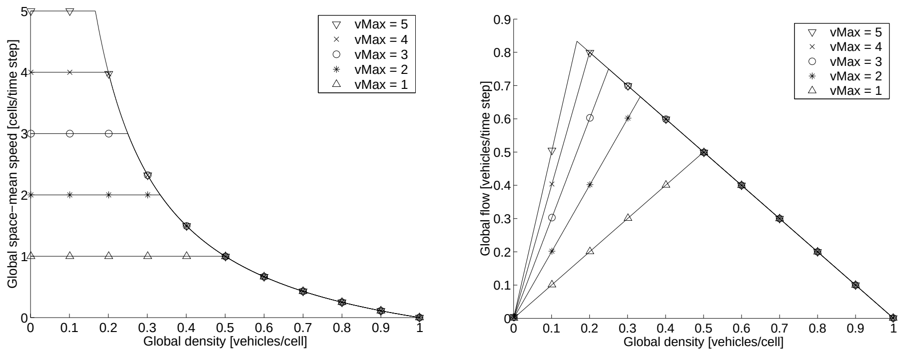


Figure 4.11: *Left:* several $(k,\overline{v}_s)$ diagrams for the deterministic DFI-TCA, each for a different $v_{\text{max}} \in \{1,\ldots,5\}$. Similarly to the CA-184, the global space-mean speed remains constant, until the critical density is reached, at which point $\overline{v}_s$ will start to diminish towards zero. *Right:* several of the DFI-TCA's (k,q) diagrams, each having a triangular shape (with the slope of the congestion branch invariant for the different v_{max}).

for single-cell models). Consequently, equation (2.11) reveals the value for the critical density as:

$$k_c = \frac{1}{\overline{h}_{s_c}} = \frac{1}{v_{\text{max}} + 1}. \tag{4.43}$$

The capacity flow is now computed by means of the fundamental relation (2.33), i.e., $q_{\text{cap}} = k_c\, v_{\text{max}}$:

$$q_{\text{cap}} = \frac{v_{\text{max}}}{v_{\text{max}} + 1}. \tag{4.44}$$

Applying equations (4.43) and (4.44), for, e.g., $v_{\text{max}} = 5$ cells/time step, results in $k_c \approx 0.167$ vehicles/cell and $q_{\text{cap}} \approx 0.83$ vehicles/time step. If we furthermore assume $\Delta X = 7.5$ m and $\Delta T = 1$ s, then both values correspond to 22 vehicles/kilometre and 3000 vehicles/hour, respectively.

As opposed to the instantaneous acceleration in rule R1, equation (4.42), we can also assume a *gradual acceleration* of one cell per time step (the braking remains instantaneous):

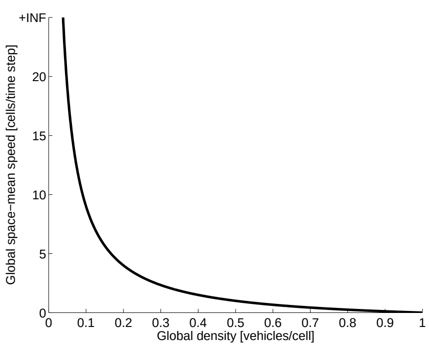
R1: *acceleration and braking*

$$v_i(t) \leftarrow \min\{v_i(t-1)+1, g_{s_i}(t-1), v_{\text{max}}\}. \tag{4.45}$$

However, our experimental observations have indicated that there is no difference in global system dynamics, with respect to either adopting gradual or instantaneous vehicle accelerations.

> There exists a strong relation between the previously discussed deterministic TCA models, and the macroscopic first-order LWR model with a triangular $q_e(k)$ fundamental diagram as mentioned in Section 3.2.1.3. Some of the finer results in this case, are the work of Nagel who extensively discusses some analytical results of both deterministic and stochastic TCA models [Nag96], and the work of Daganzo who explicitly proves an equivalency between two TCA models and the kinematic wave model with a triangular $q_e(k)$ fundamental diagram [Dag06]. More details with respect to such analytical relations, are given in Sections 4.3.2.4 and 4.5.3.

To conclude our discussion of deterministic models, we take a look at what happens in the limiting case where $v_{\max} \rightarrow +\infty$. As can be seen in Figure 4.12, the congested branches in both $(k,\overline{v}_s)$ and (k,q) diagrams grow, at the cost of the free-flow branches which disappear. Interestingly, these diagrams correspond one-to-one with a triangular $q_e(k)$ fundamental diagram that is now expressed in a *moving coordinate system*, as explained by Newell [New99]. In such a simplified system, the critical density $k_c = 0$, with a capacity flow $q_{cap} = 1$.



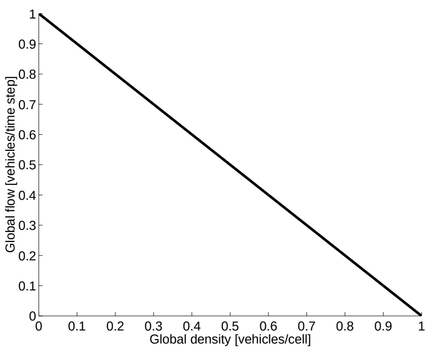


Figure 4.12: *Left:* the $(k,\overline{v}_s)$ diagram for the deterministic CA-184, with now $v_{\max} \rightarrow +\infty$. *Right:* the (k,q) diagram for the same TCA model, resulting in a critical density $k_c = 0$, with a capacity flow $q_{cap} = 1$. This type of diagram corresponds to a simplified triangular $q_e(k)$ fundamental diagram that is expressed in a moving coordinate system.

4.3.2 Stochastic models

The encountered models in the previous section were all deterministic in nature, implying that there can be no spontaneous formation of jam structures. All congested conditions produced in those models, essentially stemmed from the assumed initial conditions. In contrast to this, we now discuss stochastic TCA models (i.e., these are probabilistic CAs) that allow for the spontaneous emergence of phantom jams. As will be shown, all these models explicitly incorporate a stochastic term in their equations, in order to accomplish this kind of real-life behaviour [Nag93b].

4.3.2.1 Nagel-Schreckenberg TCA (STCA)

In 1992, Nagel and Schreckenberg proposed a TCA model that was able to reproduce several characteristics of real-life traffic flows, e.g., the spontaneous emergence of traffic jams [Nag92b; Nag95a]. Their model is called the *NaSch TCA*, but is more commonly known as the *stochastic traffic cellular automaton* (STCA). It explicitly includes a stochastic noise term in one of its rules, which we present in the same fashion as those of the previously discussed deterministic TCA models. The STCA then comprises the following three rules (note that in Nagel and Schreckenberg's original formulation, they decoupled acceleration and braking, resulting in four rules):

R1: *acceleration and braking*

$$v_i(t) \leftarrow \min\{v_i(t-1)+1, g_{s_i}(t-1), v_{\max}\}, \tag{4.46}$$

R2: *randomisation*

$$\xi(t) < p \Longrightarrow v_i(t) \leftarrow \max\{0, v_i(t)-1\}, \tag{4.47}$$

R3: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t). \tag{4.48}$$

Like in both CA-184 and DFI-TCA deterministic TCA models (see Sections 4.3.1.1 and 4.3.1.2), the STCA contains a rule for increasing the speed of a vehicle and braking to avoid collisions, i.e., rule R1, equation (4.46), as well as rule R3, equation (4.48), for the actual vehicle movement. However, the STCA also contains an additional rule R2, equation (4.47), which introduces stochasticity in the system. At each time step t, a random number $\xi(t) \in [0, 1[$ is drawn from a uniform distribution. This number is then compared with a stochastic noise parameter $p \in [0, 1]$ (called the *slowdown probability*); as a result, there is a probability of p that a vehicle will slow down to $v_i(t) - 1$ cells/time step. The STCA model is called a *minimal model*, in the sense that all these rules are a necessity for mimicking the basic features of real-life traffic flows.

> According to Nagel and Schreckenberg, the randomisation of rule R2 captures natural speed fluctuations due to human behaviour or varying external conditions. The rule introduces overreactions of drivers when braking, providing the key to the formation of spontaneously emerging jams.
>
> Although the above rationale is widely agreed upon, much criticism was however expressed due to this second rule. For example, Brilon and Wu believe that this rule has no theoretical background and is in fact introduced quite heuristically [Bri99].

To get an intuitive feeling for the STCA's system dynamics, we have provided two time-space diagrams in Figure 4.13. Both diagrams show the evolution for a global density of $k = 0.2$ vehicles/cell, but with p set to 0.1 for the left diagram, and $p = 0.5$ for the right diagram. As can be seen in both diagrams, the randomisation in the model gives rise to many unstable artificial phantom mini-jams. The downstream fronts of these jams smear out, forming *unstable interfaces* [Nag03a]. This is a direct result of the fact that the intrinsic noise (as embodied by p) in the STCA model is too strong: a jam can always form at *any* density, meaning that breakdown can (and will) occur, even in the free-flow traffic regime. For low enough densities however, these jams can vanish as they are absorbed by vehicles with sufficient space headways, or by new jams in the system [Kra99]. It has been experimentally shown that below the critical density, these jams have finite life times with a cut-off that is about 5×10^5 time steps and independent of the lattice size. When the critical density is crossed, these long-lived jams evolve into jams with an infinite life time, i.e., they will survive for an infinitely long time [Nag94b; Nag95a; Sch99a].

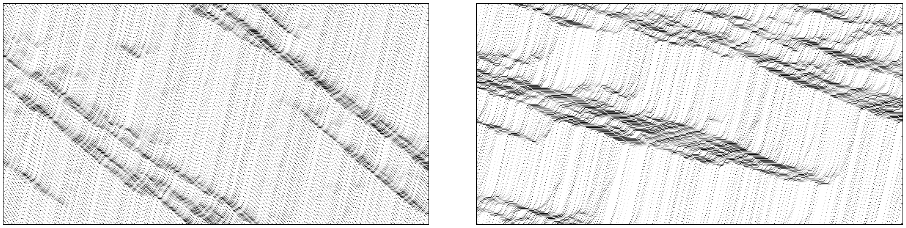


Figure 4.13: Typical time-space diagrams of the STCA model (similar setup as for the CA-184 TCA model in Figure 4.9). Both diagrams have a global density of $k = 0.2$ vehicles/cell. *Left:* the evolution of the system for $p = 0.1$. *Right:* the evolution of the system, but now for $p = 0.5$. The effects of the randomisation rule R2 are clearly visible in both diagrams, as there occur many unstable artificial phantom mini-jams. Furthermore, the speed w of the backward propagating kinematic waves decreases with an increasing p.

In free-flow traffic, a vehicle's speed will fluctuate between v_{max} and $v_{\text{max}} - 1$, due to the randomisation rule R2. We can compute the space-mean speed in the free-flow regime by means of a weighted average. This average corresponds to the probability $1 - p$ for driving with the speed v_{max} and the probability p for slowing down to the speed $v_{\text{max}} - 1$. As such, we get $\overline{v}_{\text{s}_{\text{ff}}} = \sum w_i v_i / \sum w_i = [(1-p)v_{\text{max}} + p(v_{\text{max}} - 1)]/[(1-p)+p] = v_{\text{max}} - p$. In agreement with the space-mean speed observed in the left $(k,\overline{v}_{\text{s}})$ diagram of Figure 4.14, we can state that a vehicle will drive with an average free-flow speed of $\overline{v}_{\text{ff}} = v_{\text{max}} - p$.

As mentioned in Section 4.3.1.2, the slope of the free-flow branch in a (k,q) diagram can be changed by adjusting v_{max}. Similarly, the slope of the congested branch can be changed by tuning the slowdown probability p (note that this also affects the average free-flow speed). Looking at the (k,q) diagram in the right part of Figure 4.14, we note that an increase of p will on the one hand result in a smaller $\overline{v}_{\text{ff}}$, and on the other hand the congested branch will lie lower, with a smaller critical density k_{c}. In this

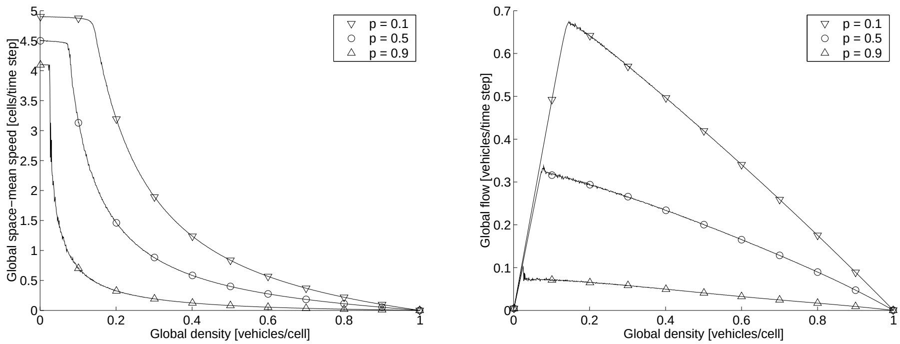


Figure 4.14: *Left:* several $(k,\overline{v}_s)$ diagrams for the STCA, each for a different $p \in \{0.1, 0.5, 0.9\}$. It is clear from the diagram, that a vehicle will drive with an average free-flow speed of $\overline{v}_{\text{ff}} = v_{\max} - p$. *Right:* several (k,q) diagrams for the same STCA models as before. The slope of the congested branch tends toward zero for an increasing slowdown probability p. Note that the seemingly small capacity drops at the critical density in the right part, are actually finite-size effects [Nag95b; Kra97b].

latter case, the speed w of the backward propagating kinematic waves will decrease, an effect that is also visible in the time-space diagrams of Figure 4.13. Note that the presence of noise in the STCA model causes both free-flow and congested branches of the (k,q) diagram to be slightly curved, as opposed to the perfectly linear branches of the deterministic models.

If we set $p = 0$, then the STCA model becomes deterministic; additionally, setting $v_{\max} = 1$ will recover the CA-184 TCA model. In the other deterministic case, when $p = 1$, the system behaves differently: in the congested state, all vehicles will come to a full stop, thereby reducing the global flow in the system to zero. As a result, the congested branch in the (k,q) regime will coincide with the horizontal axis. This implies that the behaviour of a system with $v_{\max}$ and $p = 1$ is totally different than that of a system with $v_{\max} - 1$ and $p = 0$.

Considering local measurements of the density, flow, and space-mean speed, the (k,q) diagrams in Figure 4.15 reveal that an increasing slowdown probability p, results in (i) a lower value for the critical density, (ii) a lower capacity flow, and (iii) a more localised scatter of the data points.

In Figure 4.16, we have plotted a histogram of the distributions of the STCA's vehicles' space gaps, for all global densities $k \in [0, 1]$. For very low densities, the distributions have a distinct maximum, indicating that all vehicles travel with very large space gaps. At higher densities, the maxima of the distributions shift toward smaller space gaps, as more and more vehicles encounter jams, even leading to a reduction of their space gap to zero. Around the critical density however, the distributions are smeared out across consecutive densities, but for each of those densities they exhibit a bimodal structure. Because the STCA contains many jams, the system now contains both vehicles in free-flow traffic, as well as vehicles that are in a congested state (i.e., driving closer to

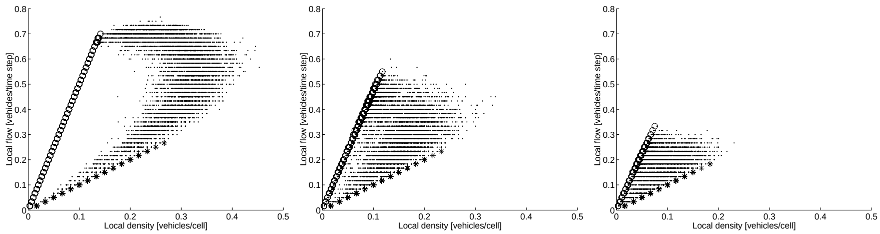


Figure 4.15: Three (k,q) diagrams based on local measurements in the STCA model with $v_{\max} = 5$ cells/time step. *Left:* $p = 0$. *Middle:* $p = \frac{1}{3}$. *Right:* $p = \frac{2}{3}$. Points obtained in the free-flow regime (i.e., for $\overline{v}_s \approx v_{\max}$ cells/time step) are marked with a $\circ$, points obtained in the congested regime with a $\cdot$, and points that imply heavy congestion (i.e., for $\overline{v}_s < 1$ cell/time step) with a $\star$. Note that for these local diagrams, the slopes of the congested branches (indicated by the points marked as $\star$) are the negative of its corresponding slope in a global diagram.

each other) [Cho98; Cho99a; Hel01b; Sch99a].

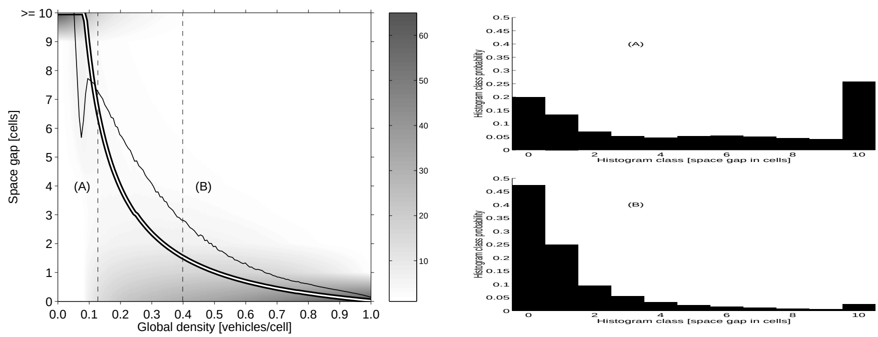


Figure 4.16: A histogram of the distributions of the vehicles' space gaps g_s, as a function of the global density k in the STCA (with $v_{\max} = 5$ cells/time step and $p = 0.5$). In the contour plot to the left, the thick solid line denotes the average space gap, whereas the thin solid line shows its standard deviation. The grey regions denote the probability densities. The histograms (A) and (B) to the right, show two cross sections made in the left contour plot at $k = 0.1325$ and $k = 0.4000$ respectively: for example, in (B), the distribution exhibits a distinct unique maximum at the histogram class $g_s = 0$ cells, corresponding to the dark region in the lower right corner of the contour plot where high global densities occur.

In similar spirit, Figure 4.17 shows the distribution of the vehicles' speeds and time gaps. Corresponding with our observations of the $(k,\overline{v}_s)$ diagrams in Figure 4.14, the left part of Figure 4.17 shows a distinct cluster of probability mass at the histogram class $v_{\max} - p$ for very low global densities. In this region, the standard deviation of the space-mean speed is more or less constant and equal to p. At higher global densities, the distributions become temporarily bimodal, after which they again tend to a unique maximum of 0 cells/time step, corresponding to severely congested traffic;

the standard deviation drastically encounters a maximum at the critical density, after which it declines steadily. With respect to the distributions of the time gaps, the right part of Figure 4.17 shows an rapidly decreasing median time gap as the critical density is approached. At this density, the time gaps settle around a local cluster at the minimum of 1 time step. Going to higher global densities, the number of stopped vehicles increases rapidly, frequently resulting in infinite time gaps. From the critical density on, all distributions exhibit a bimodal structure, corresponding to vehicles that are caught inside a jam, and other vehicles that are able to move freely (possibly at a lower speed) [Cho98; Gho98; Sch99a].

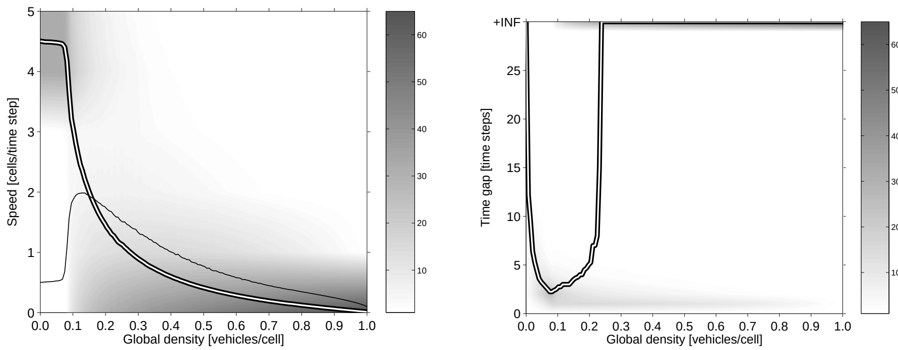


Figure 4.17: Histograms of the distributions of the vehicles' speeds v (*left*) and time gaps g_t (*right*), as a function of the global density k in the STCA (with $v_{\max} = 5$ cells/time step and $p = 0.5$). The thick solid lines denote the space-mean speed and median time gap, whereas the thin solid line shows the former's standard deviation. The grey regions denote the probability densities.

4.3.2.2 STCA with cruise control (STCA-CC)

As mentioned in the previous Section 4.3.2.1, a typical artifact of the STCA model is that it gives rise to many unstable artificial jams. Due to the noise inherent in the model, a jam can always form at any density, even inducing a local breakdown of traffic in the free-flow traffic regime. One way to remedy this, is by stabilising the free-flow branch of the (k,q) diagram. This can be done by inhibiting the randomisation for high-speed vehicles. To this end, Nagel and Paczuski considered again the rules R1 – R3 of the STCA, i.e., equations (4.46) – (4.48), but now complemented with a rule R0 [Nag95b]:

R0: *determine stochastic noise*

$$\begin{cases} v_i(t-1) = v_{\max} & \Longrightarrow \quad p'(t) \leftarrow 0, \\ v_i(t-1) < v_{\max} & \Longrightarrow \quad p'(t) \leftarrow p, \end{cases} \tag{4.49}$$

with now p replaced by $p'(t)$ in the STCA's randomisation rule R2, i.e., equation (4.47). This new rule effectively turns off the randomisation for high-speed vehicles, as only 'jammed' vehicles will now have stochastic behaviour. The resulting TCA model, is called the STCA in the *cruise-control limit*, or STCA-CC for short. If we set the maximum speed $v_{\text{max}} = 1$ cell/time step, then all jams initially present in the system will coalesce with each other, giving rise to one superjam as depicted in Figure 4.18. This superjam has been found to follow a *random walk* in the time-space diagram [Nag95b; Nag96]. Note that $v_{\text{max}} > 1$ cell/time step does not alter the critical behaviour of the model, even though jam clusters are now branching, having regions of free-flow traffic in between them [Nag96].

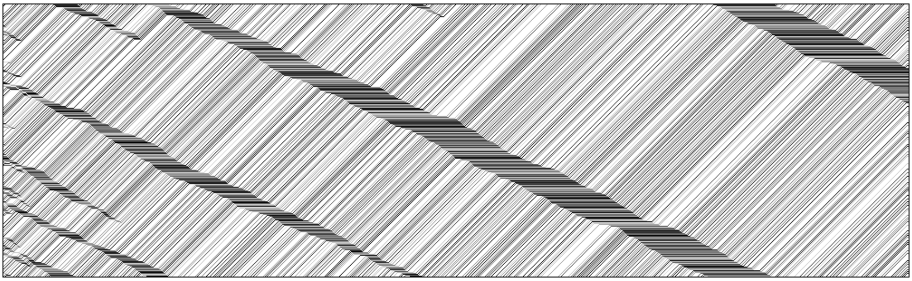


Figure 4.18: A time-space diagram of the STCA-CC model for $v_{\text{max}} = 1$ cell/time step and a global density of $k = 0.4$ vehicles/cell. The shown lattice contains 300 cells, with a visible period of 1000 time steps. We can see over ten initial jams evolving, coalescing over time into one superjam. The system exhibits two distinct phases, i.e., a free-flow and a congested regime with $\overline{v}_s = 1$ and $\overline{v}_s = 0$ cells/time step respectively.

In Figure 4.19, we show the (k,$\overline{v}_s$) and (k,q) diagram of the STCA-CC with $v_{\text{max}} = 5$ cells/time step and $p = 0.2$. As can be seen in the right part, the (k,q) diagram has a typical inverted λ shape, corresponding to the diagram in the left part of Figure 2.11 in Section 2.5.3. The STCA-CC is said to be *bistable*, in that both the free-flow as well as the congested branches of the (k,q) diagram are stable (the former because it is noise-free). Vehicles going from the free-flow to the congested regime encounter at the critical density a phenomenon much like a capacity drop. The reverse transition to the free-flow branch proceeds via a lower density and, correspondingly, a lower flow (which is the outflow q_{out} of a jam). Comparing the right parts of Figure 4.14 and Figure 4.19, it is evident that a destabilisation of the free-flow branch forms the main reason for a lower capacity flow, reached at a lower critical density.

To conclude our discussion of the STCA-CC, we note that the use of cruise control as an ADAS can have unintended consequences. The traffic system can be perceived as having an underlying critical point, at which the life times of jams switch from finite to infinite (see our discussion at the beginning of Section 4.3.2.1). The existence of this point is closely tied to the *self-organised criticality* (SOC) of the STCA model: the outflow from an infinite jam automatically self-organises to a state of maximum attainable flow [Bak88; Nag93b; Nag95a; Tur99]. Stabilising the free-flow branch with cruise-control measures, results on the one hand in traffic higher achievable flows

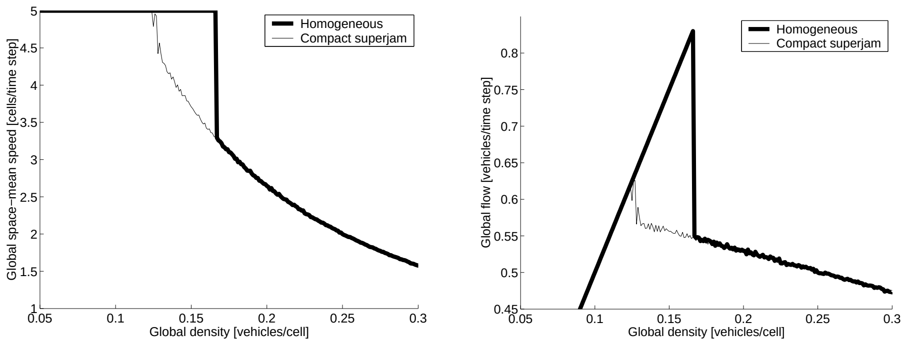


Figure 4.19: Two $(k,\overline{v}_s)$ (*left*) and (k,q) (*right*) diagrams for the STCA-CC model, with $v_{\max} = 5$ cells/time step and $p = 0.2$. The thick solid line denotes global measurements that were obtained when starting from homogeneous initial conditions; the thin solid line is based on a compact superjam as the initial condition (see Section 4.3 for an explanation of these conditions). The right part clearly shows a typical reversed λ shape, which indicates a capacity drop. Note that the observed smaller drop in flow for the compact superjam, is actually a finite-size effect [Nag95b; Kra97b].

which is beneficial, but on the other hand the system is driven closer to its critical point which is more dangerous. At this stage, travel times will experience a high degree of variability, thereby reducing its predictability [Nag94c; Nag95b; Nag95a].

4.3.2.3 Stochastic Fukui-Ishibashi TCA (SFI-TCA)

In Section 4.3.1.2, we discussed the deterministic FI-TCA which is a generalisation of the CA-184 TCA model. In their original formulation, Fukui and Ishibashi also introduced stochasticity, but now only for vehicles driving at the highest possible speed of $v_{\max}$ cells/time step [Fuk96]. We can express the rules of this model, by considering the rules R2 and R3 of the STCA, i.e., equations (4.47) and (4.48), but now complemented with the DFI-TCA's rule R1 for instantaneous accelerations, i.e., equation (4.42) of Section 4.3.1.2, and, as in the STCA-CC model, an extra rule R0:

R0: *determine stochastic noise*

$$\begin{cases} v_i(t-1) = v_{\max} & \Longrightarrow \quad p'(t) \leftarrow p, \\ v_i(t-1) < v_{\max} & \Longrightarrow \quad p'(t) \leftarrow 0, \end{cases} \tag{4.50}$$

with now p replaced by $p'(t)$ in the randomisation rule R2. It can be seen that for $v_{\max} = 1$, the SFI-TCA and STCA models are the same. Furthermore, for $p = 0$ the SFI-TCA becomes fully deterministic, and in contrast to the STCA's zero-flow behaviour (see Section 4.3.2.1), the SFI-TCA's $p = 1$ case corresponds to the STCA with $p = 0$ and $v_{\max} - 1$.

The rationale behind the specific randomisation in the SFI-TCA model, is that drivers who are moving at a high speed, are not able to focus their attention indefinitely. As a consequence, there will be fluctuations at these high speeds. As such, this corresponds to the opposite of a cruise-control limit, e.g., the STCA-CC model. There will be no capacity drop, but the effect on the (k,$\overline{v}_{\text{s}}$) diagram is that its free-flow branch will become slightly downward curving, starting at $\overline{v}_{\text{s}} = v_{\text{max}} - p$ for $k = 0$.

To conclude, we mention the related work of Wang et al., who studied the SFI-TCA both analytically and numerically, providing an exact result for $p = 0$, and a close approximation for the model with $p \neq 0$ [Wan98]. Based on the SFI-TCA, Wang et al. developed a model that is subtly different. They assumed that drivers do not suffer from concentration lapses at high speeds, but are instead only subjected to the random deceleration when they are driving close enough to their direct frontal leaders [Wan01]. And finally, we mention the work of Lee et al., who incorporate anticipation with respect to a vehicle's changing space gap g_{s} as its leader is driving away. This results in a higher capacity flow, as well as the appearance of a synchronised-traffic regime, in which vehicles have a lower speed, but are *all* moving [Lee02].

4.3.2.4 Totally asymmetric simple exclusion process (TASEP)

The simple exclusion process is a simplified well-known particle transport model from non-equilibrium statistical mechanics, defined on a one-dimensional lattice. In the case of open boundary conditions (i.e., the bottleneck scenario), particles enter the system from the left side at an *entry rate* α, move through the lattice, and leave it at an *exit rate* β. The term 'simple exclusion' refers to the fact that a cell in the lattice can only be empty, or occupied by one particle[9]. When moving through the lattice, particles move one cell to the left with probability γ, and one cell to the right with probability δ. When $\gamma = \delta$, the process is called the *symmetric simple exclusion process* (SSEP); if $\gamma \neq \delta$, then it is called the *asymmetric simple exclusion process* (ASEP) [Der92]. Finally, if we set $\gamma = 0$ and $\delta = 1$, the system is called the *totally asymmetric simple exclusion process* (TASEP). If we consider the TASEP as a TCA model, then all vehicles move with $v_{\text{max}} = 1$ cell/time step to their direct right-neighbouring cell, on the condition that this cell is empty.

Updating the configuration of CA essentially amounts to updating the states of all its cells. In general, there are two methods for the update procedure [Raj96; Raj98; Wöl05]:

Sequential update
This updating procedure considers each cell in the lattice one at a time. If all cells are considered consecutively, two updating directions are possible: *left-to-right* and *right-to-left*. There is also a third possibility, called *random sequential update*. Under this scheme and with N particles in the

[9] Note that some authors refer to the TASEP as a 'stochastic' exclusion process, which actually is an incorrect terminology.

lattice, each time step is divided in N smaller substeps. At each of these substeps, a random cell (or vehicle) is chosen and the CA rules are applied to it. As a consequence of the updating procedure, each particle is on average updated after N smaller substeps, which introduces a certain amount of noise in the system. We have depicted several typical time-space diagrams for the ASEP with $\gamma = 1 - \delta$ in Figure 4.20. Furthermore note that a hidden assumption here is that, after completing a substep, the local information is immediately available to the whole system, which can violate causality (as information is now transmitted through the lattice at an infinite speed).

Parallel update
This is the classical update procedure that is used for all TCA models discussed in this dissertation. For a parallel update, all cells in the system are updated in one and the same time step. Compared to a sequential updating procedure, this one is computationally more efficient (note that it is equivalent to a left-to-right sequential update). There is however one peculiarity associated with this updating scheme: because all particles are considered simultaneously, certain lattice configurations can not exist, i.e., the *Garden of Eden* (GoE) states mentioned in Section 4.1.2. An example of such a *paradisiacal state*, is two vehicles right behind each other, with the following having a non-zero speed. This state would imply that in single-lane traffic, the FIFO property was violated and consequently a collision occurred. Such GoE states do not exist when using a random sequential update.

In Figure 4.21, we have depicted two time-space diagrams for the TASEP with a random sequential updating procedure, operating on a closed loop. As can be seen, the diagrams qualitatively look the same, and have some of the same characteristic features of the time-space diagrams in Figure 4.20. For the TASEP, there is no free-flow regime, there are no large jams in the system, and, because of the random sequential update, all vehicles continuously have the tendency to collide with each other. As a consequence, the system is littered with mini-jams in both the low and high density regimes [Nag95a; Nag96]. Note that the TASEP with open boundary conditions exhibits a very rich behaviour, depending on the values for the entry and exit rates α and β, respectively [Kol98; San99; Sch02a].

With respect to the relations between the TASEP with a random sequential update and other models, we mention the following two analogies: on the one hand, the LWR first-order macroscopic traffic flow model (see Section 3.2.1.2) corresponds to the TASEP in the hydrodynamic limit to a noisy and diffusive conservation law, which can be reduced to the LWR model [Nag95a; Nag96]. On the other hand, the TASEP corresponds to the STCA (see Section 4.3.2.1), but now with $v_{\text{max}} = 1$ cell/time step [Cho00; Hel01b].

To gain more insight into the macroscopic behaviour of the TASEP with random se-

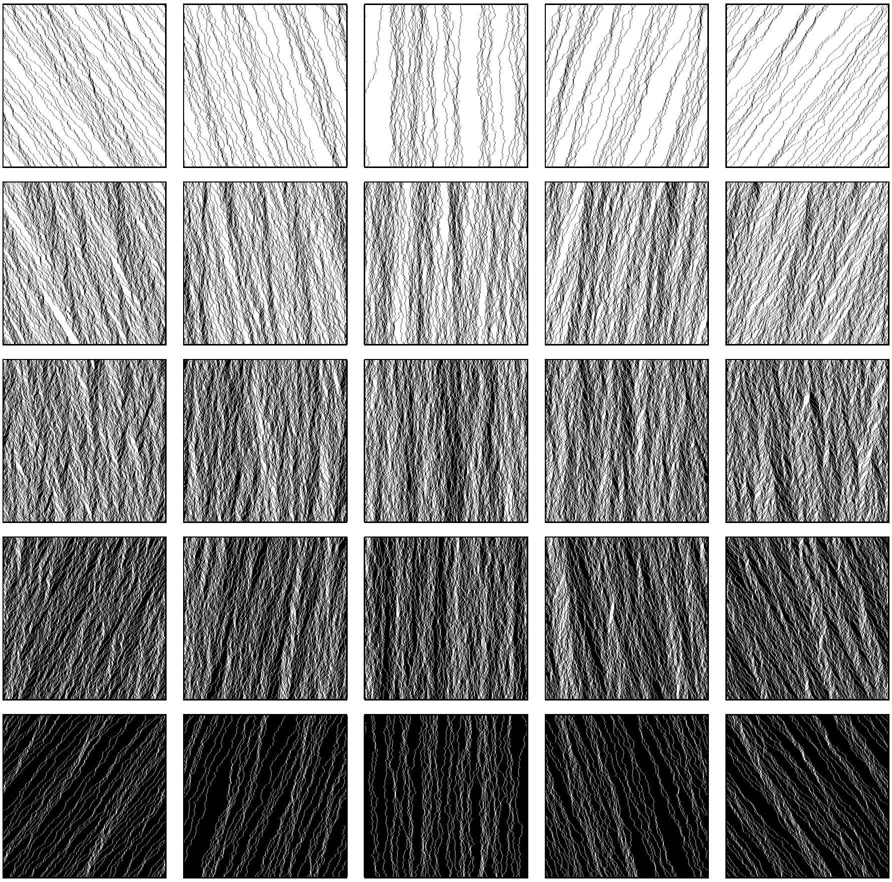


Figure 4.20: Typical time-space diagrams of the asymmetric simple exclusion process (ASEP model) with a random sequential update and $\gamma = 1 - \delta$. The shown lattices each contain 400 cells, with a visible period of 400 time steps (note that for clarity, the space and time axes are located horizontally and vertically, respectively). The global densities in the systems were set for each row to $k \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ vehicles/cell. For each column, the ASEP's probability to move to the left was set to $\gamma \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

quential update, we provide its (k,$\overline{v}_s$) and (k,q) diagrams in Figure 4.22. Looking at the (k,$\overline{v}_s$) diagram on the left part, we notice that the TASEP with $v_{max} = 1$ cell/time step corresponds exactly to Greenshields' original linear relation between the density and the mean speed (see Figure 2.6 in Section 2.5.2.1). This in fact is a further testimony of the close link between the TASEP and the LWR model with a triangular $q_e(k)$ fundamental diagram. Increasing the TASEP's maximum speed, leads to a more curved relation, intersection the vertical axis at the point (0,v_{max}). In any case, the (k,$\overline{v}_s$) diagram also reveals the absence of a distinct free-flow branch, corresponding to the observations of the large amount of mini-jams for all global densities, as could be seen in the time-space diagrams of Figure 4.21.

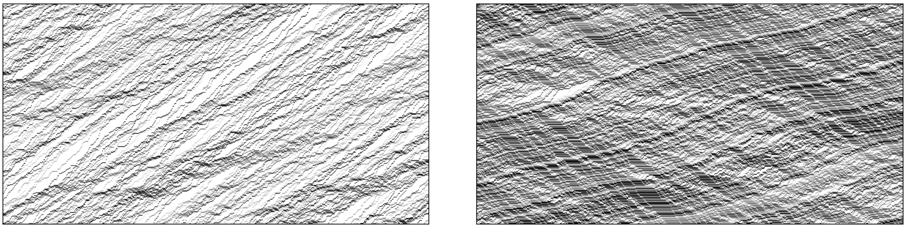


Figure 4.21: Typical time-space diagrams of the TASEP model with a random sequential update. The shown lattices each contains 300 cells, with a visible period of 580 time steps. The global density in the system was set to $k = 0.3$ vehicles/cell (*left*), and $k = 0.7$ vehicles/cell (*right*). The evolution of the system dynamics qualitatively looks the same in both diagrams: the system is littered with mini-jams in both the low and high density regimes.

Studying the (k,q) diagram in the right part of Figure 4.22, we can see that the TASEP corresponds with the STCA for $v_{\max} = 1$ and an arbitrary slowdown probability (e.g., $p = 0.1$). The diagram also shows how the CA-184 leads to a sharp transition between the free-flow and the congested regime, as opposed to the rounded peak of capacity flow at $k = 0.5$ vehicles/cell for the STCA. However, whereas the TASEP also has its capacity flow at the same value, there does not occur such a phase transition as in the other models. Finally, we can see that increasing the maximum speed $v_{\max}$ for the TASEP introduces no significant qualitative changes, except for a skewing towards lower densities [Nag95a].

> Note that with respect to the computational complexity of the implemented TCA models, most measurements in this dissertation took a few hours to obtain, using an Intel P4 2.8 GHz with 512 MB RAM, running the Java™ JDK 1.3.1 under Windows 2000. In sharp contrast to this, are the computations for the TASEP model, which took nearly *two weeks* to complete.

4.3.2.5 Emmerich-Rank TCA (ER-TCA)

Whereas the classical STCA model provided a reasonable qualitative agreement with real-world observations, Emmerich and Rank addressed the quantitative discrepancies between the model and real-world data. To this end, they proposed a variation on the STCA, extending the influence of the space gap on a vehicles updated speed [Emm97].

In their work, Emmerich and Rank fundamentally modified the STCA in two steps: (i) they changed the parallel update procedure to a *right-to-left sequential update procedure* (see Section 4.3.2.4 for more details), and (ii) they changed the behaviour of vehicles that are slowing down. In a nutshell, (i) leads to the important result that vehicles are now able to drive directly behind each other (i.e., with a zero space gap) at high speeds, because the gaps in a traffic stream are used more efficiently. The reason is that due to the specific sequential update, a downstream vehicle is moved

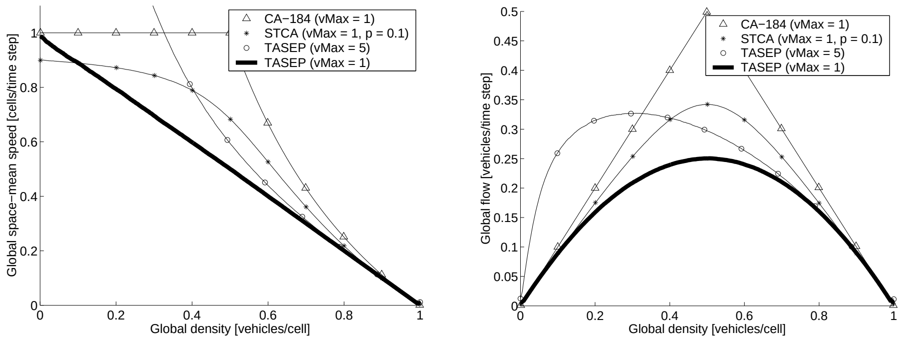


Figure 4.22: A comparison of the (k,$\overline{v}_s$) (*left*) and (k,q) (*right*) diagrams for the CA-184 with $v_{\max} = 1$ (Δ), the STCA with $v_{\max} = 1$ and $p = 0.1$ ($\star$), the TASEP with random sequential update and $v_{\max} = 5$ ($\circ$), and the TASEP with random sequential update and $v_{\max} = 1$ (thick solid line). Some distinct characteristics of the TASEP are the absence of a free-flow regime, and for $v_{\max} = 1$ cells/time step, the exact correspondence with Greenshields linear relation between the density and the mean speed.

first (for a closed loop, the vehicle with the largest space gap is chosen first), after which the next vehicle upstream will see a larger space gap.

Just as the STCA can be seen as a special case of the optimal velocity model, based on a linear optimal velocity function (see Section 3.2.3.2), the ER-TCA model generalises this function by making a vehicle's speed dependent on a variable safe distance and its current speed [Cho00]. This affects (ii), i.e., vehicles that are slowing down: when determining the new speed of a vehicle, the ER-TCA model first checks if the vehicle is within 10 cells of its direct frontal leader. If this is the case, then the vehicle will slow down according to a table lookup in a *gap-speed matrix* $M_{g_{s_i},v_i}$. This matrix is constructed in such a way that collisions are avoided (i.e., $M_{i,j} \leq \min\{i,j\}$):

$$M^{\mathrm{T}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 0 & 1 & 2 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 0 & 1 & 2 & 3 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ 0 & 1 & 2 & 3 & 4 & 4 & 4 & 4 & 4 & 4 & 5 \end{pmatrix} \tag{4.51}$$

The matrix in equation (4.51), conveys the idea that lower speeds require lower space gaps, and that vehicles tend to keep larger space gaps when travelling at higher speeds. This latter effect is also visible in the distribution of the vehicles' space gaps, as visualised in the histograms in the left part of Figure 4.23, where, in contrast to the STCA's space gaps distribution of Figure 4.16, large space gaps are observed for densities near the critical density. Furthermore, because of this mechanism, vehicles will have smoother decelerations, instead of the abrupt slowing down in the STCA model and some of its variations.

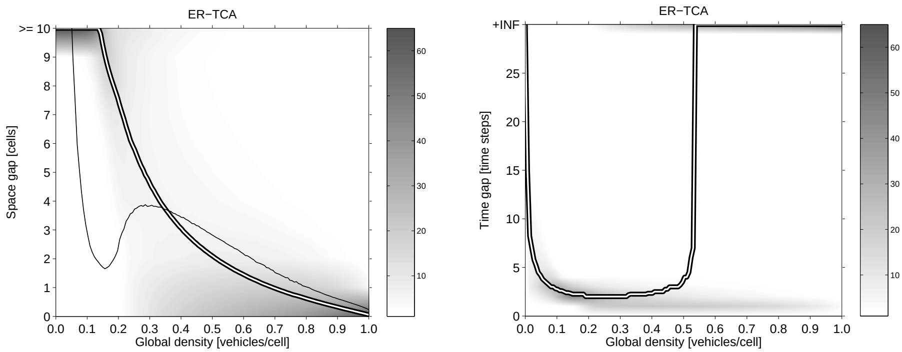


Figure 4.23: Histograms of the distributions of the vehicles' space gaps g_s (*left*) and time gaps g_t (*right*), as a function of the global density k in the ER-STCA (with $v_{max} = 5$ cells/time step and $p = 0.35$). The thick solid lines denote the mean space gap and median time gap, whereas the thin solid line shows the former's standard deviation. The grey regions denote the probability densities.

To understand some of the system dynamics of the ER-TCA model, we have provided several $(k,\overline{v}_s)$ and (k,q) diagrams in Figure 4.24. For $p = 0.35$, we can see in the (k,q) diagram in right part, that the free-flow branch gets *curved*, implying that vehicles travel at a slightly lower speed when they approach the capacity-flow regime. Because vehicles can travel at high speeds in dense platoons, the ER-TCA model can achieve very high capacity flows, even leading to $q > 1$ vehicle/time step. In order to constrain these flows to realistic values, the ER-TCA model needs a quite high slowdown probability, e.g., $p = 0.35$.

These two effects, i.e., a curving of the free-flow branch and an increased capacity flow, are basically what the ER-TCA model is all about, as there is no qualitative change in the congested branch of the (k,q) diagram. There are however some serious drawbacks to the ER-TCA model. First and foremost, the (k,q) diagram is no longer non-monotonic for low densities when the sequential update is replaced by a parallel one [Cho00; Kno04]. Secondly, the model exhibits too large time headways in the free-flow regime when compared with real-world data. This effect is also visible in the distribution of the vehicles' time gaps, as depicted in the histograms in the right part of Figure 4.23, where, in contrast to the STCA's time gaps distribution of Figure 4.17, a large amount of finite time gaps extends well into the region of medium densities. Third, due to the sequential update, the ER-TCA model's downstream jam dynamics are unstable, just as in the STCA model [Kno04]. Fourth, as can be seen from the $(k,\overline{v}_s)$ diagram in the left part of Figure 4.24, for small slowdown probabilities p, the resulting space-mean speed in the system is very unrealistic, even including plateaus of constant speed in the congested regime, e.g., the curve associated with $p = 0.1$ (we consider $p = 0$ as a degenerate case).

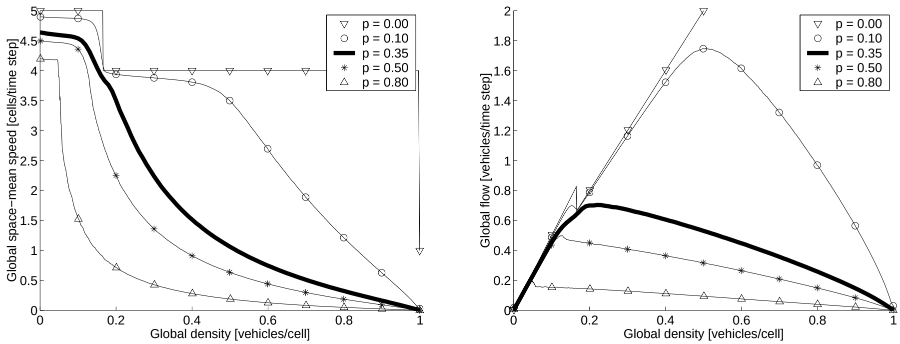


Figure 4.24: *Left:* several (k,$\overline{v}_s$) diagrams for the ER-TCA, each for a different slowdown probability p. It is clear from the diagram, that for low values of p, the resulting diagrams are unrealistic, including plateaus of constant space-mean speed in the congested regime. *Right:* several (k,q) diagrams for the same ER-TCA models as before. Due to the system dynamics in the ER-TCA, very high capacity flows are possible. To constrain these flows, the slowdown probability p has to be quite large in order to obtain realistic results. In both parts of the figure, the thick solid line denotes the original model of Emmerich and Rank, who used a value $p = 0.35$ as their best fit to experimental data.

4.3.3 Slow-to-start models

In order to obtain a correct behavioural picture of traffic flow breakdown and stable jam, it is necessary that a vehicle's minimum time headway or reaction time should be smaller than its escape time from a jam, or equivalently, the outflow from a jam (i.e., the queue discharge rate) must be lower than its inflow [Eis98; Kra99; Kay01; Jos02; Jos03a; Nag03a]. If this is not the case, as in, e.g., the STCA model where both times are exactly the same, then all jams will be unstable, as can be seen in the time-space diagram of Figure 4.13. Because of their unstable jamming behaviour, the previously discussed stochastic models, experience neither a capacity drop nor a hysteresis loop, for which stable jams are a necessary prerequisite. Although the STCA-CC seems to be an exception to this rule, the downstream fronts of its jams are still too unstable, in the sense that new jams can emerge all too easily, which is unrealistic behaviour with respect to real-life traffic flows [Wol99].

As just mentioned, one mechanism that deals with this, is by leaving free-flow traffic undisturbed, and by *significantly reducing the outflow from a jam* once a breakdown occurs, thereby stabilising the downstream front of a jam. Instead of just eliminating the noise in free-flow traffic in the STCA-CC, this reduced outflow can also be accomplished more intuitively, by making the vehicles wait a short while longer before accelerating again from stand still. As such, they are said to be *"slow to start"*.

Note that there exists yet another mechanism that allows for the reproduction of the capacity drop and hysteresis phenomena (we will only briefly mention it here). The approach followed by Werth, is based on the premise that drivers take into account the *speed difference* with their direct frontal leader, instead of just the space gap as was previously assumed. This leads to *Galilei invariant* vehicle-vehicle interactions (i.e., the system dynamics remain the same if a new linear moving coordinate system is substituted in the equations). Interestingly, the metastability in this model is not due to cruise control or slow-to-start rules, but rather a result of the anticipation adopted. The model can exhibit stable dense platoons of fast vehicles, resulting in a stabilisation of the free-flow branch, and consequently leading to hysteretic behaviour [Wer98; Wol99; Cho00].

> With respect to real-world units, we give some typical values associated with the capacity drop and hysteresis phenomena (based on [Wol99]): an outflow $q_{\text{out}} \approx 1800$ vehicles/hour/lane at an associated density of $k_{\text{out}} \approx 20$ vehicles/km/lane, with q_{cap}, k_{crit}, and k_{jam} equal to 2700 vehicles/hour/lane, 20 vehicles/km/lane, and 140 vehicles/km/lane, respectively.

4.3.3.1 Takayasu-Takayasu TCA (T^2-TCA)

In 1993, Takayasu and Takayasu proposed a deterministic TCA model, based on the CA-184 (see Section 4.3.1.1), that incorporated a *delay in acceleration for stopped vehicles* [Tak93]. Their motivation stems from the fact that high-speed vehicles are in general able to decelerate very quickly, but conversely, it takes them a lot longer to attain this high speed when they start from a stopped condition. As such, Takayasu and Takayasu introduced a delay, based on the rationale that a vehicle will only start to move when it recognises movement of its direct frontal leader. Translating this into a rule set, we can write the T^2-TCA's rules based on those of the CA-184, but now with the following modifications (note that $v_{\text{max}} = 1$ cell/time step):

R1: *braking*

$$v_i(t-1) > g_{s_i}(t-1) \quad \Longrightarrow \quad v_i(t) \leftarrow g_{s_i}(t-1), \tag{4.52}$$

R2: *delayed acceleration*

$$v_i(t-1) = 0 \quad \wedge \quad g_{s_i}(t-1) \geq 2 \quad \Longrightarrow \quad v_i(t) \leftarrow 1, \tag{4.53}$$

R3: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t). \tag{4.54}$$

From this rule set it follows that a vehicle will always drive at a speed of one cell/time step, unless it has to brake and stop according to rule R1, equation 4.52. Furthermore, the vehicle is only allowed to accelerate again to this speed of one cell/time step, on the condition that it has a sufficiently large space gap in front, as dictated by rule R2, equation 4.53. As a result, the introduced delay is *spatial in nature*, and it only affects stopped vehicles.

In Figure 4.25, we have depicted the resulting $(k,\overline{v}_s)$ and (k,q) diagrams for the T^2-TCA model. The observed behaviour is similar to that of the STCA-CC model in Section 4.3.2.2, in that the T^2-TCA model also exhibits *bistability*. Starting from homogeneous initial conditions, the space-mean speed in the system undergoes a sharp drop once a vehicle has to stop. The reverse process, i.e., going from the congested to free-flow regime, is accompanied by a smooth continuous transition. Takayasu and Takayasu state that this corresponds to a second-order phase transition, because their order parameter (the sum of the jamming times) follows a power-law distribution, with jam times tending to infinity once the system goes beyond the critical density. With respect to the T^2-TCA's tempo-spatial behaviour, we note that the critical density for the former transition is located at $k_c = 0.5$ vehicles/cell, at which point all vehicles travel at a speed of one cell/time step with all space gaps equal to one cell. The density at which the recovery associated with latter transition occurs, is equal to $k = \frac{1}{3}$ vehicles/cell, at which point all vehicles travel at a speed of one cell/time step, but now with all space gaps equal to two cells. Fukui and Ishibashi later modified the delaying process, resulting in a system that always relaxes to a state in which the space-mean speed oscillates between two values, both smaller than one cell/time step [Fuk97].

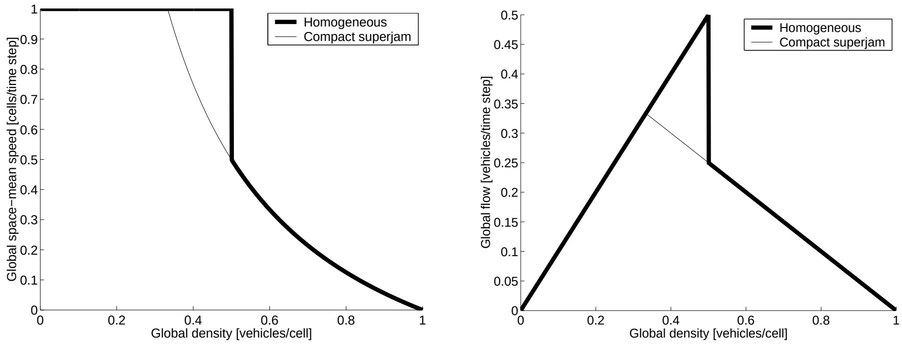


Figure 4.25: Two $(k,\overline{v}_s)$ (*left*) and (k,q) (*right*) diagrams for the T^2-TCA model, with $v_{max} = 1$ cells/time step. The thick solid line denotes global measurements that were obtained when starting from homogeneous initial conditions; the thin solid line is based on a compact superjam as the initial condition (see Section 4.3 for an explanation of these conditions). The right part clearly shows a typical reversed λ shape, which indicates a capacity drop.

> The original background for Takayasu and Takayasu's work, was based on the presence of so-called $1/f$ noise (also known as *pink noise* or *flicker noise*) in the Fourier transformed density fluctuations of motorway traffic. The seemingly random stop-and-go motions of jammed vehicles, could indicate a chaotic behaviour (as opposed to just statistical noise), closely coupled with self-organised criticality (see also the end of Section 4.3.2.2) [Nag93b]. In the free-flow regime of the T^2-TCA model, jams have a finite life time leading to a flat spectrum, as opposed to the congested regime where jams have an infinite life time, leading to a $1/f$ spectrum [Tak93].

Schadschneider and Schreckenberg later provided a generalisation of the T^2-TCA model: keeping $v_{\text{max}} = 1$ cell/time step, they now modified the braking and acceleration behaviour of a vehicle. On the one hand, they kept Takayasu and Takayasu's original acceleration rule R2, equation (4.53), and on the other hand, they allowed a vehicle with a space gap of just one cell to accelerate with a *slow-to-start probability* $1 - p_{\text{t}}$ [Sch97c]. They furthermore also introduced a randomisation for moving vehicles, similar to the STCA (see Section 4.3.2.1), making vehicles stop with a slow-down probability p. Several interesting phenomena occur for certain values of both probabilities p and p_{t}. The modified spatial slow-to-start rule can lead to the appearance of an *inflection point* in the (k,q) diagram at very high densities. The effect gets strongly exaggerated when $p_{\text{t}} \to 1$, at which point a completely blocked state of zero flow appears for all global densities $k \geq 0.5$ vehicles/cell [Sch97c; San99; Cho00].

4.3.3.2 The model of Benjamin, Johnson, and Hui (BJH-TCA)

Around the same time that Takayasu and Takayasu proposed their T^2-TCA model, Benjamin, Johnson, and Hui (BJH) constructed another type of TCA model, using a slow-to-start rule that is of a *temporal nature* [Ben96]. Their BJH-TCA model is based on the STCA (see Section 4.3.2.1), but extended it with a rule that adds a small delays to a stopped car that is pulling away from the downstream front of a queue. Benjamin et al. attribute this rule to the fact that it mimics the behaviour of a driver who momentarily looses attention, or when a vehicle's engine is slow to react. Their slow-to-start rule allows a stopped vehicle to move again with this *slow-to-start probability* $1 - p_{\text{s}}$. If the vehicle did not move, then it tries to move again but this time with probability p_{s}. Due to this peculiar acceleration procedure, all vehicles require a memory that, as mentioned before, makes the slow-start-rule temporal in nature [Cho00]. As a result of this new systematic behaviour, jams will now become less ravelled (as opposed to the STCA), because the slow-to-start rule will have the tendency to merge queues.

The BJH-TCA model was also applied to the description of a motorway with an on-ramp, leading to the conclusions that (i) it actually is beneficial to have jams on the main motorway, due to the fact that these jams homogenise the traffic streams as they compete for stopped vehicles, and (ii) it is desirable to set a maximum speed limit on this main motorway which allows to maximise the performance of the on-ramp. Note that in their discussion, Benjamin et al. used the queue length at the on-ramp as a performance measure. In our opinion, this is not a very good choice as it ignores, e.g., the total time spent in the system, which we believe is a more important measure (see also the work of Bellemans [Bel03] and Hegyi [Heg01] in this respect).

To conclude, we note that the (k,q) diagrams of the BJH-TCA and T^2-TCA models qualitatively look the same, with the exception that the former does not have the possibility of an inflection point, or a density region with zero flow, as was the case for the latter model (see Section 4.3.3.1) [Sch97c; San99].

4.3.3.3 Velocity-dependent randomisation TCA (VDR-TCA)

As already explained in the introduction of this section, reducing the outflow from a jam is responsible for the capacity drop and hysteresis phenomenon. To this end, Barlović et al. proposed a TCA model that generalises the STCA model (see Section 4.3.2.1) by employing an intuitive slow-to-start rule for stopped vehicles [Bar98; Bar03]. Similar to the STCA-CC (see Section 4.3.2.2), the complete rule set for the VDR-TCA is as follows:

R0: *determine stochastic noise*

$$\begin{cases} v_i(t-1) = 0 & \Longrightarrow \quad p'(t) \leftarrow p_0, \\ v_i(t-1) > 0 & \Longrightarrow \quad p'(t) \leftarrow p, \end{cases} \tag{4.55}$$

R1: *acceleration and braking*

$$v_i(t) \leftarrow \min\{v_i(t-1)+1, g_{s_i}(t-1), v_{\max}\}, \tag{4.56}$$

R2: *randomisation*

$$\xi(t) < p'(t) \Longrightarrow v_i(t) \leftarrow \max\{0, v_i(t)-1\}, \tag{4.57}$$

R3: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t). \tag{4.58}$$

As before, in rule R2, equation (4.57), $\xi(t) \in [0,1[$ denotes a uniform random number (specifically drawn for vehicle i at time t) and $p'(t)$ is the stochastic noise parameter, *dependent on the vehicle's speed* (hence the name 'velocity-dependent randomisation'). The probabilities p_0 and p are called the *slow-to-start probability* and the *slow-down probability*, respectively, with $p_0, p \in [0,1]$. Note that Barlović et al. only considered the case with two different noise parameters (i.e., p_0 and p), ignoring the more general case where we can have a noise parameter for each possible speed (i.e., $p_0, \ldots, p_{v_{\max}}$). Their model was also considered for systems with open boundary conditions [Bar02c].

Depending on their speed, vehicles are subject to different randomisations: typical metastable behaviour results when $p_0 \gg p$, meaning that stopped vehicles have to wait longer before they can continue their journey. This has the effect of a reduced outflow from a jam, so that, in a closed system, this leads to an equilibrium and the formation of a *compact jam*. For such a typical situation, e.g., $p_0 = 0.5$ and $p = 0.01$, the tempo-spatial evolution is depicted in Figure 4.26. We can see an initially homogeneous traffic pattern (one *metastable* phase) breaking down and kicking the system into a *phase-separated state*, consisting of a compact jam surrounded by free-flow traffic. In such a state, traffic jams in the system will absorb as many vehicles as is necessary, in order to have a free-flow phase in the rest of the system [Hel01b]. For rather small values of p_0 (i.e., a weak slow-to-start property), the system will exhibit a single phase, just as in the classical STCA model. For larger values of p_0 (i.e., a weak slow-to-start property), the system typically exhibits two phases like in a gas-liquid analogy [Jos03a; Jos03b]. Note that the VDR-TCA can also be equipped with a cruise control, by turning of fluctuations for vehicles driving at the maximum speed $v_{\max}$, thereby stabilising the free-flow branch.

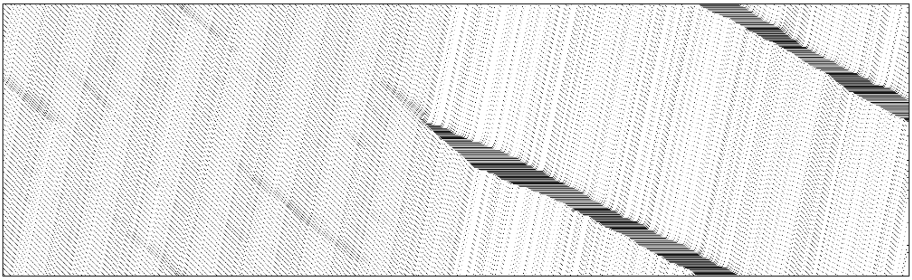


Figure 4.26: A time-space diagram of the VDR-TCA model for $v_{\max} = 5$ cells/time step, $p_0 = 0.5$, $p = 0.01$, and a global density of $k = \frac{1}{6}$ vehicles/cell. The shown lattice contains 300 cells, with a visible period of 1000 time steps. We can see the breakdown of an initially homogeneous traffic pattern. As the phase separation takes place, a persistent compact jam is formed, surrounded by free-flow traffic. The significant decrease of the density in the regions outside the jam results from the jam's reduced outflow.

In the left part of Figure 4.27, we have plotted a histogram of the distributions of the vehicles' speeds, for all global densities $k \in [0, 1]$. Here we can clearly see the distinction between the free-flow and the congested regime: the space-mean speed remains more or less constant at a high value, then encounters a sharp transition (i.e., the capacity drop), resulting in a steady declination as the global density increases. Note that as the critical density is encountered, the standard deviation jumps steeply; this means that vehicles' speeds fluctuate wildly at the transition point (because they are entering and exiting the congestion waves). Once the compact jam is formed, the dominating speed quickly becomes zero (because vehicles are standing still inside the jam). Although most of the weight is attributed to this zero-speed, there is a non-negligible maximum speed present for intermediate densities. If the global density is increased further towards the jam density, this maximum speed disappears and the system settles into a state in which all vehicles either have speed zero or one (i.e., systemwide stop-and-go traffic).

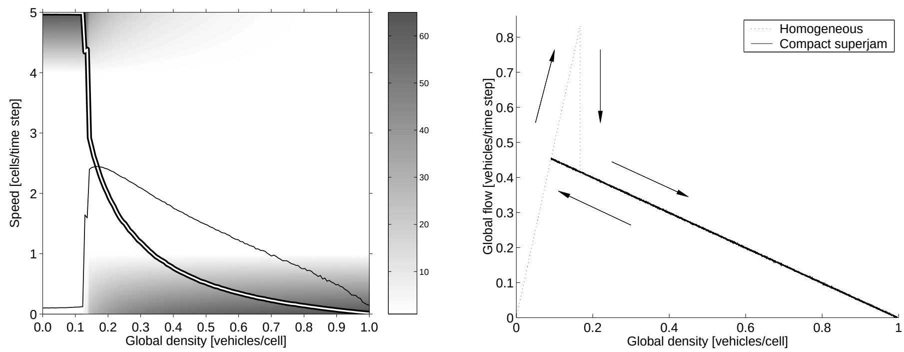


Figure 4.27: *Left:* a contour plot containing the histograms of the distributions of the vehicles' speeds v as a function of the global density k in the VDR-TCA (with $v_{\max} = 5$ cells/time step, $p_0 = 0.5$ and $p = 0.01$). The thick solid line denotes the space-mean speed, whereas the thin solid line shows its standard deviation. The grey regions denote the probability densities. *Right:* a (k,q) diagrams for the same TCA model. The dotted line denotes global measurements that were obtained when starting from homogeneous initial conditions; the solid line is based on a compact superjam as the initial condition. The right part clearly shows a typical reversed λ shape, which indicates a capacity drop.

Studying the (k,q) diagram in the right part of Fig. 4.27, gives us another view of this phase transition. We can see a capacity drop taking place at the critical density, where traffic in its vicinity behaves in a metastable manner. This metastability is characterised by the fact that sufficiently large disturbances of the fragile equilibrium can cause the flow to undergo a sudden decrease, corresponding to a first-order phase transition. The state of very high flow is then destroyed and the system settles into a phase separated state with a large megajam and a free-flow zone. The large jam will persist as long as the density is not significantly lowered, thus implying that recovery of traffic from congestion follows a hysteresis loop. In contrast to the STCA-CC's bistability, the VDR-TCA model is truly *metastable*, because now the free-flow branch in the (k,q) diagram becomes unstable for large enough perturbations. Furthermore, the spontaneous formation of jams in the downstream front that troubled the STCA, is suppressed in the VDR-TCA model.

> Note that if $p_0 \ll p$, then the behaviour of the system will be drastically different. Four distinct traffic regimes emerge in the limiting case where $p_0 = 0$ and $p = 1$; in this case, the model is called *fast-to-start* [Gra01]. In these four regimes, moving vehicles can never increase their speed once the system has settled into an equilibrium. Furthermore, there exists a regime which experiences forward propagating density waves, corresponding to a non-concave region in the system's flow-density relation. For more information, we refer to our work in [Mae04f] and [Mae04g].

4.3.3.4 Time-oriented TCA (TOCA)

Considering the STCA model (see Section 4.3.2.1), Brilon and Wu acknowledged the fact that it is quite capable of reproducing traffic dynamics in urban street networks. However, they also recognised the fact that the model performed rather inadequate when it comes to correctly describing the characteristics of traffic flows on motorways, e.g., compared to field data of a German motorway. Brilon and Wu blamed the unrealistic car-following behaviour of the STCA model for its inferior capabilities. At the core of their argument, they attributed this to the fact that the STCA model is exclusively based on spatial variables (e.g., space headways). In order to alleviate these problems, they proposed to use a model that was based on temporal variables (e.g., time headways), leading to more realistic vehicle-vehicle interactions [Bri99]. The rule set for this time-oriented TCA model (TOCA) is as follows:

R1: *acceleration*

$$\begin{array}{l} g_{s_i}(t-1) > (v_i(t-1) \cdot \overline{g}_{t_s}) \quad \wedge \\ \xi_1(t) < p_{\text{acc}} \\ \quad \Longrightarrow \quad v_i(t) \leftarrow \min\{v_i(t-1)+1, v_{\max}\}, \end{array} \tag{4.59}$$

R2: *braking*

$$v_i(t) \leftarrow \min\{v_i(t), g_{s_i}(t-1)\}, \tag{4.60}$$

R3: *randomisation*

$$\begin{array}{l} g_{s_i}(t-1) < (v_i(t-1) \cdot \overline{g}_{t_s}) \quad \wedge \\ \xi_2(t) < p_{\text{dec}} \\ \quad \Longrightarrow \quad v_i(t) \leftarrow \max\{v_i(t)-1, 0\}, \end{array} \tag{4.61}$$

R4: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t). \tag{4.62}$$

In the above rules, $\xi_1(t), \xi_2(t) \in [0, 1[$ are random numbers drawn from a uniform distribution, $\overline{g}_{t_s} \geq \Delta T$ is the *safe time gap*, p_{acc} is the *acceleration probability*, and p_{dec} is the *deceleration probability*. Because all interactions between vehicles in the STCA are bounded by the update time step, their speeds will never oscillate, leading to a rigid and stable system. As a consequence of the TOCA's temporal rules however, vehicles will now behave more *elastically*, taking a safe time gap into account that allows them to adapt their speeds with a relaxation. In this case, a vehicle will resort to emergency braking (i.e., an instantaneous deceleration) only if it gets too close to its direct frontal leader [Nag03a]. Typical parameter values for the TOCA are $\overline{g}_{t_s} = 1.2$ time steps and $p_{\text{acc}} = p_{\text{dec}} = 0.9$. Brilon and Wu also extended their model with rudimentary rules that allowed for lane changes on unidirectional multi-lane roads.

In the left part of Figure 4.28, we can see a similar tempo-spatial behaviour as with the VDR-TCA (see Section 4.3.3.3), in that an initially homogeneous traffic pattern breaks down, resulting in *dilute jam* that is surrounded by free-flow traffic. The major difference between jamming in the VDR-TCA and TOCA models however, is that in the former model, vehicles come to a complete stop when entering a jam (see Figure 4.26). They remain stationary until they can leave the downstream front of the queue. In contrast to this, the jams in the TOCA model contain moving vehicles. Pushing the global density even further to $k = 0.5$ vehicles/cell as was done in the right part of Figure 4.28, results in a fully developed jam that dominates the entire system and contains temporarily stopped vehicles.

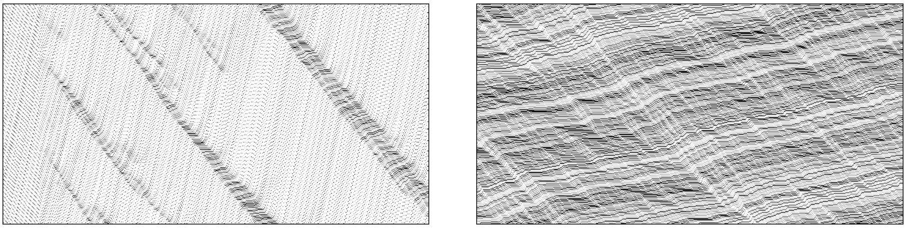


Figure 4.28: Typical time-space diagrams of the TOCA model for $v_{\text{max}} = 5$ cells/time step, $\overline{g}_{t_s} = 1.2$ time steps, and $p_{\text{acc}} = p_{\text{dec}} = 0.9$. The global density was set to $k = \frac{1}{6}$ vehicles/cell (*left*) and $k = 0.5$ vehicles/cell (*right*). The shown lattices each contain 300 cells, with a visible period of 580 time steps. In the left part, we can see the breakdown of an initially homogeneous traffic pattern, resulting in dilute jam that is surrounded by free-flow traffic. In the right part, we see a fully developed jam, dominating the entire system. As can be seen, for moderately light densities, the jams in the TOCA model contain moving vehicles.

Figure 4.29 depicts two groups of (k,q) diagrams for the TOCA model, with $v_{\text{max}} =$ 5 cells/time step. The left part shows four diagrams for different combinations of p_{acc} and $p_{\text{dec}} \in \{(0.9, 0.1), (0.9, 0.9), (0.1, 0.1), (0.1, 0.9)\}$, each time with $\overline{g}_{t_s} = 1.2$ time steps. As can be seen, the default case with $p_{\text{acc}} = p_{\text{dec}} = 0.9$ leads to an inflection point at a moderately high density of $k = 0.5$ vehicles/cell, resulting in two different slopes for the congested branch of the TOCA's (k,q) diagram. At this point, vehicles will have average space gaps less than one cell, and because p_{dec} is rather high, vehicles will have the tendency to slow down (and p_{acc} is smaller then one, so their acceleration is somewhat inhibited). As a result, a large jam, comparable to the system's size, will dominate tempo-spatial evolution. Furthermore, the acceleration probability p_{acc} should take on rather high values, otherwise the global flow in the system is too low because vehicles are not accelerating anymore. In the right part of Figure 4.29, we have shown a large amount of diagrams for different g_{t_s} with $p_{\text{acc}} = p_{\text{dec}} = 0.9$. Here we can see that, for $\overline{g}_{t_s} < \Delta T$, the resulting density-flow curves are non-monotonic. Higher values for $\overline{g}_{t_s}$ in more vehicles that drive more cautiously, apparently leading to higher values for the critical density and the capacity flow. Note that the seemingly small capacity drops at the end of each free-flow branch are in fact finite-size effects [Nag95b; Kra97b].

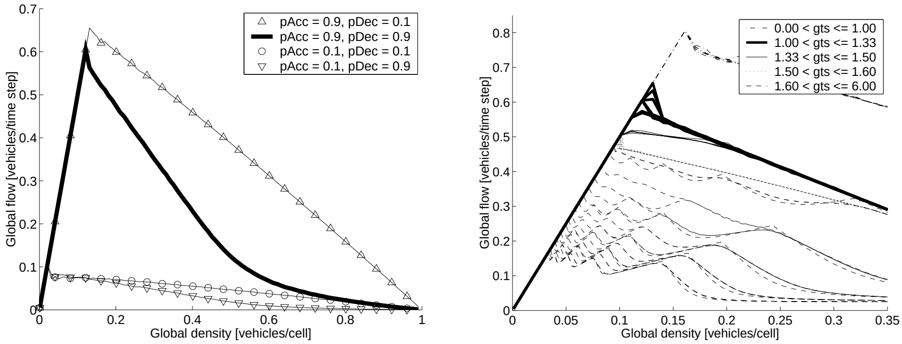


Figure 4.29: Two groups of (k,q) diagrams for the TOCA model, with $v_{\text{max}} = 5$ cells/time step. *Left:* four diagrams for different combinations of p_{acc} and p_{dec}, with $\overline{g}_{t_s} = 1.2$ time steps. *Right:* a large amount of diagrams for different $\overline{g}_{t_s}$ with $p_{\text{acc}} = p_{\text{dec}} = 0.9$. For $\overline{g}_{t_s} < \Delta T$, the resulting density-flow curves are non-monotonic. Note that the seemingly small capacity drops at the end of each free-flow branch are in fact finite-size effects [Nag95b; Kra97b].

In their original paper, Brilon and Wu claim that their TOCA model results in a better agreement with empirical data, a fact which is based on a qualitative comparison of the $(q,\overline{v}_s)$ diagrams[10] [Bri99]. Despite this optimistic view, Knospe et al. later investigated the TOCA model's capabilities more thoroughly. Their conclusions state that a quite large value for the deceleration probability p_{dec} is necessary in order to obtain realistic capacity flows. Although the time headway distribution of a jam's downstream front in the TOCA model is correct with respect to real-life observations, its downstream front moves too fast due to the large deceleration probability. As a result, the jams in the TOCA model are more dilute, as could be seen in Figure 4.28 [Kno04].

4.3.3.5 TCA models incorporating anticipation

One of the models related to anticipative driving (i.e., only taking a leaders' reactions into account, without predicting them), can be found in the work of Krauß et al., who derived a collision-free model based on the STCA (see Section 4.3.2.1), but which uses *continuous* vehicle speeds. Their model can be considered as a simplified version of the Gipps model (see Section 3.2.3.1). Although the model restricts vehicles' deceleration capabilities, it is still able to correctly reproduce the capacity drop and hysteresis phenomena [Kra97b].

Another model with anticipation was proposed by Eissfeldt and Wagner [Eis03]. Their model is based on Krauß's work (see Section 3.2.3.1), and employs a next-nearest-neighbour interaction, which stabilises dense flows and results in a non-unique flow-density relation.

[10]Note that, after personal communication with the authors, it seems they performed a minimisation of the square errors in the $(k,\overline{v}_s)$ diagram. However, in order to get the correct values for calibrating the TOCA's parameters, they just manually guessed, without performing a thorough numerical optimisation.

Recently, Lárraga et al. introduced a TCA model that includes a driver's *anticipation* of the leading vehicle's speed [Lár04]. In contrast to the STCA model (see Section 4.3.2.1), the acceleration and braking rules are decoupled. As a first rule, the standard acceleration towards the maximum speed is applied, after which the randomisation is performed by means of a second rule. Only then, the model considers braking in its third rule; however, the deceleration is not only based on the space gap between both vehicles, but also on an anticipation of the leading vehicle's speed:

R3: *anticipation and braking*

$$v_i(t) \leftarrow \min \left\{ v_i(t), \underbrace{g_{s_i}(t-1) + \left[(1-\alpha_i) \cdot v_{i+1}(t-1) + \frac{1}{2} \right]}_{\text{safe distance}} \right\}, \tag{4.63}$$

with $v_i(t)$ on the right-hand side corresponding to the computed speed after applying rule R2, $[x]$ denoting x rounded to the nearest integer, $v_{i+1}(t-1)$ the speed of the leading vehicle at the current time step, and $\alpha_i \in [0,1]$ an anticipatory driving parameter for the i^{th} vehicle. In their work, Lárraga et al. considered all α_i to be equal.

The interesting aspect of this anticipatory TCA model, is that for certain values of α, it can result in *dense platoons of vehicles*, travelling coherently and thereby leading to forward propagating density structures. In the free-flow regime, the (k,q) diagram also exhibits a slight curvature near the capacity flow, similar to the ER-TCA model (see Section 4.3.2.5). Del Rió and Lárraga later also extended the model to accommodate for multi-lane traffic flows [del05].

4.3.3.6 Ultra discretisation, slow-to-accelerate, and driver's perspective

It is also possible to derive a cellular automaton model, based on the discretisation of a partial differential equation. Starting from a PDE (e.g., the Burgers equation (3.3) from Section 3.2.1.1), we can obtain an finite difference equation by discretising the spatial and temporal dimensions, resulting in a model that still has continuous state variables. As a further step, we can now also discretise these state variables, using a process called the *ultra-discretisation method* (UDM) [Tok96]. The result of the UDM can be interpreted as a cellular automaton in the *Euler representation*. The latter means that for a TCA model, a road is considered to be a field, whereby the individual cars are not distinguished [Nis01a]. The interesting part of this type of CA is that its cells are allowed to hold multiple vehicles, which makes it possible to implicitly model multi-lane traffic in a simplified sense (because the effects of lane changes are neglected) [Cho00]. As a next step, this obtained CA can be cast in its *Lagrangian representation*, by means of an *Euler-Lagrange transformation* [Nis01a; Mat03]. The resulting Lagrange representation treats the positions of all vehicles individually, thus leading to the well-know position-based rule sets of the TCA models discussed in this dissertation.

Nishinari proposed an interesting TCA model, based on the above UDM scheme. Their discretisation leads to the so-called *Burgers cellular automaton* (BCA), which is for single-lane traffic equivalent to the CA-184 TCA model (see Section 4.3.1.1) [Nis99; Nis01b]. Emmerich et al. also provided a TCA model, by applying the UDM scheme to a Korteweg-de Vries equation. In contrast to the BCA model, their work resulted in a second-order TCA model because the CA's global map not only needs the configuration at the previous time step $t-1$, but also the configuration at time step $t-2$ [Emm98; Cho00].

Nishinari et al. recently extended the BCA model, thereby allowing for slow-to-start effects with $v_{\text{max}} > 1$ cell/time step [Nis04]. Their model contains a rule similar to the classical notion of slow-to-start rules, but now generalised for moving vehicles, leading to the terminology of a *slow-to-accelerate* rule. Taking the idea of anticipation one step further, they also incorporated a *driver's perspective*, meaning that a vehicle will base its acceleration and braking decisions not only on the basis of its space gap and the anticipated speed of the vehicle ahead, but also on the space gap with the *next* leading vehicle (or even a vehicle located more downstream). As a result, the model exhibits *multiple metastable branches* in the (k,q) diagram, as can be seen in Figure 4.30. For the lowest metastable branch, vehicles inside jams will come to a complete stop. In contrast to this, vehicles will still be able to move forward inside jams for the higher branches. Note that depending on the strength of a local perturbation, traffic will shift from the highest branch to one of the lower branches. Finally, Nishinari et al. also combined the model with the classical STCA (see Section 4.3.2.1), thereby allowing for stochasticity in both the acceleration and braking rules.

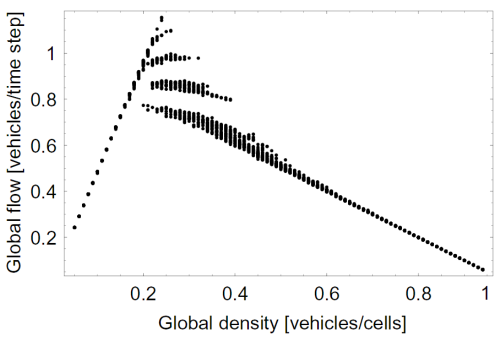


Figure 4.30: A (k,q) diagram of Nishinari et al.'s extended BCA model, with $v_{\text{max}} = 5$ cells/time step, $\Delta T = 1.3$ s, $\Delta X = 7.5$ m, and a driver's perspective of two vehicles ahead. The resulting diagram exhibits multiple metastable branches. Vehicles inside jams come to a complete stop only for the lowest metastable branch; for the higher branches, vehicles inside jams are still able to move forward. Depending on the strength of a local perturbation, traffic will shift from the highest branch to one of the lower branches (image reproduced after [Nis04]).

4.4 Multi-cell models

Whereas all the previously discussed TCA models were based on a single-cell setup, this section introduces some of the existing multi-cell TCA models (still for single-lane traffic). In a multi-cell model, a vehicle is allowed to span a number of consecutive cells in the longitudinal direction, i.e., $l_i \geq 1$ cell.

In the subsequent sections, we discuss several multi-cell TCA models encountered in literature. We first start with an overview of the artifacts that can be introduced when switching to a multi-cell setup. Subsequently, we describe three multi-cell TCA models, which have more intricate rule sets than the simple models of Section 4.3:

- Helbing-Schreckenberg TCA (HS-TCA)
- Brake-light TCA (BL-TCA)
- The model of Kerner, Klenov, and Wolf (KKW-TCA)

Note that with respect to the measurements performed on the TCA models' lattices, we assume homogeneous traffic flows, i.e., all vehicles have the same length. This allows us, after suitable adjustment with the average vehicle length $\overline{l} = l_i$, to express the global density as $k_{\text{g}} \in [0, 1]$.

4.4.1 Artifacts of a multi-cell setup

It might seem that a translation of the classical STCA model (see Section 4.3.2.1) into a multi-cell version would be straightforward. However, using a finer discretisation introduces a very specific artifact, i.e, *hysteresis*. In order to investigate this phenomenon, we have performed several experiments based on a multi-cell translation of the STCA model (now called the MC-STCA). In what follows, we assume a closed-loop lattice consisting of 10^5 cells. The simulations ran each for 5×10^5 time steps, with $\Delta T = 1$ s.

Setting the slowdown probability to $p = 0.5$, the left part of Figure 4.31 shows the resulting (k,q) diagrams for different spatial discretisations, each time for homogeneous initial conditions. The average vehicle length was set to $\overline{l} \in \{2, 4, 8, 16, 32, 64\}$ cells. In these experiments, we also scaled the maximum speed v_{max} correspondingly (e.g., if $\overline{l} = 4$ cells, then v_{max} would become $5 \times 4 = 20$ cells/time step), as can be seen from the coinciding free-flow branches in the left part in Figure 4.31. We also notice that an increase of the average vehicle length apparently results in a higher critical density, with an associated higher capacity flow. Furthermore, the flow seems to encounter a *capacity drop* at this critical density.

What causes this capacity drop ? To answer this question, we must first consider what happens in the deterministic case where $p = 0$. Here, our experiments have shown that there is no difference between a single-cell and a multi-cell setup. Setting

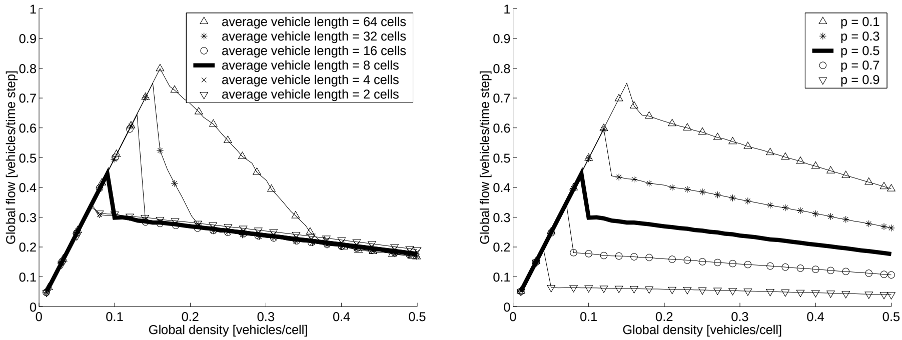


Figure 4.31: *Left:* several (k,q) diagrams of the MC-STCA, for $\bar{l} \in \{2, 4, 8, 16, 32, 64\}$ cells and $p = 0.5$. As can be seen, an increase of the average vehicle length apparently results in a higher critical density, with an associated higher capacity flow (followed by a capacity drop). *Right:* the same setup for the MC-STCA, but now with a fixed $\bar{l} = 8$ cells and $v_{\max} = 5 \times 8 = 40$ cells/time step. The (k,q) diagrams depict the results of changing the slowdown probability $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$: an increase of p, leads to decrease of both the critical density and the capacity flow.

$p > 0$, the randomisation rule R2, equation (4.47), introduces fluctuations in the high speeds of vehicles in free-flow traffic. However, these speed fluctuations are actually small compared to the vehicles' speeds themselves. Because of this limited influence, the free-flow branch of the (k,q) diagrams remains *very stable*. The smaller the discretisation, i.e., the larger the average vehicle length, the more stable the free-flow branch becomes for larger densities (note however that the capacity drop gets less pronounced for increasing average vehicle lengths). This capacity drop behaviour due to a stabilisation effect, is akin to the observations in the STCA's cruise-control limit (see Section 4.3.2.2), and thus different from the VDR-TCA (see Section 4.3.3.3), where a reduced outflow from a jam causes the drop in flow [Kno04]. In contrast to this, random initial conditions or a superjam to start the simulations with, will always lead to the congested branch, thereby indicating a hysteretic phase transition. As the left part of Figure 4.31 indicates, changing the discretisation level of the STCA, by adjusting the average vehicle length and relatively keeping the same maximum speed, has only an effect on the length of the free-flow branch; the traffic dynamics in the congested regime remain the same.

Holding $\bar{l}$ fixed at 8 cells and $v_{\max} = 5 \times 8 = 40$ cells/time step, the right part of Figure 4.31 shows the resulting (k,q) diagrams for different values of the slowdown probability $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. It is clear that an increase of p, leads to a decrease of both the critical density and the capacity flow. Note that the size of the capacity drop remains approximately the same for the different p.

To conclude, we mention the work of Grabolus who performed extensive numerical studies on the STCA. He also noted that it is possible to translate any multi-cell STCA variant into an *equivalent* single-cell STCA model, by suitably adjusting the values of the density and the maximum speed [Gra01].

> Interestingly, the use of a smaller discretisation was already considered by Barrett et al. in the early course of the TRANSIMS project (see Section 3.1.3.3) [Bar95]. In their work, they introduce the terminology of *multi-resolution* TCA models, corresponding to our multi-cell setup. Although they discuss several methods for integral refinements of the TCA's lattice, they do not make any mention of the observed unexpected hysteresis phenomenon introduced by a finer discretisation.

4.4.2 Advanced multi-cell models

Having discussed the repercussions of switching to a multi-cell setup, we now illustrate three TCA models that have more complex rule sets. We discuss their properties by means of time-space diagrams, fundamental diagrams of global and local measurements, and histograms of the distributions of the space and time gaps.

4.4.2.1 The model of Helbing and Schreckenberg (HS-TCA)

Similar in spirit as the STCA (see Section 4.3.2.1) and the ER-TCA (see Section 4.3.2.5), Helbing and Schreckenberg proposed their HS-TCA model in analogy with the optimal velocity model (see Section 3.2.3.2) [Hel99b]. In fact, their model can be seen as a direct discretisation of the OVM, with the following rule set:

R1: *acceleration and braking*

$$v_i(t) \leftarrow v_i(t-1) + \lfloor \alpha \left(V(g_{\mathrm{s}_i}(t-1)) - v_i(t-1) \right) \rfloor, \tag{4.64}$$

R2: *randomisation*

$$\xi(t) < p \Longrightarrow v_i(t) \leftarrow \max\{0, v_i(t) - 1\}, \tag{4.65}$$

R3: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t). \tag{4.66}$$

The function $V(g_{\mathrm{s}_i})$ in rule R1, equation (4.64), is the discrete version of the optimal velocity function; it is specified in the form of a lookup table, containing speed entries for each space gap (see Table 4.1). The parameter α is similar to the one in equation (3.30) and has the following meaning: higher values indicate an almost instantaneous adaptation of the vehicle's speed to the OVF, whereas lower values denote an increasing inertia and longer adaptation times [Hel99b]. However, as stated by Chowdhury et al. and Knospe et al., the role of α is a bit unclear as it does not exactly correspond to the time scale of the adaptation to the OVF (which is the case for the original optimal velocity model) [Cho00; Kno04]. Furthermore, as mentioned in Section 3.2.3.2, certain values for α can, in combination with the OVF, lead to collisions between vehicles (because α reduces a vehicle's braking capability). Knospe et al. later provided the necessary conditions that guarantee collision-free driving, and avoid the possible backward moving of vehicles [Kno04]. Note that, similar to the Fukui-Ishibashi models (see Sections 4.3.1.2 and 4.3.2.3), vehicles are allowed to accelerate instantaneously in the HS-TCA model. The model is stochastic, in that it introduces randomisation by means of rule R2, equation (4.65), with $\xi(t) \in [0, 1[$ a random number drawn from a uniform distribution.

g_{s_i} (cells)	$V(g_{s_i})$ (cells/time step)	g_{s_i} (cells)	$V(g_{s_i})$ (cells/time step)
0, 1	0	11	8
2, 3	1	12	9
4, 5	2	13	10
6	3	14, 15	11
7	4	16 – 18	12
8	5	19 – 23	13
9	6	24 – 36	14
10	7	$\geq$ 37	15

Table 4.1: A possible optimal velocity function (OVF) for the TCA model of Helbing and Schreckenberg (HS-TCA). The OVF is represented as a table, giving the optimal speed $V(g_{s_i})$ associated with each possible space gap g_{s_i}.

In Figure 4.32, we have given two time-space diagrams of the HS-TCA for global densities $k = 0.25$ and $k = 0.40$ vehicles/cell. The length of a vehicle was $l = 2$ cells, $p = 0.001$, $\alpha = 1 \div 1.3$, $v_{\text{max}} = 15$ cells/time step, $\Delta T = 1$ s, and $\Delta X = 2.5$ m. Due the small slowdown probability, the system dynamics are strongly deterministic, totally dependent on the initial (homogeneous) conditions. In the left diagram we can observe how vehicles can accelerate instantaneously when exiting a jam. Note that for higher densities, all jams become dense and compact, always containing stopped vehicles, as is depicted in the right diagram. Because of the non-linearity introduced by the discretised optimal velocity function, all tempo-spatial patterns in the system are of a chaotic nature (i.e., nonlinear with stochastic noise) [Kno04].

The $(k,\overline{v}_s)$ and (k,q) diagrams in Figure 4.33 are based on local and global measurements. A feature of these diagrams is that the local measurements tend to form clusters around certain space-mean speeds (see the left part of Figure 4.33): these clusters correspond to the speeds dictated by the discretised optimal velocity function of Table 4.1, each time associated with an average space gap corresponding to the inverse of the locally measured density. As a result, the (k,q) diagram in the right part of Figure 4.33 shows several branches, each one with a different OVF speed. The lowest branch corresponds to the speed of the backward propagating waves, i.e., the jam speed. Even more striking, is that from a certain finite density $k \ll 1$ vehicle/cell on, all vehicles always come to a full stop and the flow in the system becomes zero [Hel99b].

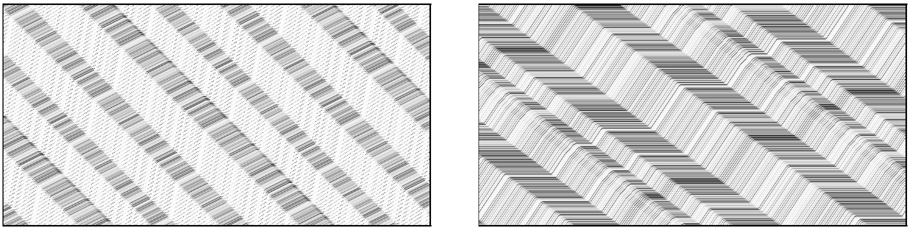


Figure 4.32: Typical time-space diagrams of the HS-TCA model, with $l = 2$ cells, $p = 0.001$ $\alpha = 1\div 1.3$, and $v_{\text{max}} = 15$ cells/time step. The shown closed-loop lattices each contain 300 $\times 2 = 600$ cells, with a visible period of 580 time steps. The global density k was set to 0.25 vehicles/cell (*left*) and 0.40 vehicles/cell (*right*). The formation of congestion waves leads to dense, compact jams containing stopped vehicles. Vehicles strive to decelerate smoothly, but are allowed to accelerate instantaneously when exiting jams fronts.

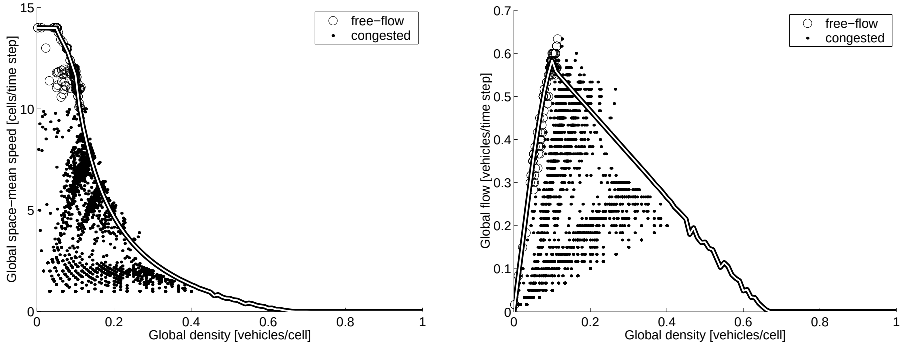


Figure 4.33: The $(k,\overline{v}_{\text{s}})$ (*left*) and (k,q) (*right*) diagrams for the HS-TCA, obtained by local and global measurements. The local measurements tend to form clusters around certain space-mean speeds, corresponding to the speeds dictated by the discretised optimal velocity function of Table 4.1. These clusters are visible in the right diagram as branches with different slopes. Remarkably, from a certain finite density $k \ll 1$ vehicle/cell on, all vehicles always come to a full stop and the flow in the system becomes zero.

To conclude our discussion of the HS-TCA, we give the histograms of the distributions of the space and time gaps in the left and right parts, respectively, of Figure 4.34. The most prominent features of these histograms, are that (i) there exist small clusters of probability mass between certain space gaps (i.e., 15 – 20, 25 – 25, and 35 – 40 cells), corresponding to groups of vehicles, (ii) for higher densities, we can observe a spread-out cluster of probability mass, corresponding to the lowest local measurements in the left part of Figure 4.33, and (iii) in contrast to the previous TCA models, the median of the time gap for the HS-TCA is already very small for densities $k < 0.1$.

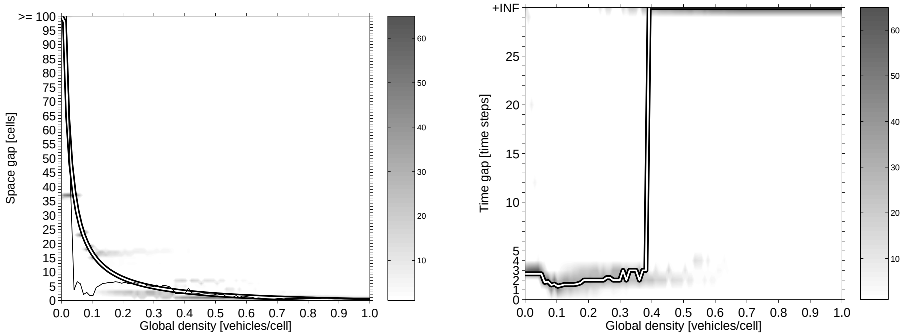


Figure 4.34: Histograms of the distributions of the vehicles' space gaps g_s (*left*) and time gaps g_t (*right*), as a function of the global density k in the HS-TCA. The thick solid lines denote the mean space gap and median time gap, whereas the thin solid line shows the former's standard deviation. The grey regions denote the probability densities.

The HS-TCA might seem an interesting improvement, as it is being based on a discretisation of the optimal velocity model. But although its authors state that it *"reproduces many of the empirically observed features"* [Hel99b], Knospe et al. showed several shortcomings in the model [Kno04]: care must be taken to avoid collisions, and the model fails to reproduce the synchronised-flow regime entirely. This latter can be understood by looking at the dense, compact structure of jams in the time-space diagrams of Figure 4.32, and the occurrence of branches with distinct speeds as in the right part of Figure 4.33.

4.4.2.2 Brake-light TCA (BL-TCA)

Recently, an interesting idea was pursued by Knospe et al.; their TCA model includes *anticipation* effects (see also Section 4.3.3.5), introduced by equipping the vehicles with *brake lights* [Kno00]. The focus of this (and the following) TCA model lies in a correct reproduction of the three phases of traffic as introduced by Kerner et al. (see Section 2.5.4). In a sense, the BL-TCA incorporates many of the features encountered in previously discussed single-cell TCA models. First of all, the BL-TCA has randomisation for spontaneous braking. Secondly, it has slow-to-start behaviour for the capacity drop and hysteresis phenomena. Moreover, it incorporates anticipation which can lead to a stabilisation of the free-flow branch. Finally, it includes elements for reproducing synchronised traffic. These latter two aspects clearly go beyond the standard incentive for drivers to avoid collisions. As such, it is the desire for smooth and comfortable driving (which resembles *human behaviour*), that is responsible for the occurrence of traffic states like, e.g., synchronised traffic [Kno02b]. To achieve all this, the rule set of the BL-TCA becomes quite complex, in comparison with some of the more standard single-cell TCA models of Section 4.3:

R0: *determine stochastic noise*

$$\begin{cases} b_{i+1}(t-1) = 1 \quad \wedge \quad g_{\mathrm{t}_i}(t-1) < t_{\mathrm{s}_i}(t-1) & \Longrightarrow \quad p(t) \leftarrow p_\mathrm{b}, \\ v_i(t-1) = 0 & \Longrightarrow \quad p(t) \leftarrow p_0, \\ \text{else} & \Longrightarrow \quad p(t) \leftarrow p_\mathrm{d}, \\ b_i(t) \leftarrow 0, \end{cases} \tag{4.67}$$

R1: *acceleration*

$$\begin{aligned} &(b_i(t-1) = 0 \quad \wedge \quad b_{i+1}(t-1) = 0) \quad \vee \quad g_{\mathrm{t}_i}(t) \geq t_{\mathrm{s}_i}(t) \\ &\quad \Longrightarrow v_i(t) \leftarrow \min\{v_i(t+1), v_\mathrm{max}\}, \end{aligned} \tag{4.68}$$

R2a: *determine effective space gap*

$$\begin{aligned} &g^*_{\mathrm{s}_i}(t) \leftarrow \\ &\quad g_{\mathrm{s}_i}(t-1) + \\ &\quad \max\{\underbrace{\min\{v_{i+1}(t-1), g_{\mathrm{s}_{i+1}}(t-1)\}}_{\text{anticipated speed of leading vehicle}} - g_{\mathrm{s}_\mathrm{security}}, 0\}, \end{aligned} \tag{4.69}$$

R2b: *braking*

$$\begin{aligned} &v_i(t) \leftarrow \min\{v_i(t), g^*_{\mathrm{s}_i}(t)\}, \\ &v_i(t) < v_i(t-1) \\ &\quad \Longrightarrow b_i(t) \leftarrow 1, \end{aligned} \tag{4.70}$$

R3: *randomisation*

$$\begin{aligned} &\xi(t) < p(t) \Longrightarrow \\ &\quad p(t) = p_\mathrm{b} \quad \wedge \quad v_i(t) = v_i(t-1) + 1 \Longrightarrow b_i(t) \leftarrow 1, \\ &\quad v_i(t) \leftarrow \max\{0, v_i(t) - 1\}, \end{aligned} \tag{4.71}$$

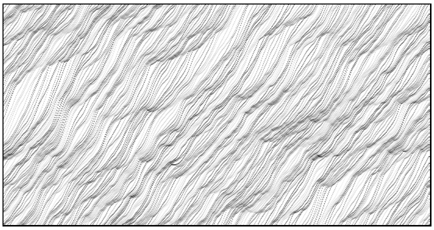
R4: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t), \tag{4.72}$$

where $b_i(t)$ denotes the state (0 or 1) of the brake light of the i^{th} vehicle at time step t, $g_{\mathrm{t}_i} = g_{\mathrm{s}_i}/v_i$ and $t_{\mathrm{s}_i} = \min\{v_i, h\}$ with h the interaction range of the brake light. As such, g_{t_i} is the time to reach the leading vehicle, which gets compared with an *interaction horizon* t_{s_i} that depends on the speed v_i and is constrained by h. If the leading vehicle is far away, its brake light should not influence the following vehicle. Furthermore, rule R0 also takes into account that drivers are more alert when they are travelling at high speeds. The slowdown probability p in rule R0, equation (4.67), corresponds to either the *braking probability* p_b, the slow-to-start probability p_0, or the classical slowdown probability p_d for decelerations. Finally, $g^*_{s_i}(t)$ in rules R2a and R2b, equations (4.69) and (4.70), respectively, denotes the *effective space gap*, based on the *anticipated speed* of the leading vehicle and taking into account a *security constraint* $g_{s_{security}}$. Just as the previous TCA models, the BL-TCA is stochastic, in that it introduces randomisation by means of rule R3, equation (4.71), with $\xi(t) \in [0,1[$ a random number drawn from a uniform distribution. If a vehicle was in the process of braking due to the previous rules, then its brake light b_i is turned on. Note that Knospe also extended the BL-TCA with rules that allow asymmetric lane changing on a two-lane road (unidirectional), incorporating a right-lane preference as well as an overtaking prohibition on the right lane. As such, the model correctly reflects the density inversion phenomenon (see also Section 4.5.1) [Kno02b; Kno02a].

In the remainder of this discussion, we set $p_b = 0.94$, $p_0 = 0.5$, $p_d = 0.1$, $h = 6$ time steps, $g_{s_{security}} = 7$ cells, $v_{max} = 20$ cells/time step, with a vehicle length of $l = 5$ cells, $\Delta T = 1$ s, and $\Delta X = 1.5$ m [Kno00; Kno04]. With respect to the calibration of the BL-TCA model's parameters, Knospe et al. provide a nice overview, giving intuitive analogies for each of these parameters (e.g., p_0 is associated with the speed of the backward propagating waves) [Kno04].

In Figure 4.35, we have given two time-space diagrams of the BL-TCA for global densities $k = 0.25$ and $k = 0.40$ vehicles/cell. As can be seen in the time-space diagram in the left part, the anticipation and synchronisation phenomena lead to forward propagating density waves, where vehicles carry the density downstream. Going to higher densities, we can see stable jams, indicative of the wide-moving jam phase (see Section 2.5.4).



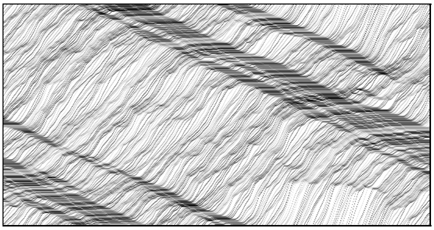


Figure 4.35: Typical time-space diagrams of the BL-TCA model (refer to the text for the used parameter values). The shown closed-loop lattices each contain 300 $\times 5 = 1500$ cells, with a visible period of 580 time steps. The global density k was set to 0.25 vehicles/cell (*left*) and 0.40 vehicles/cell (*right*). The visible forward propagating density waves are a result of the anticipation and synchronisation phenomena. At higher densities, stable jams occur, indicative of the wide-moving jam phase.

Looking at the $(k,\overline{v}_s)$ and (k,q) diagrams in Figure 4.36, we can use the local measurements to discriminate between the free-flow ($\circ$), synchronised-flow ($\cdot$), and jammed regimes ($\star$). The synchronised regime is visible as a wide scatter in the data points, having various speeds but relatively high flows. The data points in the wide-moving jam correspond to Kerner's line J in Figure 2.12 in Section 2.5.4.1. The use of a finer discretisation can lead to metastable states (see Section 4.4.1), but as Knospe et al.

note, the slow-to-start behaviour in rule R0, equation (4.67), is necessary in order to produce the correct speed of the backward propagating wave, as a result of a reduced outflow from a jam [Kno04].

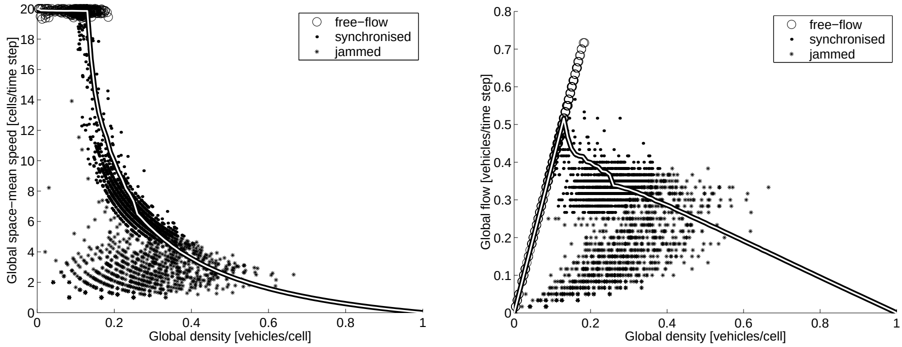


Figure 4.36: The (k,$\overline{v}_{\mathrm{s}}$) (*left*) and ($k$,$q$) (*right*) diagrams for the BL-TCA model, obtained by local and global measurements. The local measurements discriminate between the free-flow (○), synchronised-flow (·), and jammed regimes (⋆). The synchronised regime is visible as a wide scatter in the data points, having various speeds but relatively high flows. The data points in the wide-moving jam correspond to Kerner's line J in Figure 2.12 in Section 2.5.4.1.

Finally, Figure 4.37 depicts the histograms of the distributions of the space and time gaps in the left and right parts, respectively. In contrast to the HS-TCA, there are no more clusters for the space gap (see left part of Figure 4.34), but rather a smooth region of probability mass: as the global density of the system increases, the average space gap diminishes continuously and monotonically. The observations for the distributions of the time gaps correspond to those encountered in literature [Kno00; Kno04]: from the right part of Figure 4.37, we can see a wide range of probability mass at low densities (free-flow traffic), corresponding to a wide distribution of time gaps. At intermediate densities (synchronised flow), the distribution tends to peak, leading to a small dense cluster at approximately $k = 0.15$ vehicles/cell, with a median time gap of 1 time step. Finally, at higher densities (jammed traffic), the distribution of the time gaps gets more peaked, as is illustrated by the narrowing of the grey region of probability mass.

4.4.2.3 The model of Kerner, Klenov, and Wolf (KKW-TCA)

Based upon the BL-TCA of Knospe et al., Kerner, Klenov, and Wolf (KKW) refined this approach by extending it. As already mentioned in Section 2.5.4.4, their work resulted in a family of models that incorporate the notion of a *synchronisation distance* for individual vehicles [Ker03]. Derived from this model class, Kerner et al. proposed discretised versions in the form of traffic cellular automata models. In this dissertation, we consider the KKW-1 TCA model, of which the complex rule set is as follows [Ker02]:

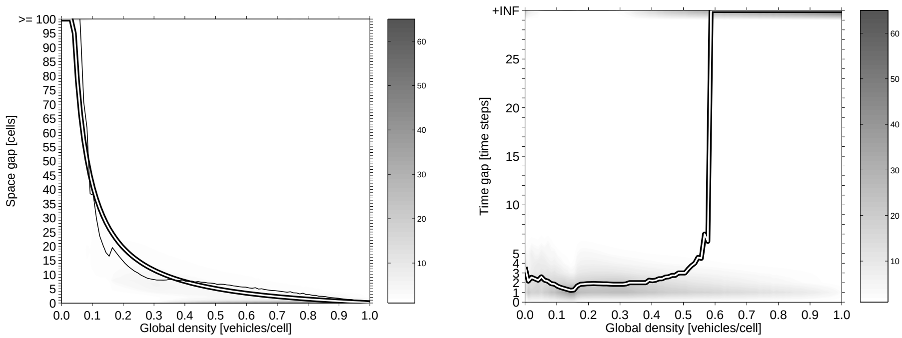


Figure 4.37: Histograms of the distributions of the vehicles' space gaps g_s (*left*) and time gaps g_t (*right*), as a function of the global density k in the BL-TCA model. The thick solid lines denote the mean space gap and median time gap, whereas the thin solid line shows the former's standard deviation. The grey regions denote the probability densities.

R1a: *determine synchronisation distance*

$$D_i(t) \leftarrow D_0 + D_1 v_i(t-1), \tag{4.73}$$

R1b: *determine acceleration and deceleration*

$$\begin{cases} v_i(t-1) < v_{i+1}(t-1) \Longrightarrow \Delta_{\text{acc}_i}(t) \leftarrow a, \\ v_i(t-1) = v_{i+1}(t-1) \Longrightarrow \Delta_{\text{acc}_i}(t) \leftarrow 0, \\ v_i(t-1) > v_{i+1}(t-1) \Longrightarrow \Delta_{\text{acc}_i}(t) \leftarrow -b, \end{cases} \tag{4.74}$$

R1c: *determine desired speed*

$$\begin{cases} g_{\text{s}_i}(t-1) > (D_i(t) - l_i) \Longrightarrow v_{\text{des}_i}(t) \leftarrow v_i(t-1) + a, \\ g_{\text{s}_i}(t-1) \leq (D_i(t) - l_i) \Longrightarrow v_{\text{des}_i}(t) \leftarrow v_i(t-1) + \Delta_{\text{acc}_i}(t), \end{cases} \tag{4.75}$$

R1d: *determine deterministic speed*

$$v_i(t) \leftarrow \max\{0, \min\{v_{\max}, g_{\text{s}_i}(t), v_{\text{des}_i}(t)\}\}, \tag{4.76}$$

R2a: *determine acceleration probability*

$$\begin{cases} v_i(t) < v_{\text{p}} \Longrightarrow p_{\text{a}}(t) \leftarrow p_{\text{a}_1}, \\ v_i(t) \geq v_{\text{p}} \Longrightarrow p_{\text{a}}(t) \leftarrow p_{\text{a}_2}, \end{cases} \tag{4.77}$$

R2b: *determine braking probability*

$$\begin{cases} v_i(t) = 0 \Longrightarrow p_{\text{b}}(t) \leftarrow p_0, \\ v_i(t) > 0 \Longrightarrow p_{\text{b}}(t) \leftarrow p_{\text{d}}, \end{cases} \tag{4.78}$$

R2c: *determine stochastic noise*

$$\begin{cases} \xi(t) < p_\text{a}(t) & \Longrightarrow \quad \eta_i(t) \leftarrow a, \\ p_\text{a}(t) \leq \xi(t) < p_\text{a}(t) + p_\text{b}(t) & \Longrightarrow \quad \eta_i(t) \leftarrow -b, \\ \xi(t) \geq p_\text{a}(t) + p_\text{b}(t) & \Longrightarrow \quad \eta_i(t) \leftarrow 0, \end{cases} \tag{4.79}$$

R2d: *determine stochastic speed*

$$v_i(t) \leftarrow \max\{0, \min\{v_\text{max}, v_i(t) + \eta_i(t), v_i(t) + a\}\}, \tag{4.80}$$

R3: *vehicle movement*

$$x_i(t) \leftarrow x_i(t-1) + v_i(t). \tag{4.81}$$

As can be seen from this overview, the KKW-TCA model's rule set is mainly composed of a *deterministic* part (rules R1a – R1d) and a *stochastic* part (rules R2a – R2d). In the deterministic part, the synchronisation distance D_i is computed first with rule R1a, which uses a linear function (other forms, e.g., quadratic functions, are also possible). The parameters D_0 and D_1 need to be estimated. Rule R1c determines the desired speed v_{des_i}: the first part of the rule allows the vehicle to accelerate, whereas the second part of the rule uses an acceleration Δ_{acc_i} defined by rule R1b (a and b are parameters denoting the acceleration, and respectively braking, capabilities). As such, a vehicle will tend to adapt its speed to that of its direct frontal leader, whenever the vehicle is within a zone of interaction (i.e., the synchronisation distance). The deterministic speed is then computed by means of rule R1d, which takes into account the maximum speed v_max, the space gap g_{s_i} to avoid a collision, and the previously computed desired speed of rule R1c.

In the stochastic part for computing the speed, a randomisation is introduced in rule R2d by means of a stochastic acceleration η_i. The values of η_i are obtained in rule R2c with probability p_a for accelerating, and probability p_b for braking. The former is dependent on the vehicles computed deterministic speed and the parameters v_p, p_{a_1}, and p_{a_2} with $p_{\text{a}_1} > p_{\text{a}_2}$ and $p_{\text{a}_1} + p_{\text{a}_2} \leq 1$. The latter, p_b is dependent on the vehicles computed deterministic speed and the slowdown probability p_d and the slow-to-start probability p_0 with $p_0 > p_\text{d}$.

In the remainder of this discussion, we set $D_0 = 60$, $D_1 = 2.55$, $a = b = 1$, $v_\text{p} = 28$, $p_{\text{a}_1} = 0.2$, $p_{\text{a}_2} = 0.052$, $p_0 = 0.425$, $p_\text{d} = 0.04$, $v_\text{max} = 60$ cells/time step, with a vehicle length of $l = 15$ cells, $\Delta T = 1$ s, and $\Delta X = 0.5$ m [Kno04].

Considering the KKW-TCA models' time-space diagrams in Figure 4.38, we can see that, in contrast to the BL-TCA (see Section 4.4.2.2), there are less spontaneous formations of small traffic jams. The forward propagating density waves in Figure 4.35 are absent in the KKW-TCA model. However, the two models show good correspondence with respect to the speed of the backward propagating waves.

Similar as in the BL-TCA model's effective space gap $g^*_{\text{s}_i}(t)$, the synchronisation distance D is responsible for producing the typical two-dimensional scatter in the $(k,\overline{v}_\text{s})$

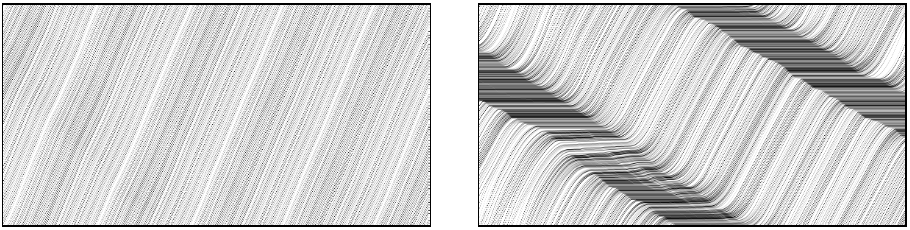


Figure 4.38: Typical time-space diagrams of the KKW-TCA model (refer to the text for the used parameter values). The shown closed-loop lattices each contain 300 $\times$15 $=$ 4500 cells, with a visible period of 580 time steps. The global density k was set to 0.25 vehicles/cell (*left*) and 0.40 vehicles/cell (*right*). Note the stable flow of vehicles surrounding the dense and compact superjams.

and (k,q) diagrams in Figure 4.39. When a driver who is within the synchronisation distance adapts the vehicle's speed, the only factors taken into account are the current speed of the direct frontal leader and a safety criterion (in the form of the current space gap); it is this effect that produces the scatter in the data, because the exact specification of this speed is absent. In both diagrams of Figure 4.39, the local measurements discriminate between the free-flow ($\circ$), synchronised-flow ($\cdot$), and jammed regimes ($\star$). One of the major differences between these two models, is that the flow in the synchronised regime is almost a factor two larger for the KKW-TCA than the BL-TCA. The KKW-TCA also experiences a capacity drop similar as in the BL-TCA, but also undergoes an abrupt transition when going from the synchronised-flow to the wide-moving jam regime around a global density of some 0.4 vehicles/cell (see the left part of Figure 4.39). Because the model is built around the assumption that vehicles tend to approximate the behaviour of their direct leader within a certain synchronisation distance, the resulting traffic regimes correspond well to Kerner's empirical observations of Section 2.5.4.1 (note that Nishimura et al. also seemed to achieve an indication of these regimes, by using the classical STCA model of Section 4.3.2.1, but in a multi-segment setup [Nis05]).

In Figure 4.40, we have depicted the histograms of the distributions of the space and time gaps in the left and right parts, respectively. The distributions are similar to those of the BL-TCA, but there are some important differences. With respect to the space gaps in the left part of Figure 4.40, there is a high variance in the jammed regime, due to the fact that there are vehicles in free-flow traffic, as well as inside the wide-moving jams (although most of the probability mass is assigned to the zero space gap inside the dense jams). Considering the time gaps in the right part of Figure 4.40, we can see that they always form a tight cluster around the median of the distribution, indicating very narrow distributions with an pronounced peak. This is completely different behaviour than in the BL-TCA model (see the right part of Figure 4.40). The main reason is probably due to the lack of an anticipation effect in the KKW-TCA model. Even more severe, is the fact that the KKW-TCA model, despite its elaborate construction based on a synchronisation distance, completely fails to describe the microscopic structure of

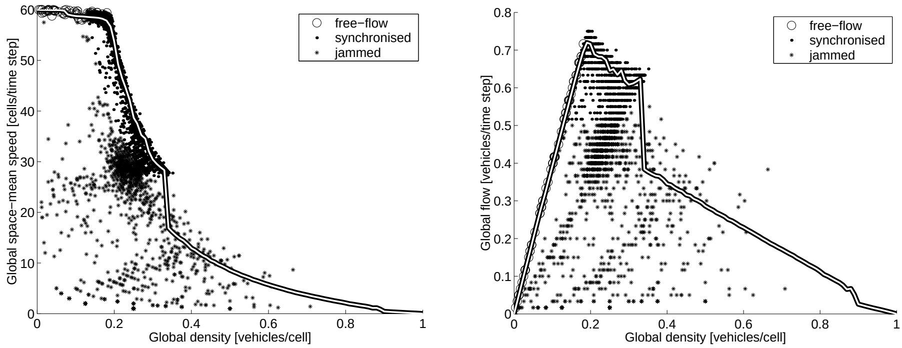


Figure 4.39: The $(k,\overline{v}_s)$ (*left*) and (k,q) (*right*) diagrams for the KKW-TCA model, obtained by local and global measurements. The local measurements discriminate between the free-flow ($\circ$), synchronised-flow ($\cdot$), and jammed regimes ($\star$). The synchronised regime is visible as a wide scatter in the data points, having various speeds but flows comparable to the capacity flow.

motorway traffic. The BL-TCA model however succeeds in having a good fit on both macroscopic and macroscopic scales, as stated according to Knospe et al. [Kno02a; Kno04].

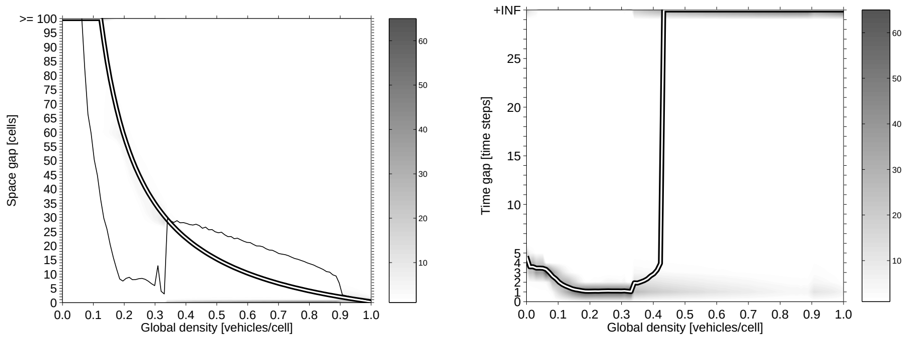


Figure 4.40: Histograms of the distributions of the vehicles' space gaps g_s (*left*) and time gaps g_t (*right*), as a function of the global density k in the KKW-TCA model. The thick solid lines denote the mean space gap and median time gap, whereas the thin solid line shows the former's standard deviation. The grey regions denote the probability densities.

4.5 Multi-lane traffic, city traffic, and analytical results

In this final section on traffic cellular automata models, we take a look at some other aspects related to TCA models. We first discuss some properties and methodologies for modelling multi-lane traffic in the context of a cellular automaton, after which we briefly consider several approaches for dealing with city traffic. The final part of the section concludes with an overview of different analytical treatments of TCA models.

4.5.1 Multi-lane traffic

In this section, we briefly discuss some properties and methodologies for modelling multi-lane traffic in the context of a cellular automaton. To this end, we illustrate the types of lane changes that are possible, then discuss the general setup for a lane-changing model. We conclude with a short overview on the implementation of lane-change rules and explain the phenomenon of ping-pong traffic, an artifact introduced by an inferior implementation.

4.5.1.1 Types of lane changes

As has already been mentioned in Section 3.2.3.1, there are commonly two types of lane changes identified: *mandatory lane changes* (MLC) and *discretionary lane changes* (DLC). In the former case, a vehicle is obliged to execute a lane change, e.g., because it needs to exit the motorway at an off-ramp, or because the vehicle is by law obliged to drive in the right shoulder lane. In the latter case, a vehicle changes a lane at its own discretion, e.g., when approaching and overtaking a slow-moving leading vehicle.

With respect to the rules for lane changing, there are also two approaches: *symmetric* and *asymmetric*. In the US, the symmetric approach is more applicable: this is embodied by the fact that motorways have a large number of lanes (i.e., more than three), with vehicles driving at lower speeds (e.g., 60 miles/hour, corresponding to some 100 kilometres/hour), effectively using all lanes more homogeneously. Such a system is typically called *"keep-your-lane"*, as frequent lane changes are discouraged. In contrast to this, people in most European countries are obliged by law to drive on the outer right shoulder lane whenever possible. Motorways have fewer lanes (typically either two or three, unidirectional), operating at higher speeds of, e.g., 120 kilometres/hour. In addition, most of these countries have instituted an overtaking prohibition on the right lane, with large trucks restricted to the two most right lanes.

With respect to this latter system of asymmetric lane changes, the phenomenon of *density* or *lane inversion* plays an important role, especially on the numerous 2x2 motorways in Europe (see also the beginning of Section 4.3 for a discussion of this phenomenon). Another aspect that has a significant influence, is the change of driver behaviour, e.g., near on-ramps. Here, drivers might avoid the shoulder lane to allow traffic to enter, or because of their increased attention, they might induce a more subtle effect such as the capacity funnel (see Section 2.5.5 for more details on this phenomenon).

4.5.1.2 General setup for lane changing

Deciding on whether or not to perform a lane change, is typically split in two separate steps: first, a vehicle checks if it is *desirable* to change lanes, i.e., making the distinction between a mandatory or discretionary lane change. If a lane change is indeed desirable, then the second step proceeds to check whether or not such a lane change can be performed at all with respect to safety and collision avoidance. Thus, there is a check for *gap acceptance*.

One of the first approaches to model such lane-changing behaviour an a two-lane road in a TCA model, is due to Nagatani. His work was based on the deterministic CA-184 model (see Section 4.3.1.1) [Nag93a]. One of the artifacts of his lane-changing rules, was the existence of states in which blocks of vehicles alternated from one lane to another, without moving at all. To circumvent this problem, Nagatani randomised the lane-changing behaviour [Nag94a]. Rickert et al. later applied this lane-changing methodology, by extending the STCA model (see Section 4.3.2.1) to handle two-lane unidirectional traffic [Ric96a]. Wagner et al. later assessed the previous work of Rickert et al., concluding that it did not capture certain aspects (e.g., density inversion) of traffic flows very well [Wag97]. To this end, they built upon the previous work, adding a more specialised security constraint that takes into account the fact that vehicles should also consider the following vehicles in the target lane, thereby avoiding severe disruptions. As a final comment, they state that the lane-changing rules in a TCA model typically do not provide a realistic microscopic model, but they rather lead to a good correspondence with respect to observed macroscopic features (e.g., the frequency of lane changes).

In order to address the correct reproduction of the density inversion phenomenon, Nagel et al. artificially introduced a *slack parameter*, capturing the inclination of a driver to change back to the right lane. They furthermore also provided an extensive classification of some 10 lane changing rules and criteria encountered in literature [Nag98d]. Another excellent overview of multi-lane traffic is given by Chowdhury et al. [Cho00].

As all the previous work dealt with unidirectional roads, it seems logical to consider *bidirectional traffic*, i.e., traffic with adjacent but opposing lanes. Simon and Gutowitz were among the first to consider a TCA model of such traffic, with vehicles driving on two lanes [Sim98a]. Central to their approach, is the notion of a *local density* that each driver must assess before attempting to complete an overtaking manoeuvre. When a driver encounters a slower moving vehicle, a check is made whether or not there is enough space *in front* of this leading vehicle (this is the local density). If the check is positive, then a lane change can be performed (under the condition of course that there is a safe gap in the opposing lane). With this scheme in mind, high density traffic thus excludes such overtaking manoeuvres, due to the fact that the local density is too low to complete them.

Note that some authors, e.g., Gundaliya et al. [Gun04], Mallikarjuna and Ramachandra Rao [Mal05], use a peculiar variant of a multi-lane setup. Their models have essentially a multi-cell structure, but now the multi-cell concept is extended in the lateral direction. So cells not only get smaller, but also 'thinner', allowing *variable-width vehicles*, e.g., motor cycles that can more easily pass other vehicles in the same lane. In our opinion, this leads to unnecessary complexity, giving little benefits. In fact, we believe that such a scheme directly opposes the idea behind a CA model, as explained at the introduction of this chapter. We strongly feel that heterogeneity in a TCA model should *only* be incorporated by means of different lengths, maximum speeds, acceleration characteristics, anticipation levels, and stochastic noise for distinct classes of vehicles and/or drivers. Any other approach would be better off with a continuous microscopic model.

4.5.1.3 Implementation of lane-changing rules and the phenomenon of ping-pong traffic

The basic implementation of a lane-changing model in a TCA setting, leads to two substeps that are consecutively executed at each time step of the CA:

- first, the lane-changing model is executed, exchanging vehicles between *laterally* adjacent lanes,
- then, all vehicles are moved forward (i.e., *longitudinal*) by applying the car-following part of the TCA model's rules.

One immediate result from this approach, is that a lane change in a TCA model is completed within one time step (i.e., ΔT). This is in contrast to real-life traffic, where lane changes have a duration of several seconds [Nag98d].

For more than two lanes, care must be taken to avoid so-called *scheduling conflicts* during the first substep. Consider for example three lanes, with two vehicles driving in the outer left, respectively outer right, lane at the same longitudinal position. If the cell in the middle lane is empty, then the vehicles may decide to move to this location, resulting in a lateral collision. In order to compensate this, one possibility is to choose a vehicle at random (or by preference), thereby allowing it to perform its requested lane change. Another possibility is to perform left-to-right lane changes in even time steps, and right-to-left lane changes in odd time steps.

As hinted earlier, the 'correctness' of a lane-change model should be judged on the basis of certain macroscopic observations. Examples of these are the frequency of lane changes with respect to different densities, the capacity flows for all lanes separately and combined, the critical density at which a breakdown occurs in each of the lanes, ... Good indicators can be found in the many small fluctuations typically exhibited by multi-lane TCA models, instead of the large jams in single-lane traffic. Traffic flows get more fluid if vehicles are allowed to pass moving bottlenecks [Wag97; Nag98d]. However, under certain conditions, Helbing and Huberman have shown the existence of coherent states, where vehicles' speeds are synchronised across adjacent lanes. For heterogeneous traffic flows, this can lead to a moving 'solid block' of vehicles [Hel98b].

When implementing lane-change rules in a TCA model, care must however be taken that the implementation does not introduce any unrealistic artifacts. A prominent example of this, plaguing many TCA models, is a phenomenon called *ping-pong traffic*. Nagatani was among the first to observe this peculiar behaviour of vehicles in traffic flows (see Section 4.5.1.2). In ping-pong traffic, vehicles typically alternate between lanes during successive time steps. As explained earlier, one way to resolve this behaviour is by randomising the lane-change decision, thereby quickly destroying any such artificial patterns [Nag94a; Ric96a].

4.5.2 City traffic and intersection modelling

When modelling city traffic, essentially two approaches can be followed: either the entire road network is considered as a two-dimensional lattice (i.e., a *grid*), or each road in the network is a single longitudinal lattice (single- or multi-lane) with explicitly modelled intersections. The former was historically used in the context of phase transitions in a CA, whereas the latter is more applicable to describe real-life traffic flows in populated cities.

In this section, we illustrate both approaches, starting with a classical grid layout as embodied by the Biham-Middleton-Levin (BML) and Chowdhury-Schadschneider (ChSch) TCA models, after which we briefly comment on explicit descriptions of intersections in TCA models.

4.5.2.1 Grid traffic

The first model of 'city traffic' was proposed by Biham, Middleton, and Levine (BML). It was developed around the same time Nagel and Schreckenberg presented their STCA (see Section 4.3.2.1). The BML-TCA, is a two-dimensional model that describes traffic on a square grid in a toroidal setup (i.e., opposing sides are identified), with vehicles distributed randomly over the lattice [Bih92]. The model is in fact a very simplistic model, in that it assumes that all vehicles either move from the south to the north direction, or from the west to the east. Each cell of the lattice is assumed to contain a traffic light, in the sense that all west-east vehicles try to move during even time steps, and all south-north vehicles during odd time steps (thus $v_{\max} = 1$ cell/time step for all vehicles). The BML-TCA constitutes a fully deterministic model, where the only randomness is introduced through the initial conditions. Note that its one-dimensional version corresponds to the CA-184 and the TASEP (see Sections 4.3.1.1 and 4.3.2.4).

Depending on the global density of vehicles in the lattice, the model results in two distinct traffic regimes, with a *sharp first-order phase transition* between them. The first regime, i.e., free-flow traffic, corresponds to a state with alternate moving vehicles (i.e., west-east and south-north moving); an example is depicted in the left part of Figure 4.41. In the congested regime, a self-organised global cluster emerges, completely composed of blocked vehicles (see, e.g., the middle part of Figure 4.41). When the phase transition between both regimes occurs, the space-mean speed changes abruptly from one to zero cells/time step [Bih92; Ang05]. Cuesta et al. also showed that the phase transition persists even when vehicles are allowed to turn [Cue93]. Fukui and Ishibashi studied the repercussions of a local disruption in the lattice (e.g., a crashed vehicle that remains stopped for an eternal period), and found that it provides the seed of a growing global cluster [Fuk93]. Biham et al. also considered a less restrictive version of the above model, in which now all vehicles try to move at each time step. In case of conflicts between a west-east and a south-north vehicle, one of them is chosen at random. Another variation considers also opposing traffic, which can lead to *gridlocked* situations where no vehicles are able to move at all. A generalisation of the BML-TCA, was provided by Freund and Poschel who consider a similar setup, but now with traffic moving in all four directions [Fre95]. With respect to the critical densities at which the previously mentioned phase transitions occur, Shi was able to obtain analytical expressions for them [Shi99]. Finally, the recent work of D'Souza showed that, contrary to the general belief in a sharp first-order phase transition, the BML-TCA can exhibit intermediate stable phases that contain both free-flow and congested regimes [D'S05].

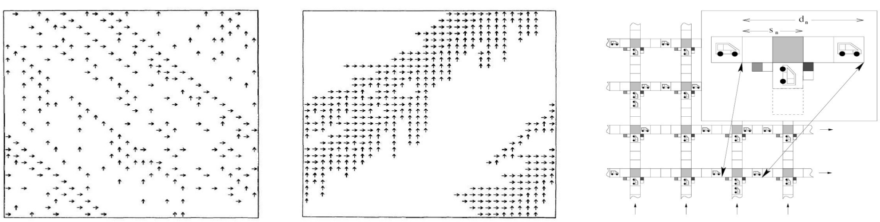


Figure 4.41: *Left:* snapshot of the spatial structure in the BML-TCA for $k = 0.25$. In this free-flow regime, all vehicles move alternatingly, with the right-oriented arrows denoting west-east travelling vehicles, and the upward-oriented arrows denoting south-north travelling vehicles. *Middle:* same setup as before, but now for $k \approx 0.4082$. In this congested regime, a global cluster emerges, completely composed of blocked vehicles. *Right:* an overview of the ChSch-TCA, showing the street segments of finite length between the BML-TCA's original intersections. The first two images are reproduced after [Bih92], the third after [Bar03].

In the work of Chowdhury et al., a comprehensive overview is given, describing extensions to the BML-framework [Cho00]. This overview includes asymmetric distributions of the west-east and south-north vehicles, unequal maximum speeds, two-level crossings (where two vehicles can share the same cell), faulty traffic lights (here, either a west-east or south-north vehicle is chosen at random to occupy a cell, irrespective of the current time step), road blocks, line- and point-defects (i.e., a crowded 'street' of the model, corresponding to a dense horizontal or vertical row of cells), random turning of vehicles, cut-off streets (similar to a row of two-level crossings), and so forth and so on.

Chowdhury and Schadschneider later extended the BML-TCA model to incorporate randomisation effects like in the STCA model, having the result that jamming can now occur spontaneously [Cho99b]. Their model furthermore contains street segments of finite length between the cells, with vehicles driving according to the STCA's rules on these streets. The original cells in the BML-TCA model form the signallised intersections of the Chowdhury-Schadschneider model (ChSch-TCA), as can be seen in the right part of Figure 4.41. At sufficiently large densities, a transition can occur that leads to a self-organising state of completely gridlocked traffic. Barlović later provided a solution to this problem, making the model well-suited for assessing the results of different traffic light control policies in a city [Bar03].

4.5.2.2 Explicit intersection modelling

In contrast to the previous section were all traffic operations were essentially defined on a two-dimensional lattice, it is also possible to consider a complete road network, consisting of *separate links* that are connected to each other by means of *intersections*. These intersections can either be signallised, or unsignallised, turning priorities can be defined, as well as different geometrical layouts (e.g., roundabouts).

Road networks based on the above assumptions, typically combine a set of basic building blocks. As such, the network is logically decomposed in a set of *nodes* and *links*. The former denote the intersections, whereas the latter can, depending on the implementation, refer to individual lanes, a group of adjacent lanes, or even a road with two-way traffic. In general, traffic operations on motorways are primarily influenced by the behaviour of vehicles on links, i.e., their car-following and lane-changing behaviour. Conversely, traffic operations in cities and denser street networks, are primarily defined by the behaviour of vehicles at intersections, i.e., queueing delays at traffic lights, priority turns, . . . In many cases, the intersection logic is simplified, such that all decisions (conflict resolving et cetera) are taken *before* a vehicle enters the intersection [Hel04].

Several non-exhaustive examples include the work of Esser and Schreckenberg with applications to the city of Duisburg [Ess97], the work of Simon and Nagel who primarily focussed on single-lane traffic in combination with several setups for controlling traffic lights, applying their work to the city of Dallas (different links have different slowdown probabilities associated with them, thus enabling to model different street capacities) [Sim98b], the work of Diedrich et al. who consider the effects of various implementations of on- and off-ramps in the classical STCA model [Die00], and all the references on TRANSIMS, the travel behaviour in Switzerland, the region of Dallas, the city of Portland, and the city of Geneva (where all intersections are replaced by generalised roundabouts), mentioned at the end of Section 3.1.3.3.

> All these examples have in common that they are based on simple building blocks. Despite this elegance, most of them however, do not provide satisfactory information regarding the calibration and validation of their underlying models (this for example with respect to the correct observed queueing delays at intersections). A popular technique is to use *sources* and *sinks*, where vehicles are added and removed, allowing tuning of the simulator in order to agree with incoming on-line measurements. Clearly, we feel that besides a need for elaborate descriptions of the employed models, there is perhaps even a bigger need for correct information with respect to these models' fidelity and accuracy.

4.5.3 Analytical results

Because most studies based on TCA models heavily rely on numerical simulations, this creates the danger of introducing artifacts (e.g., finite-size effects) that obscure the true dynamics of the systems under consideration. Although most of these problems should resolve in the so-called *thermodynamic limit* where $K_{\mathcal{L}}, T_{\text{mp}} \rightarrow +\infty$ (i.e., a lattice with infinite length considered over an infinite time period), resorting to this approach is computationally not feasible. As a result, researchers have focussed on analytical methods. Except for the most trivial cases with a deterministic (i.e., noiseless) TCA model, these analytical methods most of the time provide approximations at best.

In this section, we illustrate several of these analytical methods encountered in literature. Our discussion focusses on the concept of a mean-field theory, after which we elaborate on some of its improvements that lead to better agreement with numerical results.

> Note that other avenues for analytical treatments of CA models, and TCA models in particular, are also explored. In this section, we will however not go into detail about them. For more information, we refer the reader to the interesting work of Fukś and Boccara [Fuk98; Fuk99; Boc00; Fuk01; Fuk04].

4.5.3.1 Mean-field theory

As mentioned in the introduction of this section, for the case of arbitrary v_{max} and $p = 0$ or $p = 1$, or for $v_{\text{max}} = 1$ cell/time step, the analytical solution of the resulting TCA model is exactly known. This solution, expressed as its (k,q) diagram, corresponds to the set of diagrams as depicted in Figure 4.11 (see Section 4.3.1.2) for the DFI-TCA.

The problem is to find an analytical description of how the system evolves in time through the state space, i.e., what are the occurring configurations ? The evolution of a system, can be described by what is called a *master equation*. For cellular automata, this equation is a first-order differential equation, describing the change in probability of a system's lattice to be in a certain configuration. The downside is that, in general, this master equation can not be solved exactly.

For the TASEP model (see Section 4.3.2.4) with open boundary conditions and random sequential update, the master equation can be solved exactly [Raj96; San99]. In a first step, the master equation is elegantly written in vector form, comprising a *transfer matrix* that contains the time-evolution of the probabilities. By assuming the *matrix-product ansatz* (MPA) formalism, the transfer matrix can be rewritten as a product of local transfer matrices, operating on sets of cells. This provides a algebra that can be solved exactly, thereby solving the TASEP analytically. Note that for the TASEP with a parallel update however, obtaining the exact solution is difficult, because no simple MPA decomposition into local matrices is possible.

In contrast to this promising result, obtaining an analytical solution becomes harder to even intractable for the STCA model (see Section 4.3.2.1) with $v_{\text{max}} > 1$ cell/time step and $0 < p < 1$. In the master equation, probabilities of cluster of cells will occur, making its solution very hard [Sch02a]. One well-known method that is suitable for dealing with many-particle systems in statistical mechanics, is the construction of a *mean-field theory* (MFT) of the model. Such a MFT can provide an approximation of the master equation; in some cases, the MFT turns out to be an exact solution.

The idea behind a MFT, is that all correlations between neighbouring cells are neglected. For TCA models, such a *site-oriented mean-field theory* (SOMF) assumes that all cluster probabilities are replaced by single cell probabilities. The MFT now replaces the effects of these individual cells with an average effect (the 'mean field'), which simplifies computations considerably. When translating the STCA's rules R1 – R3, i.e., equations (4.46) – (4.48), R1 is decoupled into separate acceleration and braking rules R1a and R1b, after which their order is changed to R1b, R3, R4, R1a. The upshot of this is that there are no stopped vehicles in the system, thereby reducing the number of possible states for a cell by one. If $v_{\text{max}} = 1$ cell/time step, then the system can be fully described by cell occupancies. Applying this SOMF theory to the STCA model, results in considerably underestimation of the flow in the system (even for the restricted case of $v_{\text{max}} = 1$ cell/time step) [Sch95; Sch99a; Sch02a]. Finally, note that Wang et al. also applied the mean-field theory to the case of two-dimensional traffic such as the BML-TCA (see Section 4.5.2.1) [Wan96].

4.5.3.2 Improving the SOMF theory

As mentioned in the previous section, setting $v_{\text{max}} = 1$ cell/time step leads to an underestimation of the flow. However, when switching from a parallel update procedure to a random sequential one, the resulting SOMF theory becomes exact ! It turns out that the reason for the underestimation, can be traced back to its neglecting of all correlations between cells (which are a consequence of the parallel update procedure). As explained in the beginning of Section 4.3.2.4, using a parallel update excludes certain Garden of Eden states. However, the SOMF theory naively includes these paradisiacal states. As a solution, these GoE states can be eliminated, resulting in a *paradisiacal mean-field theory* (pMFT). In systems with higher maximum speeds, more GoE states occur, making it difficult to derive a pMFT. Even then, the theory still remains an approximation (albeit a better one) when using a parallel update procedure [Sch98; Sch99a; Sch02a].

Taking into account short-range correlations, can be done by considering a *car-oriented mean-field theory* (COMF). Instead of dealing with cells and their occupancies, the COMF theory computes the probabilities $P_n(v)$ of finding a space gap of n cells for a vehicle driving with speed v [Sch97b]. In a sense, the COMF theory approximates the problem by neglecting the correlations between space gaps of successive vehicles [Cho00]. As such, it gives qualitatively good approximations for $p \rightarrow 0$; in all other cases, the COMF theory starts to fail, because there are also correlations between the space gaps [San99; Cho00]. Note that the COMF theory has also been applied to the BJH-TCA and VDR-TCA models (see Sections 4.3.3.2 and 4.3.3.3, respectively) [Sch97c].

Another approach to analytically solve the master equation, is to explicitly take into account the correlations between neighbouring cells, by considering *clusters* composed of n consecutive cells [Sch99a; Sch02a]. Such a *site-oriented cluster-theoretic approach* proves to perform better than the COMF theory from the previous section [Sch95]. The improvement of the approximation is even better when considering larger clusters; it is exact for $n \rightarrow +\infty$ [Sch97a; San99; Cho00].

4.6 Conclusions

This chapter introduced all the necessary material for understanding traffic cellular automata (TCA) models, which are a class of computationally efficient microscopic traffic flow models. TCA models arise from the physics discipline of statistical mechanics, having the goal of reproducing the correct macroscopic behaviour based on a minimal description of microscopic interactions.

We began with an overview of cellular automata (CA) models, their background and physical setup. Applying this technique to the modelling of traffic flows, we discretise a road into a number of small cells (a procedure called coarse graining), having a width of, e.g., $\Delta X = 7.5$ m. Time is also discretised into units of approximately $\Delta T = 1$ s. After introducing the mathematical notations, we showed how to perform measurements on a TCA model's lattice of cells, and how to convert these quantities into real-world units and vice versa.

Subsequently, we gave an extensive account of the behavioural aspects of several TCA models encountered in literature. Already, several reviews of TCA models exist, but none of them consider all the models exclusively from the behavioural point of view. In this respect, our overview fills this void, as it focusses on the behaviour of the TCA models, by means of time-space diagrams, (k,q) diagrams and the like, and histograms showing the distributions of vehicles' speeds, space, and time gaps. In this chapter, we have distinguished between single-cell and multi-cell models, whereby in the latter vehicles are allowed to span a number of consecutive cells. We concluded with a concise overview of TCA models in a multi-lane setting, and some of the TCA models used to describe city traffic as a two-dimensional grid of cells, or as a road network with explicitly modelled intersections. The final part of the chapter illustrated some of the more common analytical approximations to single-cell TCA models.

Considering the state-of-the-art in using TCA models, our analysis indicates that the field has evolved rapidly over the last decade. Starting from initial attempts based on rather crude models, the past few years have seen an increase in the computational complexity as well as the available computational power. More complex models are developed, of which we believe the brake-light TCA model of Section 4.4.2.2 is the most promising: it is able to faithfully reproduce the correct real-life empirical observations, and quite some work has been done at calibrating the model, see, e.g., the recent work of Knospe et al. [Kno04], and the work of Chrobok et al. in their development of the *On-Line SIMulator* (OLSIM) [Chr04; Pot04]. To conclude, we note an evolving trend of using these TCA models as the physical models underlying multi-agent systems, in part describing the behaviour of individual people in large-scale road networks as explained in Section 3.1.3.3.

In the next chapter, we present a relation between the stochastic STCA model of Section 4.3.2.1 and the macroscopic first-order LWR model of Section 3.2.1.2. Instead of considering the classical angle of using, e.g., a mean-field theory, we take a different approach at studying their relation. By considering a TCA model as a particle-based discretisation scheme for macroscopic traffic flow models, we can address the common structure between both models. This allows us to provide a means for explicitly incorporating the STCA's stochasticity into the LWR model.

Chapter 5

Relating the dynamics of the STCA to the LWR model

Considering the existing relations between the stochastic TCA (STCA) model of Nagel and Schreckenberg (see Section 4.3.2.1) [Nag92b; Nag95a] and the macroscopic first-order LWR model of Lighthill, Whitham, and Richards (see Section 3.2.1.2) [Lig55; Ric56], from traffic flow theory, there are already numerous links between both modelling approaches. An example is the so-called totally asymmetric simple exclusion process (TASEP) [Der92], which corresponds to the LWR model with a noisy and diffusive conservation law if a random sequential update is assumed (see Section 4.3.2.4) [Nag95a; Nag96]. Deriving an analytical treatment of the STCA proves to be quite hard to even intractable. One standard way for dealing with this, is an approximation by a so-called mean field theory (MFT), and its successive refinements, such as the site-oriented and car-oriented mean-field theories (SOMF and COMF), as well as the recently developed site-oriented cluster-theoretic approach (see Section 4.5.3) [Sch95; Sch97b; Sch98; San99; Sch99a; Cho00; Sch02a]. Except for the most trivial cases with a deterministic (i.e., noiseless) TCA model, these analytical methods most of the time provide approximations at best. In summary, we can say that there already exist several methods for bridging both the microscopic STCA and the macroscopic LWR model[1]*.*

In this chapter, we reconsider the STCA and LWR models, but we take a different approach at studying their relation: we consider a TCA model as a particle-based discretisation scheme for macroscopic traffic flow models. It is from this latter point of view that our work addresses the common structure between both models. Our main goal is therefore to provide a means for implicitly incorporating the STCA's stochasticity into the LWR model, which is in fact deterministic in nature [Mae03a].

[1] Note that we do not consider the hybrid models of Section 3.2.5, as we are only interested in direct analogies between both microscopic and macroscopic models.

> *Note that we use the term* implicit *to denote the fact that the STCA's stochasticity is not introduced in the equations by means of* explicit *noise terms. Rather, our methodology implies that the stochasticity is introduced through the shape of the LWR's fundamental diagram. At the heart of our procedure, lies a transformation of the STCA's rules into a deterministic fundamental diagram that is specified to the LWR model.*

In the remainder of this chapter, we first consider a methodology for implicitly incorporating the STCA's stochasticity into the LWR's triangular fundamental diagram. We then apply this technique to a small case study, which points us to some discrepancies between both modelling approaches. We then highlight some of the resulting artifacts, after which we investigate the main reason for the difference in behaviour. Continuing this train of thought, we present an alternate derivation of the fundamental diagram. The chapter concludes with a summary of our findings.

5.1 Implicitly incorporating the STCA's stochasticity

As mentioned in the introduction, we reconsider the STCA and LWR models, taking a different approach at studying their relation. Our main goal is to provide a means for implicitly incorporating the STCA's stochasticity into the LWR model. To this end, we provide a practical methodology for specifying the fundamental diagram to the LWR model. Assuming that a *stationarity condition* holds on the STCA's rules, we incorporate the STCA's stochasticity directly into the LWR's fundamental diagram.

Relating both the STCA and the LWR models is now done using a simple two-step approach, in which we first rewrite the STCA's rules into a single rule, leading to a set of *linear inequalities*. These constraints can be considered as a $\overline{v}_{\mathrm{s_e}}(\overline{h}_{\mathrm{s}})$ fundamental diagram (see, e.g., Figure 2.7). This latter diagram can then be converted into an equivalent triangular flow versus density $q_{\mathrm{e}}(k)$ fundamental diagram.

5.1.1 Rewriting the STCA's rule set

Considering a vehicle's average speed, the STCA's rules R1 and R2, equations (4.46) and (4.47) respectively, state that a vehicle slows down with probability p, and that it does not slow down with probability $1 - p$. As such, they can be rewritten into the following single rule that is expressed in *continuous* speeds and space gaps:

$$\begin{aligned} v_i(t) \quad \leftarrow \quad & p \cdot \min\{v_i(t-1) \cancel{+1} \cancel{-1}, g_{\mathrm{s}_i}(t-1) - 1, v_{\max} - 1\} + \\ & (1-p) \cdot \min\{v_i(t-1) + 1, g_{\mathrm{s}_i}(t-1), v_{\max}\}, \end{aligned} \tag{5.1}$$

with $v_i(t) \leftarrow \max\{0, v_i(t)\}$. Furthermore, the following two algebraic relations always hold:

$$
\begin{aligned}
a \cdot \min\{b,c\} &= \min\{ab, ac\}, &(5.2)\\
\min\{a,b\} + \min\{c,d\} &= \min\{a+c, a+d, b+c, b+d\}. &(5.3)
\end{aligned}
$$

Applying relation (5.2) to our rule (5.1), yields the following result:

$$
\begin{aligned}
v_i(t) \leftarrow\ & \min\{pv_i(t-1), p(g_{s_i}(t-1)-1), p(v_{\max}-1)\} + \\
& \min\{(1-p)(v_i(t-1)+1), (1-p)g_{s_i}(t-1), (1-p)v_{\max}\}. \qquad (5.4)
\end{aligned}
$$

Using relation (5.3) to this result, allows us to obtain a formulation with a single minimum-operator:

$$
\begin{aligned}
v_i(t) \leftarrow \min\{\quad & pv_i(t-1) + (1-p)(v_i(t-1)+1),\\
& pv_i(t-1) + (1-p)g_{s_i}(t-1),\\
& pv_i(t-1) + (1-p)v_{\max},\\
& p(g_{s_i}(t-1)-1) + (1-p)(v_i(t-1)+1),\\
& p(g_{s_i}(t-1)-1) + (1-p)g_{s_i}(t-1),\\
& p(g_{s_i}(t-1)-1) + (1-p)v_{\max},\\
& p(v_{\max}-1) + (1-p)(v_i(t-1)+1),\\
& p(v_{\max}-1) + (1-p)g_{s_i}(t-1),\\
& p(v_{\max}-1) + (1-p)v_{\max}\}. \qquad (5.5)
\end{aligned}
$$

Expanding all the terms between parentheses gives the following result:

$$
\begin{aligned}
v_i(t) \leftarrow \min\{\quad & \cancel{pv_i(t-1)} + v_i(t-1) + 1 - \cancel{pv_i(t-1)} - p,\\
& pv_i(t-1) + g_{s_i}(t-1) - pg_{s_i}(t-1),\\
& pv_i(t-1) + v_{\max} - pv_{\max},\\
& pg_{s_i}(t-1) - p + v_i(t-1) + 1 - pv_i(t-1) - p,\\
& \cancel{pg_{s_i}(t-1)} - p + g_{s_i}(t-1) - \cancel{pg_{s_i}(t-1)},\\
& pg_{s_i}(t-1) - p + v_{\max} - pv_{\max},\\
& pv_{\max} - p + v_i(t-1) + 1 - pv_i(t-1) - p,\\
& pv_{\max} - p + g_{s_i}(t-1) - pg_{s_i}(t-1),\\
& \cancel{pv_{\max}} - p + v_{\max} - \cancel{pv_{\max}}\}. \qquad (5.6)
\end{aligned}
$$

And finally, regrouping for p yields:

$$
\begin{aligned}
v_i(t) \leftarrow \min\{ \quad & v_i(t-1) + 1 - p, \\
& p(v_i(t-1) - g_{s_i}(t-1)) + g_{s_i}(t-1), \\
& p(v_i(t-1) - v_{\max}) + v_{\max}, \\
& p(g_{s_i}(t-1) - v_i(t-1) - 2) + v_i(t-1) + 1, \\
& g_{s_i}(t-1) - p, \\
& p(g_{s_i}(t-1) - v_{\max} - 1) + v_{\max}, \\
& p(v_{\max} - v_i(t-1) - 2) + v_i(t-1) + 1, \\
& p(v_{\max} - g_{s_i}(t-1) - 1) + g_{s_i}(t-1), \\
& v_{\max} - p\}. \qquad (5.7)
\end{aligned}
$$

If we now assume traffic is *stationary* (see, e.g., Daganzo's description of stationary traffic in Section 2.3.4.2), then we can assert that the state of a vehicle at time t is the same as its state at time $t-1$, i.e., $v_i(t) = v_i(t-1)$ and $g_{s_i}(t) = g_{s_i}(t-1)$. As a result, equation (5.7) gets transformed into the following set of *linear inequalities* that express constraints on the relations between $v_i(t)$, $g_{s_i}(t)$, p, and $v_{\max}$:

$$
\begin{aligned}
\cancel{v_i(t)} + 1 - p \geq \cancel{v_i(t)} &\qquad \text{(C1)}, \\
p(v_i(t) - g_{s_i}(t)) + g_{s_i}(t) \geq v_i(t) &\qquad \text{(C2)}, \\
p(v_i(t) - v_{\max}) + v_{\max} \geq v_i(t) &\qquad \text{(C3)}, \\
p(g_{s_i}(t) - v_i(t) - 2) + \cancel{v_i(t)} + 1 \geq \cancel{v_i(t)} &\qquad \text{(C4)}, \\
g_{s_i}(t) - p \geq v_i(t) &\qquad \text{(C5)}, \\
p(g_{s_i}(t) - v_{\max} - 1) + v_{\max} \geq v_i(t) &\qquad \text{(C6)}, \\
p(v_{\max} - v_i(t) - 2) + \cancel{v_i(t)} + 1 \geq \cancel{v_i(t)} &\qquad \text{(C7)}, \\
p(v_{\max} - g_{s_i}(t) - 1) + g_{s_i}(t) \geq v_i(t) &\qquad \text{(C8)}, \\
v_{\max} - p \geq v_i(t) &\qquad \text{(C9)}.
\end{aligned}
$$

Let us now examine each of these nine constraints C1 – C9.

- Constraint C1 states that $1 - p \geq 0$, i.e., $p \leq 1$. This logically follows from the STCA's condition that $p \in [0, 1]$.
- Constraint C2 states that $p(v_i(t) - g_{s_i}(t)) + g_{s_i}(t) \geq v_i(t)$, i.e., $g_{s_i}(t)\cancel{(1-p)} \geq v_i(t)\cancel{(1-p)}$. This corresponds to $v_i(t) \leq g_{s_i}(t)$, which states that vehicles strive for collision-free driving.
- Constraint C3 states that $p(v_i(t) - v_{\max}) + v_{\max} \geq v_i(t)$, i.e., $v_{\max}\cancel{(1-p)} \geq v_i(t)\cancel{(1-p)}$. This corresponds to $v_i(t) \leq v_{\max}$, which logically follows from the STCA's condition that $v_i(t) \in \{0, \ldots, v_{\max}\}$.

- Constraint C4 states that $p(g_{s_i}(t) - v_i(t) - 2) + 1 \geq 0$, i.e., $v_i(t) \leq g_{s_i}(t) - 2 + \frac{1}{p}$ (for $p \neq 0$).
- Constraint C5 states that $g_{s_i}(t) - p \geq v_i(t)$, i.e., $v_i(t) \leq g_{s_i}(t) - p$, which is a more stringent constraint than C2 and C4.
- Constraint C6 states that $p(g_{s_i}(t) - v_{\text{max}} - 1) + v_{\text{max}} \geq v_i(t)$, i.e., $v_i(t) \leq v_{\text{max}}(1-p) + p(g_{s_i}(t) - 1)$.
- Constraint C7 states that $p(v_{\text{max}} - v_i(t) - 2) + 1 \geq 0$, i.e., $v_i(t) \leq v_{\text{max}} - 2 + \frac{1}{p}$ (for $p \neq 0$).
- Constraint C8 states that $p(v_{\text{max}} - g_{s_i}(t) - 1) + g_{s_i}(t) \geq v_i(t)$, i.e., $v_i(t) \leq g_{s_i}(t)(1-p) + p(v_{\text{max}} - 1)$.
- Constraint C9 states that $v_{\text{max}} - p \geq v_i(t)$, i.e., $v_i(t) \leq v_{\text{max}} - p$, which is a more stringent constraint than C3 and C7.

Taking the previous considerations into account, we can see that constraints C1, C2, and C3 are always satisfied. The remaining three pairs of similar constraints on the relations between $v_i(t)$, $g_{s_i}(t)$, p, and v_{max}, are the following: constraints C5 and C9, C4 and C7, and C6 and C8.

In order to gain insight into the more difficult constraints C6 and C8, we first rewrite them as follows:

$$\text{(C6)} \qquad v_i(t) \leq \underbrace{p}_{\text{slope}} g_{s_i}(t) + \underbrace{(1-p)v_{\text{max}} - p}_{\text{intercept}},$$

$$\text{(C8)} \qquad v_i(t) \leq \underbrace{(1-p)}_{\text{slope}} g_{s_i}(t) + \underbrace{p(v_{\text{max}} - 1)}_{\text{intercept}},$$

where we have separated the terms containing $g_{s_i}(t)$. Plotting the speed $v_i(t)$ versus the space gap $g_{s_i}(t)$ in Figure 5.1, allows us to more easily interpret the combined effects of these two constraints. On the one hand, if we continuously change $p = 0 \rightarrow 1$, then constraint C6 goes from a horizontal line at $v_i(t) = v_{\text{max}}$, to a slanted line with a slope of $+1$, intercepting the horizontal and vertical axes at $+1$ and -1, respectively. In all cases, the point at ($v_{\text{max}} + 1$,v_{max}) remains invariant. On the other hand, changing $p = 0 \rightarrow 1$ turns constraint C8 from a slanted line with a slope of $+1$, passing through the origin, into a horizontal line at $v_i(t) = v_{\text{max}} - 1$. In all cases, the point at ($v_{\text{max}} - 1$,$v_{\text{max}} - 1$) remains invariant.

5.1.2 Deriving the fundamental diagram

The next step of our approach, considers the most determining linear inequalities C5, C6, C8, and C9 as boundaries in a $\overline{v}_{s_e}(\overline{g}_s)$ fundamental diagram. As such, we note the following observations:

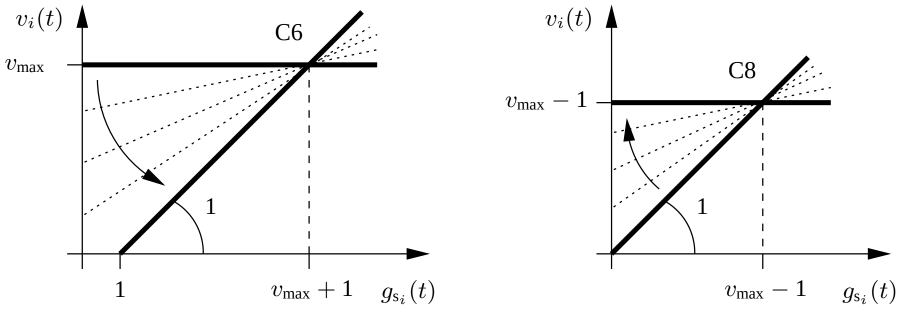


Figure 5.1: A visual representation of the constraints C6 and C8. *Left:* as $p = 0 \rightarrow 1$, C6 changes from a horizontal line at $v_i(t) = v_{\max}$, to a slanted line with a slope of $+1$, intercepting the horizontal and vertical axes at $+1$ and -1, respectively. *Right:* at the same time, constraint C8 changes from a slanted line with a slope of $+1$, passing through the origin, into a horizontal line at $v_i(t) = v_{\max} - 1$.

- **Increasing the slowdown probability p, holding $v_{\max}$ constant**:
  - The average speed $\overline{v}_{\text{ff}}$ in the free-flow regime decreases towards $v_{\max} - p$.
  - The transition point at the critical space gap g_{s_c} remains invariant.
  - The space gap g_{s_j}, corresponding to the jam density, increases.
- **Decreasing the maximum speed $v_{\max}$, holding p constant**:
  - The average speed $\overline{v}_{\text{ff}}$ in the free-flow regime decreases towards $v_{\max} - p$.
  - The transition point at the critical space gap g_{s_c} decreases.
  - The space gap g_{s_j}, corresponding to the jam density, remains invariant.

Using equation (2.1) from Section 2.2.2, i.e., $h_s = g_s + l$, the derived $\overline{v}_{s_e}(\overline{g}_s)$ fundamental diagram can be converted into a $\overline{v}_{s_e}(\overline{h}_s)$ fundamental diagram. Because we originally started from a single-cell TCA model (i.e., the STCA), we can use our convention which states that a vehicle's length $l_i \geq 1$ cell $\propto \Delta X$ (see Section 4.2.4 for more details).

Based on equation (2.11) from Section 2.3.1.1, we note that $\overline{h}_s = k^{-1}$. This allows us to effectively transform the $\overline{v}_{s_e}(\overline{h}_s)$ fundamental diagram into a $\overline{v}_{s_e}(k)$ fundamental diagram. As can be seen in the left part of Figure 5.2, increasing the stochasticity while holding $v_{\max}$ constant, leads to the same observations that we previously mentioned. Finally, using the fundamental relation of traffic flow theory (2.33) (see Section 2.3.4.2), our constraints are transformed into an equivalent *triangular* $q_e(k)$ fundamental diagram. Applying this technique results in the following analytical expressions for the parameters of the LWR's fundamental diagram (based on the conversions in Section 4.2.4.1):

$$\overline{v}_{\text{ff}} = (v_{\text{max}} - p)\,\frac{\Delta X}{\Delta T}\,3.6, \tag{5.8}$$

$$k_{\text{crit}} = \frac{1000}{(v_{\text{max}} + l)\,\Delta X}, \tag{5.9}$$

$$k_{\text{jam}} = \frac{1000}{(l + p)\,\Delta X}, \tag{5.10}$$

with $l = 1$ cell as the length of all vehicles in the single-cell STCA model. Note that equations (5.8) – (5.10) contain expressions for transforming the STCA's units into their LWR real-world equivalents, i.e., from cells/time step to km/h ($\times$ 3.6 $\Delta X/\Delta T$) and from vehicles/cell to vehicles/km ($\times$ $1000/\Delta X$). Equation (5.10) deserves special attention: from the linear inequality C6 it follows that an increase of the stochasticity p leads to a larger space gap in a jam, i.e., $g_{s_{\text{jam}}} = l + p$ with l the vehicle length taken to be 1 cell, i.e., ΔX. As such, the jam density (proportional to $(l + p)^{-1}$) decreases for higher p. Because of our specific transformation of the rule set, two important things are noticed: firstly, the jam density is different for the *transformed* LWR deterministic and stochastic systems (in contrast to this, they are taken to be the same in the STCA model, i.e., 1 vehicle per cell). And secondly, the LWR model is a strictly deterministic system: jams can occur only due to the imposed boundary conditions (e.g., an increased demand, a narrowing of the road, ...). So the phenomenon of spontaneous emergence of jams in the STCA model is not carried over when transforming its rules into a fundamental diagram for the LWR model (this corresponds with our notion of implicitly introducing the stochasticity to the LWR model, as mentioned in the introduction of this chapter).

The capacity flow is calculated using the fundamental relation (2.33), resulting in the following expression that is automatically expressed in vehicles/hour:

$$q_{\text{cap}} = k_{\text{crit}}\,\overline{v}_{\text{ff}}. \tag{5.11}$$

As is visible in the right part of Figure 5.2, an increase of the stochasticity leads to a lower capacity flow q_{cap}, an invariant critical density k_{c}, and a smaller jam density k_{j}.

> In conclusion, we note how rewriting the STCA's rule set allowed us to obtain a stationary triangular $q_{\text{e}}(k)$ fundamental diagram. This fundamental diagram, which implicitly incorporates the STCA's stochasticity, can then be specified as a parameter to the macroscopic first-order LWR model of Section 3.2.1.2. Finally note that the shape of the derived fundamental diagram is dictated by the inequality constraints C1 – C9. As such, it actually represents an 'outer envelope', that is to say, all possible fundamental diagrams lie beneath this envelope. This includes curved fundamental diagrams, more generally piecewise-linear fundamental diagrams, et cetera.

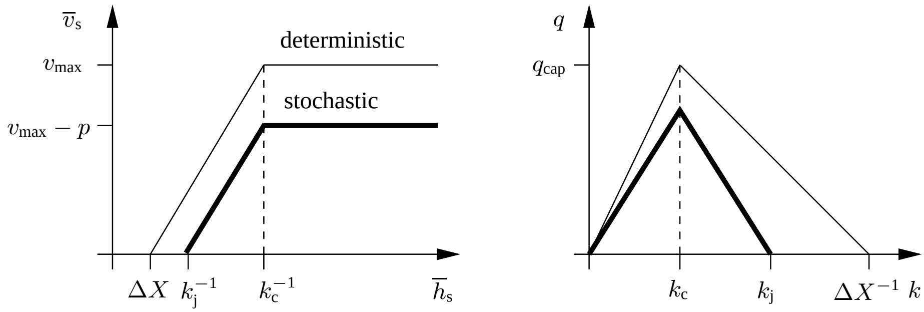


Figure 5.2: *Left:* deriving a stationary $\overline{v}_{s_e}(\overline{h}_s)$ fundamental diagram from the STCA's constraints C1 – C9. The stochastic diagram has a higher inverse jam density, but the same inverse critical density as its deterministic counterpart (for the same v_{max}). *Right:* an equivalent triangular $q_e(k)$ fundamental diagram.

5.2 Application to an illustrative case study

After deriving a relation between the STCA and LWR models by means of the process explained in the previous Section 5.1, we now apply our methodology to a small case study. We first describe the setup of the test scenario, after which we interpret and discuss our obtained results.

5.2.1 Description of the case study

The case study we consider, is modelled as a single-lane road that has a middle section with a reduced maximum speed (corresponding to, e.g., an elevation, a speed limit, ...). This road consists of three consecutive segments A, B, and C, as depicted in Figure 5.3, whereby vehicles enter the road at segment A, travel through segment B, and exit it at the end of segment C. For the STCA, we assume a temporal and spatial discretisation of $\Delta T = 1$ s and $\Delta X = 7.5$ m, respectively. The first road segment A then consists of 1500 cells (11.25 km), while the second and third segments B and C each consist of 750 cells (i.e., each approximately 5.6 km long). The maximum speed for segments A and C is $v_{max}^{A,C} = 5$ cells/time step, whereas it is $v_{max}^{B} = 1$ cell/time step for segment B. The capacity flows for all three segments are denoted as $q_{cap}^{A,C}$ and q_{cap}^{B}.

This road is simulated using both the STCA and the LWR model, each time for 3000 time steps. As for the boundary conditions, we assume an overall inflow of $q_{cap}^{B}/2$, except from time step 200 until time step 600, where we have created a short *traffic burst* of increased demand, with an inflow of $(q_{cap}^{A,C} + q_{cap}^{B})/2$. Figure 5.4 shows a close up of the individual vehicle trajectories for the STCA in a time-space diagram, near the border between segments A and B. As can be seen from the trajectories, heavy congestion sets in and flows upstream into segment A, where it starts to dissolve at

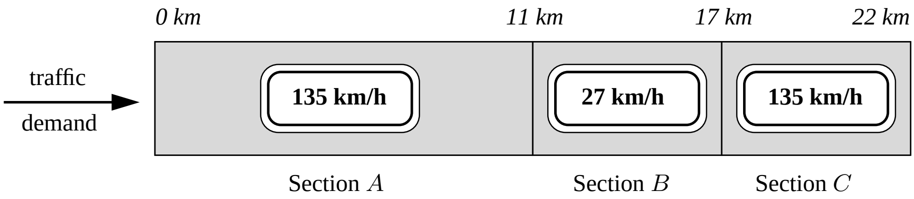


Figure 5.3: The single-lane road of the case study we consider, consisting of three consecutive segments A, B, and C. Assuming temporal and spatial discretisations of $\Delta T = 1$ s and $\Delta X = 7.5$ m, respectively, segment A is composed of 1500 cells, while segments B and C are each composed of 750 cells. The maximum speed for segments A and C is $v_{\text{max}} = 5$ cells/time step, whereas it is $v_{\text{max}} = 1$ cell/time step for segment B.

the end of the traffic burst. The result is a typical triangular-shaped region that contains a queue of slow-moving vehicles (the backward propagating waves are clearly distinguished as the pattern of parallel black and white stripes).

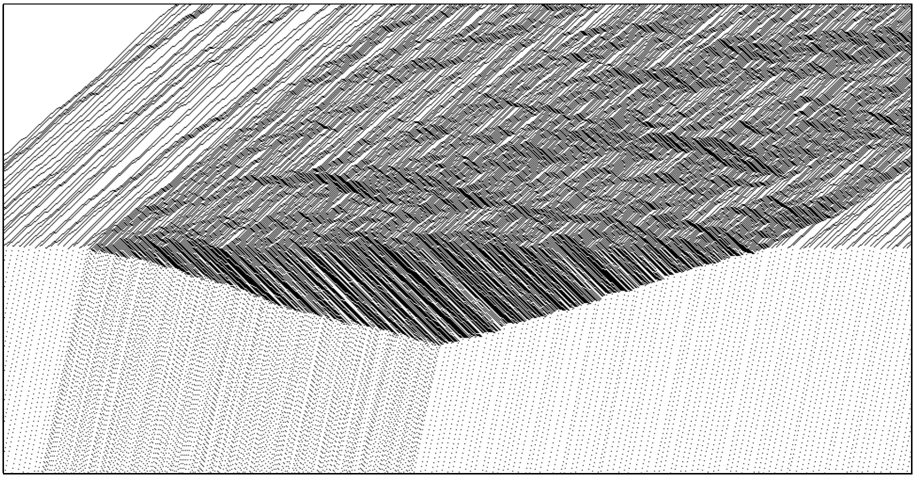


Figure 5.4: A close up of the individual vehicle trajectories for the STCA in a time-space diagram, near the border between segments A ($v_{\text{max}}^{A,C} = 5$ cells/time step) and B ($v_{\text{max}}^{B} = 1$ cell/time step), for $p = 0.1$ everywhere in the system. We can see the formation and dissolution of an upstream growing congested region at the end of segment A, related to the short traffic burst.

Applying our previously discussed methodology, we construct a stationary triangular $q_{\text{e}}(k)$ fundamental diagram. Its parameters are calculated by means of equations (5.8) – (5.11). The results are listed in Table 5.1, with the TCA's parameters expressed in cells/time step, vehicles/cell, and vehicles/time step, respectively, and the LWR's parameters expressed in kilometres/hour, vehicles/kilometre, and vehicles/hour, respectively.

$v_{\max} = 1 \quad (p = 0.1)$			$v_{\max} = 5 \quad (p = 0.1)$		
	STCA	LWR		STCA	LWR
$\overline{v}_{\text{ff}}$	0.90	24.30	$\overline{v}_{\text{ff}}$	4.90	132.30
k_{crit}	0.50	66.67	k_{crit}	0.17	22.22
k_{jam}	0.91	121.20	k_{jam}	0.91	121.20
q_{cap}	0.45	1620.08	q_{cap}	0.83	2939.71

$v_{\max} = 1 \quad (p = 0.5)$			$v_{\max} = 5 \quad (p = 0.5)$		
	STCA	LWR		STCA	LWR
$\overline{v}_{\text{ff}}$	0.50	13.50	$\overline{v}_{\text{ff}}$	4.50	121.50
k_{crit}	0.50	66.67	k_{crit}	0.17	22.22
k_{jam}	0.67	88.89	k_{jam}	0.67	88.89
q_{cap}	0.25	900.05	q_{cap}	0.77	2699.73

Table 5.1: The resulting parameters for the triangular fundamental diagrams, as calculated by means of equations (5.8) – (5.11). The STCA's parameters are expressed in cells/time step, vehicles/cell, and vehicles/time step, respectively, whereas the LWR's parameters are expressed in kilometres/hour, vehicles/kilometre, and vehicles/hour, respectively.

5.2.2 Results and discussion

The result of numerically solving the LWR model for the case of $p = 0.1$ using the Godunov method [Dag95b; Leb96] (see Section 3.2.1.4 for more details), is depicted in the right part of Figure 5.5. Note that for the LWR model, each cell in the Godunov scheme corresponds to 5 (i.e., $v_{\max}^{A,C}$) consecutive cells of the STCA model. Comparing the tempo-spatial behaviour of the LWR model to that of the microscopic system dynamics of the STCA model (i.e., the left part of Figure 5.5), we find a good *qualitative* agreement between the two approaches. The time-space diagram of the STCA is an average taken over 100 stochastic realisations of the system; in contrast to this, the diagrams of the LWR model are always deterministic in nature. With respect to the first-order traffic flow characteristics, we note that the buildup and dissolution of congestion queues are fairly analogous for both techniques.

In Figure 5.6, we show the results when repeating the same experiment, but this time with the stochastic noise p set to 0.5 for all three segments. As revealed by the shape of the dark triangular region in the LWR model (right part), the buildup and dissolution of congestion queues seems to be *exaggerated*, especially in the upstream flowing queue of segment A.

> It is interesting to note that the STCA model reveals a *higher-order effect* that is not visible in the LWR model: there exists a fan of forward propagating density waves in segment B (see the left parts of Figure 5.5 and Figure 5.6). As such, in its tempo-spatial diagram, the STCA seems to be able to visualise the characteristics that constitute the solution of the LWR model (as was visualised in Figure 3.7 in Section 3.2.1.3).

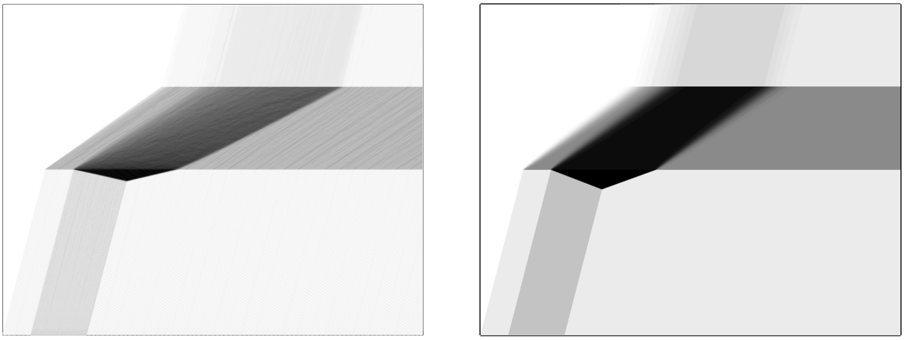


Figure 5.5: Time-space diagrams showing the propagation of densities during 3000 time steps for the road in the case study. *Left:* the microscopic system dynamics of the STCA model (the resulting diagram is an average taken over 100 stochastic realisations). *Right:* the results for the LWR model. In both cases, $p = 0.1$, with darker regions corresponding to more congested traffic conditions. There is a qualitatively good agreement between the two approaches on the level of first-order traffic flow characteristics: the buildup and dissolution of congestion queues are fairly analogous for both techniques.

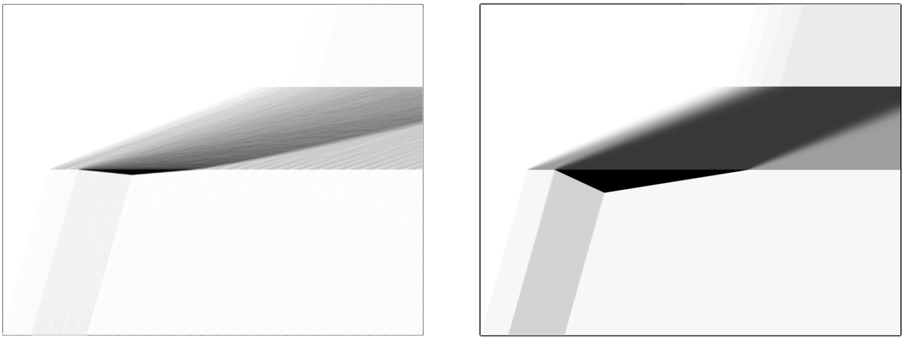
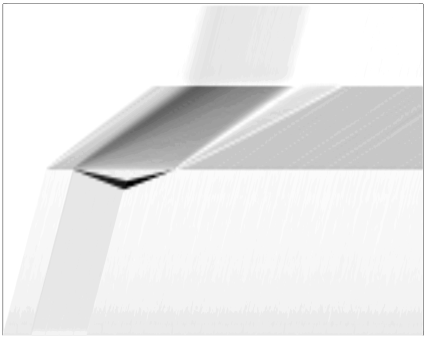


Figure 5.6: Time-space diagrams showing the propagation of densities during 3000 time steps for the road in the case study. *Left:* the microscopic system dynamics of the STCA model (the resulting diagram is an average taken over 100 stochastic realisations). *Right:* the results for the LWR model. In both cases, $p = 0.5$, with darker regions corresponding to more congested traffic conditions. As revealed by the shape of the dark triangular region in the LWR model (right part), the buildup and dissolution of congestion queues is exaggerated, especially in the upstream flowing queue of segment A.

In order to more rigourously quantify the discrepancies between the time-space diagrams of both STCA and LWR models, we provide their *absolute differences* in Figure 5.7 (i.e., we compute and plot the differences between the average densities of 100 stochastic realisations of the STCA model and the densities of the LWR model, at each grid point in the time-space diagrams). The left part shows the differences for $p = 0.1$, whereas the right part shows the differences for $p = 0.5$. The most important features to look at, are the dark coloured regions which indicate larger differences between both modelling approaches. We can clearly see that there is a problem with respect to a *quantitative* agreement between both STCA and LWR models. It appears as though the LWR model *overestimates* the STCA's capacity flows. As a result, it dissolves its jams more quickly in segment B, and it predicts a more severe onset of congestion in segment A (i.e., the triangular-shaped region containing the spill-back queue is more pronounced in the LWR's case). The sharply pronounced darker regions in the tempo-spatial left part of segment B, are due to the fact that the LWR model does not visualise the characteristics of its solution, in contrast to the STCA model which is able to give a clear indication of them.



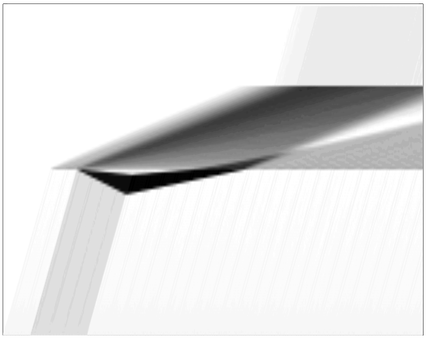


Figure 5.7: Time-space diagrams showing the differences in densities for the STCA and LWR models, during 3000 time steps for the road in the case study. Darker regions indicate large differences between both modelling approaches. *Left:* the differences for $p = 0.1$ are less pronounced, showing only a dark edge at the bottom triangular-shaped region in segment A. *Right:* the differences for $p = 0.5$, showing significant discrepancies in the bottom of the triangular-shaped region in segment A.

One of the main reasons for this discrepancy between both modelling approaches, lies in the derivation of a triangular $q_{\text{e}}(k)$ fundamental diagram for the LWR model, as was explained in Section 5.1. Because we assumed a stationarity condition on the STCA's rule set, the resulting constraints implied an invariant critical density, and always overestimated the STCA's capacity flows. In our opinion, the different behaviour of both models, mainly stems from this artifact. As a result, the discrepancies will become more articulated when increasing the stochastic noise p.

5.3 Alternate derivation of the fundamental diagram

Considering the results of the previous approach, i.e., deriving the LWR's fundamental diagram based on the STCA's rule set, and the problems related to it, the next step is to specify the fundamental diagram *directly*, based on the empirically observed behaviour of the STCA model. In the following two sections, we first discuss the effects of explicitly adding noise to the LWR's fundamental diagram, after which we discuss our obtained results when specifying the fundamental diagram directly.

5.3.1 The effect of adding noise to the LWR's fundamental diagram

Adding noise to the LWR model can mainly be accomplished via two ways: either by explicitly incorporating noise terms in the LWR equations (e.g., the conservation equation), or as a noise term in the $q_{\mathrm{e}}(k)$ relation (i.e., the fundamental diagram). We refrain from changing the LWR's conservation equation, because this amounts to introducing some form of numerical diffusion, similar to the viscosity terms in the conservation equation's right-hand side as explained in Section 3.2.1.3.

In Figure 5.8, we show the results of supplying uniformly distributed additive noise of 0.1 (left part) and 0.5 (right part). As can be seen, the introduction of noise in the fundamental diagram, leads to a 'spreading' of the solution. For small noise levels, some of the characteristics are revealed; for larger noise levels, the characteristics are clearly pronounced, including long jam dissolution times.

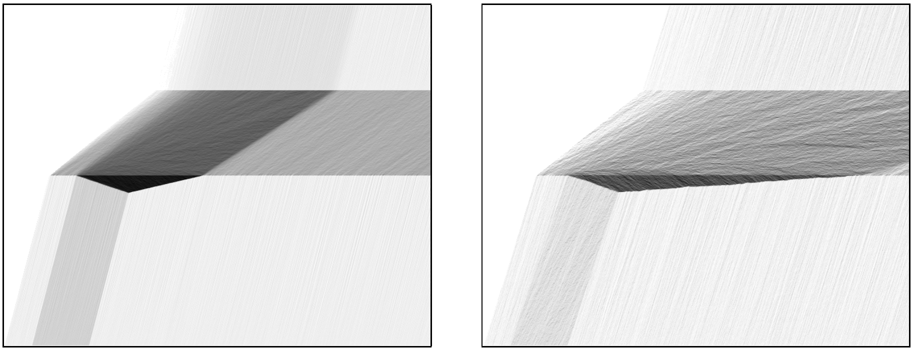


Figure 5.8: Time-space diagrams showing the propagation of densities during 3000 time steps for the road in the case study. Depicted are the results for the LWR model, with noise levels of 0.1 (*left*) and 0.5 (*right*). Higher noise levels clearly reveal the typical characteristics of the solution, and introduce longer jam dissolution times.

5.3.2 Specifying the fundamental diagram directly

Instead of deriving the fundamental diagram based on the approach taken in Section 5.1, we now try to obtain the values for the critical densities and capacity flows directly, by looking at the STCA's (k,q) diagrams in Figure 5.9 (an equivalent procedure would be to measure the capacity flows directly from the time-space diagram in Figure 5.4). Considering the STCA's (k,q) diagrams in Figure 5.9, we can estimate its capacities at approximately $q_{\text{cap}}^{B} = 0.34$ vehicles/time step $\approx$ 1220 vehicles/hour, and

$q_{\text{cap}}^{A,C} = 0.67$ vehicles/time step $\approx$ 2400 vehicles/hour for $v_{\max}^{B} = 1$ and $v_{\max}^{A,C} = 5$ cells/time step, respectively. The stochastic noise p was set to 0.1 for all three segments. Changing p to 0.5 for these segments, we can estimate the capacities at approximately $q_{\text{cap}}^{B} = 0.15$ vehicles/time step $\approx$ 540 vehicles/hour, and $q_{\text{cap}}^{A,C} = 0.34$ vehicles/time step $\approx$ 1220 vehicles/hour for $v_{\max} = 1$ and $v_{\max} = 5$ cells/time step, respectively.

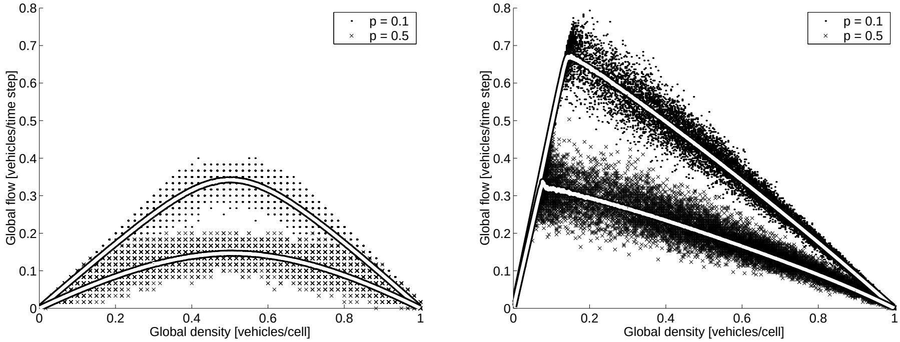


Figure 5.9: The (k,q) fundamental diagrams for the STCA model. *Left:* two diagrams for $v_{\max}^{B} = 1$ cells/time step. *Right:* two diagrams for $v_{\max}^{A,C} = 5$ cells/time step. Each time, the slowdown probability $p \in \{0.1, 0.5\}$. Small dots and crosses denote short-term averages taken over one-minute intervals, while the thick solid white lines denote long-term averages. Note how a slower maximum speed makes the diagrams more curved, and how an increasing slowdown probability leads to both a lower critical density and capacity flow.

Instead of calculating the capacity flows from the average free-flow speeds and the critical densities, as was done by means of equation (5.11), we now specify these capacity flows directly to the LWR's fundamental diagrams and calculate the critical densities from them. The results we obtained, are visualised in the time-space diagrams of Figure 5.10. Because the STCA's capacity flows are now better approximated, and not overestimated as with the previous methodology, there seems to be a better qualitative agreement for both noise levels with the STCA's time-space diagrams in the left parts of Figure 5.5 and Figure 5.6.

In Figure 5.11 we have depicted the absolute differences between this approach and the STCA's time-space diagrams. Comparing this to the previous results of Figure 5.7, we can see that in both cases the buildup and dissolution of congestion queues is in good qualitative agreement for both noise levels. *As such, we come the conclusion that it is vital to correctly capture the capacity flows of the STCA model.* Neglecting this property, can result in severe distortion of the system dynamics for higher noise levels.

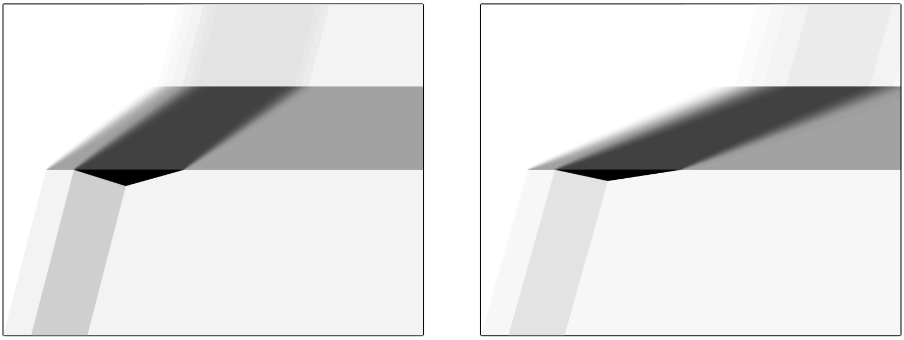


Figure 5.10: Time-space diagrams showing the propagation of densities during 3000 time steps for the road in the case study. *Left:* the results for the LWR model with $p = 0.1$. *Right:* the results for the LWR model with $p = 0.5$. In both diagrams, the fundamental diagram was specified directly to the LWR model, by explicitly stipulating the capacity flows of the STCA model. As a result, there seems to be a better qualitative agreement for both noise levels with the STCA's time-space diagrams in the left parts of Figure 5.5 and Figure 5.6.

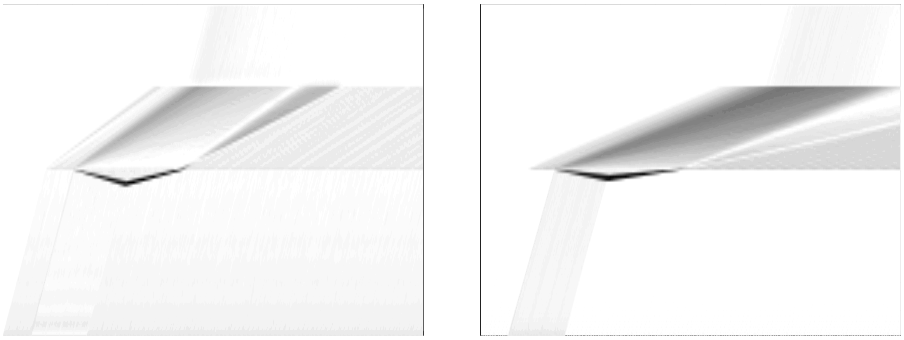


Figure 5.11: Time-space diagrams showing the differences in densities for the STCA and LWR models, during 3000 time steps for the road in the case study. Darker regions indicate large differences between both modelling approaches. *Left:* the differences for $p = 0.1$. *Right:* the differences for $p = 0.5$. In both cases, the differences are less pronounced, showing only dark edges at the bottom of the triangular-shaped region in segment A.

Note that the LWR model is able to correctly capture the first-order effects of jam buildup and dissolution, and that, due to its microscopic treatment, the STCA model allows us to visualise the higher-order effects inside jam. However, as is evidenced by this and the previous section, it is very important to correctly capture the capacity flows in the STCA model, otherwise a growing discrepancy between the LWR and STCA model is introduced with higher noise levels.

5.4 Conclusions

In this chapter, we presented an alternate methodology for implicitly incorporating the STCA's stochasticity into the macroscopic first-order LWR model. The innovative aspect of our approach, is that we derive the LWR's fundamental diagram directly from the STCA's rule set, by assuming a stationarity condition that converts the STCA's rules into a set of linear inequalities. In turn, these constraints define the shape of the fundamental diagram that is then specified to the LWR model.

For noise-free systems, our method is exact. In the presence of noise, however, the capacity flows in the derived fundamental diagram are overestimates of those of the STCA model. This discrepancy can be explained as follows: the underlying assumption for the LWR model, is that the fundamental diagram is assumed to be exact, and implicitly obeyed, i.e., the existing equilibrium relation is representative for the real traffic situation. In the original LWR formulation, this relation was also assumed to hold also for non-stationary traffic (which is a more or less reasonable assumption if we consider long and crowded roads). Our calculations have shown that a direct translation of the STCA's rule set into the LWR's fundamental diagram, does not always result in a valid fundamental diagram, especially for higher noise levels. As such, there can be a significant difference between an *average* fundamental diagram (STCA) and a *stationary* fundamental diagram (LWR). As a result, the STCA model is able to temporarily operate under larger flows and densities than those possible for the LWR's stationary fundamental diagram. A logical course of action would be to better approximate the STCA's fundamental diagram. By doing so however, we lose the advantage gained through an explicit derivation of the fundamental diagrams' outer envelope, almost certainly leading to extra conditions that need to make further assumptions about its shape. Directly specifying the STCA's capacity flows to the triangular LWR fundamental diagram, effectively remedies most of the mismatches between both STCA and LWR models.

Our methodology sees the STCA complementary to the LWR model and vice versa, so the results can be of great assistance when interpreting the traffic dynamics in both models. Especially appealing, is the fact that the STCA can visualise the higher-order characteristics of traffic stream dynamics, e.g., the fans of rarefaction waves. Nevertheless, because the LWR model is only a coarse representation of reality, there are still some mismatches between the two approaches. One of the main concerns the authors discovered, is as hinted at earlier, the fact that using a stationary fundamental diagram (i.e., an equilibrium relation between density and flow), always overestimates the practical capacity of a stochastic cellular automaton model. As such, it is vital to correctly capture the capacity flows in both STCA and LWR models, a remark that we feel is valid for all case studies.

Part III

Numerical Analysis of Traffic Data

Chapter 6

Data quality, travel time estimation, and reliability

We now turn our attention towards what is called exploratory data analysis *(EDA) of all traffic flow measurements gathered on Flanders' motorways. In a first section, we describe how all these measurements are obtained by detectors either embedded in the concrete, or by cameras positioned alongside the road, how they are stored in a central database, and how we can query this database, e.g., to give a visualisation of weekly patterns.*

We then discuss the quality of the measurements, from a statistical point of view. To this end, we both give a technique that tracks outliers and some pointers for dealing with missing values. Subsequently, we provide a methodology for quickly assessing structural and incidental detector malfunctioning; this is done by creating maps that give a clear visual indication of when and where the problems occurred.

The final section of this chapter provides a methodology for the off-line estimation of travel times, based on flow measurements (as opposed to the much used technique based on speed measurements). To conclude, we provide some reliability and robustness properties related to travel times and traffic flow dynamics, which establishes an extra instrument for the analysis of recurrent congestion.

6.1 Acquisition of traffic flow measurements

Since the last decade, a tremendous amount of traffic data is being gathered by detectors in Flanders' motorway road network (see Table 6.1 for a summary of the total length of all roads in Flanders); this data stems from over 1600 sensors in total (see Figure 6.1) [VVC03]. Until now, only data collection has been performed, but recently the Flemish government expressed interest in analysing this data. More specifically, due to the presumed high level of data corruption, it becomes worthwhile to perform quality assessments of the available data and provide corrections if possible. This will allow the Flemish motorway operating agency to use the detector data for fine tuning certain control measures pertaining to optimal flows and incident avoidance, as well as on- and off-line travel time prediction and the assessment of network reliability.

Province	*Motorways*	*On-/Off-ramps*	*Normal roads*	*Total*
Antwerpen	230 km	98 km	970 km	1298 km
Oost-Vlaanderen	203 km	80 km	1030 km	1313 km
West-Vlaanderen	187 km	94 km	1286 km	1567 km
Vlaams-Brabant	194 km	104 km	604 km	902 km
Limburg	102 km	44 km	1062 km	1208 km
Total	916 km	420 km	4952 km	6288 km

Table 6.1: The total length of all roads in Flanders, the Dutch-speaking northern part of Belgium (information cited from [AWV04]).

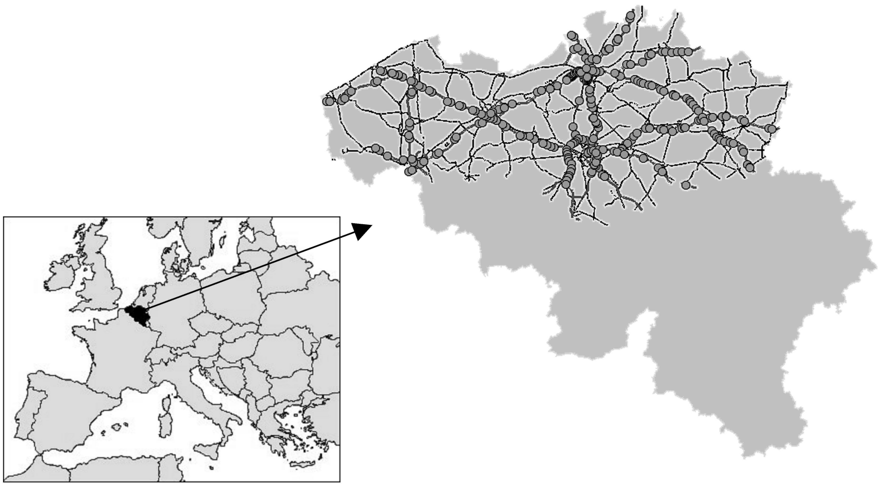


Figure 6.1: Flanders' motorway road network and its underlying secondary network of national roads, located in the northern part of Belgium. All motorways are equipped with more than 1600 sensors in total, as indicated by the locations of the gray circles (mostly single inductive loop detectors and some cameras), each minute measuring local flows, occupancies and time-mean speeds (for all lanes separately).

6.1.1 Aggregation procedures

In Belgium, there are mainly two types of detectors employed: single inductive loop detectors (SLD) embedded in the concrete and cameras positioned above or alongside the road. In the following two sections, we briefly discuss each of these devices. The third section and fourth section deal with the operational characteristics of single inductive loop detectors, and some remarks on the traditional way of estimating speeds, respectively.

6.1.1.1 Single inductive loop detectors

These are inductive loops of copper wire embedded in the concrete, typically in a rectangular setup (see the left part of Figure 6.3); they create an induced electromagnetic field that changes whenever a vehicle passes over the loop. Comparing the change in a loop's total inductance against a calibrated threshold, allows the associated controller logic, which energises the physical loop with a periodic signal, to count vehicles each time the current settles again to its stationary state. Counting the number of successive pulses corresponds to a vehicle's on-time (see the left part of Figure 6.2) [Sie]. In Belgium, the SLDs are provided by the company *Macq électronique*[1].

An SLD is sometimes called a *presence-type detector*, and is therefore only able to measure flows and occupancies. In order to get a reliable estimation of a vehicle's speed, a double inductive loop detector (DLD), consisting of two closely spaced single inductive loop detectors, can be used. The vehicle's speed is computed based on the distance between both loops and the time needed for the vehicle to travel this distance. As such, these DLDs are also called *speed traps*. Typical dimensions for an SLD are a width of 1.8 metres (i.e., half the width of a typical lane in Belgium and The Netherlands), with a length of 1.5 metres. The width assures that a typical vehicle can not avoid a detector when changing lanes. The length is taken large enough such that a small truck is considered as a single vehicle, and at the same time it is assumed to be small enough such that the individual vehicles are still counted under congested conditions. Double inductive loop detectors are spaced 1 metre apart [Bov00]. Each loop detector is connected to a circuit board that contains the controller logic which processes the changes in the coils' inductances as vehicles drive by (see the right part of Figure 6.2).

6.1.1.2 Cameras

These are mounted above or alongside the road and record all traffic that drives over a certain section of the road (see the middle part of Figure 6.3). As vehicles pass by, the image processing algorithms embedded in the camera's software detect and count these vehicles in real-time. Cameras are able to easily outperform inductive loop detectors in terms of quality of the measurements (which is of course dependent on the capability of the software to deal with varying road and weather conditions). In Belgium, there are some 200 cameras in use and all of them (as well as their accompanying software) are provided by *Traficon*[2]. Traficon essentially provides a detector

[1] `http://www.macqel.be`
[2] `http://www.traficon.be`

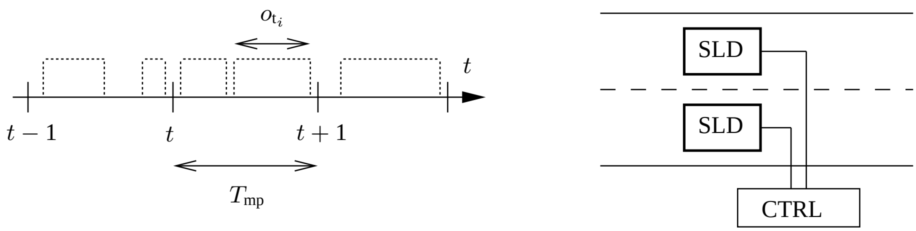


Figure 6.2: *Left:* each time a vehicle i passes over the area of a single inductive loop detector, the controller logic records the vehicle's on-time o_{t_i} as a consecutive number of pulses. The detector aggregates these on-times (see Section 2.3.3 for more details) during measurement periods of length T_{mp}. *Right:* two single inductive loop detectors (marked as SLD) embedded in a road; the coils are connected to the controller logic (marked as CTRL) which processes the changes in the coils' inductances as vehicles drive by.

board that contains a *video image processor* (VIP); this processor detects vehicles that cross lines that are superposed on the camera's picture (see the right part of Figure 6.3).

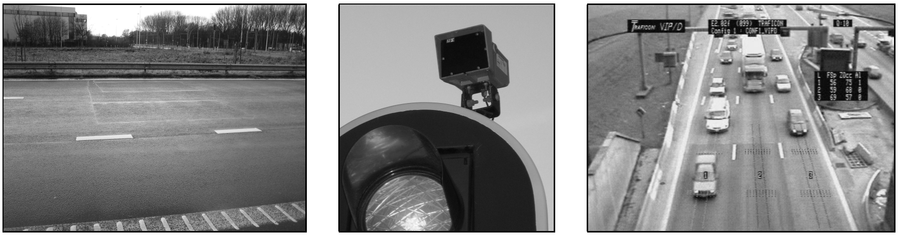


Figure 6.3: Some images of traffic detectors typically encountered in the Belgian road network. *Left:* two single inductive loop detectors embedded in the concrete. *Middle:* a traffic camera mounted on top of a traffic light. *Right:* an image sequence of a camera that is processed by a Traficon video image processor to extract local traffic data.

> Considering the measurement regions from Section 2.3 (see Figure 2.3), we note that an SLD corresponds to region R_t, a DLD corresponds to two such successive regions, whereas a camera corresponds to region $R_{t,s}$ (put more correctly, it resembles sequences of R_s regions that correspond to the video's individual frames).

Other possible detectors are pneumatic tubes which detect changes in pressure, detectors based on infrared beams, radar devices using the Doppler effect, ... Different detection schemes require different installation and maintenance costs. Nowadays, the Belgian government chooses to replace faulty single inductive loop detectors with cameras, as these latter can quickly be installed without having to completely block one or more lanes of the road.

6.1.1.3 Operational characteristics of single inductive loop detectors

Most of the detectors are located right before and after each complex of on- and off-ramps at motorways. This clearly gives a very sparse spatial distribution because there are many kilometres of road compared to the kilometres spanned by these complexes. In total, the number of detectors present in Flanders' motorway network amounts to 1654 for the year 2001 and 1800 for the year 2003: each detector is responsible for a single lane that can be located on the main road of the motorway, or on an on- or off-ramp. Measurements for each detector are aggregated every minute, i.e., $T_{\text{mp}} = 60$ s. Noting that traffic flows vary in time and space, the operations of the detectors can be seen as the stochastic sampling of these traffic flows. So we should always keep in mind that the obtained measurements are not absolute values, but samples from a stochastic distribution.

There are four *macroscopic* variables[3] that each detector i in the motorway network outputs after the elapse of each measurement period t:

- $q_{\text{c}_i}(t)$, the number of cars driving by,
- $q_{\text{t}_i}(t)$, the number of trucks driving by,
- $\rho_i(t)$, the occupancy of the detector,
- and $\overline{v}_{\text{t}_i}(t)$, the time-mean speed of all vehicles driving by.

It is important to realise that an SLD is not capable of measuring the speed of a single vehicle. This stems from the fundamental fact that the measurements are taken at a single point in space (i.e., measurement region R_{t}). And without knowing a vehicle's length, its speed can not be derived. So either the length or the speed can be calculated (provided one of the two is known), but not both. As such, only $q_{\text{c}_i}(t)$, $q_{\text{t}_i}(t)$, and $\rho_i(t)$ are measured directly; an estimate of the time-mean speed is $\overline{v}_{\text{t}_i}(t)$ is derived from these values. The detectors operate with a resolution of 50 Hz, so each $1 \div 50 = 0.02$ s the detectors record pulses due to the changing current in the loop. All detectors within one complex are connected to a counting station that contains a microprocessor to handle the signals of at most 20 SLDs (one such station controls 4 groups of at most 5 SLDs). We now explain the operation of a single inductive loop detector i that is installed in the Belgian motorway network since 1980 [Sie; Sie92].

1. All vehicles are assumed to have the same mean speed $\overline{v}_{\text{t}_i}(t-1)$, calculated during the previous measurement period.
2. When the j^{th} vehicle passes over the loop, its on-time o_{t_j} is recorded as a number of pulses (see the left part of Figure 6.2 for a schematic overview). As such, this on-time corresponds to an integer multiple of the sampling period, i.e., 2 ms.

[3]Note that in contrast to some other countries such as The Netherlands and Germany, our sensors do not measure microscopic variables such as time and space headways, … [Neu99; Kno02c].

3. The j^{th} vehicle is then classified as being either a car or a truck, based on its recorded on-time o_{t_j} and a *threshold* $\tau_i(t-1)$ that was calculated during the *previous* measurement period:

$$\begin{array}{rlcl} \text{if} & o_{\mathrm{t}_j} \leq \tau_i(t-1) & \Longrightarrow & \text{car} \\ \text{else} & o_{\mathrm{t}_j} > \tau_i(t-1) & \Longrightarrow & \text{truck} \end{array}$$

   This threshold essentially is the 'trick' that the counting stations for SLDs use to discriminate between cars and trucks. Because no vehicle lengths and speeds are known (they can not be measured by an SLD), the vehicles' on-times are used in comparison with a dynamic threshold, in order to obtain a vehicle type classification.

4. At the end of the measurement period $t + T_{\mathrm{mp}}$, the detector's controller logic has determined the number of cars $q_{\mathrm{c}_i}(t)$ and trucks $q_{\mathrm{t}_i}(t)$ counted, and the occupancy $\rho_i(t)$ according to equation (2.21) from Section 2.3.3. Furthermore, it has also calculated the *average on-time for a car*:

$$\overline{o}_{\mathrm{t}_{\mathrm{c}_i}}(t) = \frac{1}{q_{\mathrm{c}_i}(t)} \sum_{j=1}^{q_{\mathrm{c}_i}(t)} o_{\mathrm{t}_j}. \qquad (6.1)$$

5. Based on the average on-time for a car, the controller logic then determines the dynamic threshold, using a so-called *control curve*, as shown in Figure 6.4. This new threshold is to be used during the *next* measurement period. As such, there is a lag of 1 minute before the measurements can adapt to changing traffic conditions, which change the threshold.

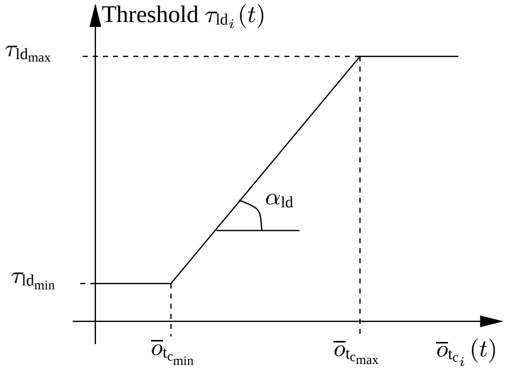


Figure 6.4: An illustration of the control curve used to calculate the dynamic threshold $\tau_{\mathrm{ld}_i}(t)$ for SLD i at measurement period t. The curve is assumed to have lower and upper boundaries, as well as a linear part that relates the threshold to the average on-time $\overline{o}_{\mathrm{t}_{\mathrm{c}_i}}(t)$ for a car.

The parameters $\overline{o}_{\mathrm{t}_{\mathrm{c}_{\min}}}$, $\overline{o}_{\mathrm{t}_{\mathrm{c}_{\max}}}$, $\tau_{\mathrm{ld}_{\min}}$, $\tau_{\mathrm{ld}_{\max}}$, and the slope α_{ld} of the linear function, are received from a central computer that sends these values every measurement period T_{mp} to *all* the single loop detectors in the motorway network. In Belgium, their values are set by an operator at respectively 9, 72, 18 and 144 (expressed in pulses of 2 ms, i.e., multiples of the sampling period). For a typical car length of 4.5 m, the values 9 and 72 correspond to speeds of approximately 120 km/h and 15 km/h, respectively. An important consequence of this, is that the detector's logic is insensitive at detecting speeds below 15 km/h. Filling in the values for the four parameters of the linear control function, the slope α_{ld} becomes equal to 2, thus giving the following equation for the function:

$$\tau_{\mathrm{ld}_i}(t) = \tau_{\mathrm{ld}_{\min}} + 2\,(\overline{o}_{\mathrm{t}_{\mathrm{c}_i}}(t) - \overline{o}_{\mathrm{t}_{\mathrm{c}_{\min}}}). \tag{6.2}$$

As indicated, the calculation of the new threshold is purely based on the number of cars, as it is assumed that the majority of the vehicles are cars, and that their individual lengths are more or less constant with an average of $\overline{l}_{\mathrm{c}}$ = 4.5 m. In case the number of counted cars $q_{\mathrm{c}_i}(t)$ is strictly less than a predefined lower bound (in Belgium, this bound is set at 6 cars), then the calculation of a new threshold value (as shown by equation (6.2) and in Figure 6.4) is omitted and the previous value is maintained, i.e., $\tau_{\mathrm{ld}_i}(t) = \tau_{\mathrm{ld}_i}(t-1)$.

6. Once the car and truck counts and the detector's occupancy are known, only one variable is missing in order to calculate the time-mean speed $\overline{v}_{\mathrm{t}_i}(t)$: the *average vehicle length* $\overline{l}(t)$ needs to be known. The effective vehicle length, *as seen by the detector*, is actually the sum of the vehicle's length l_j, and the length K_{ld} of the loop detector. This corresponds to the following equation, similar to equation (2.22):

$$l_j + K_{\mathrm{ld}} = o_{\mathrm{t}_j}\ v_j, \tag{6.3}$$

with as previously stated, $K_{\mathrm{ld}} \approx$ 1.5 m for SLDs in Belgium.

Just as the threshold $\tau_{\mathrm{ld}_i}(t)$ for the on-time is used for the classification of cars and trucks, we can look at the equivalent threshold $\lambda_{\mathrm{ld}_i}(t)$ related to the vehicle length:

$$\lambda_{\mathrm{ld}_i}(t) + K_{\mathrm{ld}} = \tau_{\mathrm{ld}_i}(t)\ \overline{v}_{\mathrm{t}_i}(t). \tag{6.4}$$

Multiplying both sides of equation (6.2) with the time-mean speed $\overline{v}_{\mathrm{t}_i}(t)$, applying equation (6.4), and filling in the values for the control function's parameters yields:

$$\lambda_{\mathrm{ld}_i}(t) + K_{\mathrm{ld}} = \alpha_{\mathrm{ld}}\ \overline{o}_{\mathrm{t}_{\mathrm{c}_i}}(t)\ \overline{v}_{\mathrm{t}_i}(t), \tag{6.5}$$

which is by equation (6.3) equivalent to:

$$\lambda_{\text{ld}_i}(t) + K_{\text{ld}} = \alpha_{\text{ld}}\,(\bar{l}_{\text{c}} + K_{\text{ld}}), \tag{6.6}$$

or:

$$\lambda_{\text{ld}_i}(t) = 2\,\bar{l}_{\text{c}} + 1.5\text{ m}. \tag{6.7}$$

In other words: the calculated threshold for the classification, based on the perceived vehicle length, corresponds to trucks which have a minimum length $\bar{l}_{\text{t}}$ of 2 × 4.5 m + 1.5 m = 10.5 m. These values can now be used to determine the average vehicle length $\bar{l}(t)$, which is a mixture of the proportions of the counted cars and trucks, and is expressed as the following weighted average, with the flows now expressed as hourly counts:

$$\bar{l}(t) = \frac{(q_{\text{c}_i}(t)\,\bar{l}_{\text{c}}) + (q_{\text{t}_i}(t)\,\bar{l}_{\text{t}})}{q_{\text{c}_i}(t) + q_{\text{t}_i}(t)}. \tag{6.8}$$

7. The final step now estimates the time-mean speed of the vehicles. It is assumed that individual vehicle lengths and speeds are uncorrelated, and that all vehicles passing the SLD during one minute have the same speed. Applying the relation between occupancy, flow, and density as expressed by equation (2.24) from Section 2.3.3, to the fundamental relation of traffic flow theory as expressed by equation (2.33) from Section 2.3.4.2, results in the following *estimation* for the time-mean speed[4]:

$$\overline{v}_{\text{t}_i}(t) = \frac{\cancel{(q_{\text{c}_i}(t) + q_{\text{t}_i}(t))}\;\dfrac{(q_{\text{c}_i}(t)\,\bar{l}_{\text{c}}) + (q_{\text{t}_i}(t)\,\bar{l}_{\text{t}})}{\cancel{(q_{\text{c}_i}(t) + q_{\text{t}_i}(t))}}}{\rho_i(t)} \tag{6.9}$$

6.1.1.4 Some remarks on speed estimation techniques

As mentioned in the previous section, the estimation of the mean speed is based on an assumed average vehicle length. The inverse of this length is called the *g-factor*, which converts occupancy to density [Hal89; Pus94]:

$$\text{space-mean speed} = \frac{\text{flow}}{\text{occupancy} \times g}. \tag{6.10}$$

It is now possible to tune the SLD's processor by estimating this g-factor, so it can be used for the calculation of the mean speed. The algorithm elaborated upon in Section 6.1.1.3 assumes constant average vehicle lengths, which implies the use of a

[4] The estimation stems from the fact that the fundamental relation is based on the space-mean speed, whereas these single inductive loop detectors are only capable of dealing with time-mean speeds.

fixed g-factor. However, it is considered bad practice to use a constant for this critical parameter, e.g., setting the mean speed fixed during free-flow conditions and estimating the g-factor, while using this fixed g-factor during congested conditions (for the latter, the fleetmix, e.g., the percentage of long vehicles, plays an important role) [Coi01; Coi03b; Kwo03]. The use of a constant g-factor can lead to flawed results, as examined by Mikhalkin et al. [Mik72], Hall and Persaud [Hal89], and Pushkar et al. [Pus94; AD94]; Dailey addresses this problem by explicitly taking into account the statistical nature of the measurements, thereby providing criteria that help to evaluate their reliability [Dai97; Dai99]. Other possible approaches are those elaborated by Coifman et al., who provide better estimations for the average vehicle length and the speed, e.g., by tuning it with estimations coming from double loop detectors [Coi01], or try to estimate the median speed instead of the mean speed at SLDs [Coi03b]. It is important to realise that flawed measurements due to a wrong g-factor can and will lead to faulty predictions of travel times based on the measured speeds [Jia01]. In light of the algorithm elaborated in the previous section, Kwon et al. describe a methodology for the real-time estimation of the portion of truck traffic on motorways, based on data from single inductive loop detectors; they assume the existence of a truck-free lane and a high lane-to-lane speed correlation [Kwo03].

> Note that the operation of the single inductive loop detectors as described above, has a negative side effect: when the average length $\bar{l}(t)$ of a vehicle is calculated, it is done using information collected during the *previous* minute. This leads to the fact that the implemented algorithm shows incorrect behaviour when the mean speed fluctuates abruptly, as the newly calculated threshold has a lag of one minute. This results in an overestimation of the number of trucks counted when the speed suddenly drops. This can be seen in Figure 6.5, where the upper part shows the mean speed and the lower part shows the percentage of detected trucks in the traffic stream: we observe a strong correlation between these two at times when the speed drops to very low values (e.g., in congestion periods). Currently, the only way to resolve this problem is to post-process the data (excluding the application of real-time corrections), as for example elaborated in the work of De Ceuster and Immers, who recalculate the average vehicle length using an exponentially weighted moving average, taken over a longer averaging period that one minute [Ceu01]. Regardless of the problem here indicated, it can be assumed that the total vehicle count, i.e., $q_{\mathrm{c}_i}(t) + q_{\mathrm{t}_i}(t)$, can be considered as the most reliable measurement from a single inductive loop detector.

6.1.2 Storage of the measurements in a central database

As already mentioned in the introduction of this section, there are over 1600 sensors located in Flanders' motorway road network (see Figure 6.1); they are mostly single inductive loop detectors, with some 200 Traficon cameras (each sensor accounts for one lane). All these sensors are grouped into *measurement posts*; these posts group sensors at a single location over all lanes in the same driving direction. They are

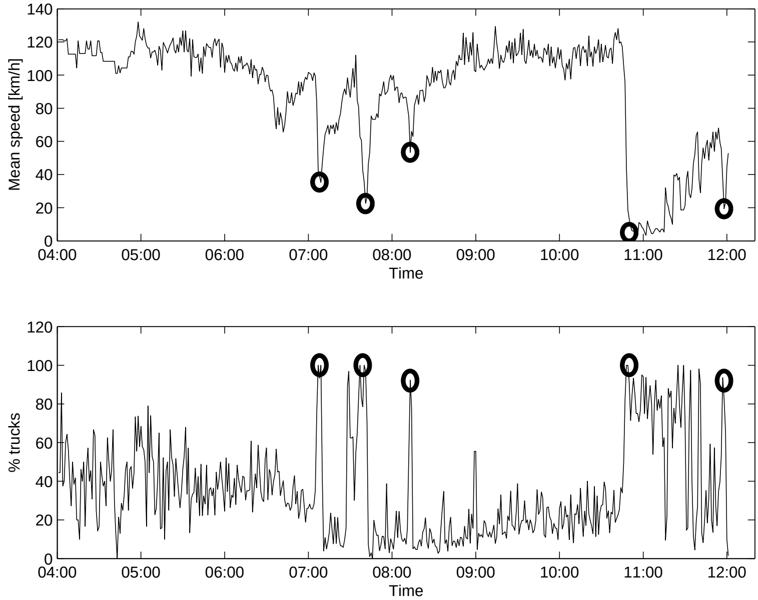


Figure 6.5: The algorithm implemented in Flanders' single inductive loop detectors shows incorrect behaviour: each time the mean speed suddenly drops, the percentage of detected trucks in the traffic stream increases (data taken from loop detectors 810 – 812 at Ternat near Brussel, during September 12, 2001 between 04:00 and 12:00).

typically located right before and right after motorway on-ramps, off-ramps, merges, and diverges. A collection of measurements posts is called a *measurement complex* [VVC03].

Several front-end computers query these sensor complexes each minute, after which the measurements are relayed to a central computer in Flanders' traffic centre (located in Antwerp). This central processor creates exchange files that get sent to the traffic centres in the Flemish and Walloon[5] regions in Belgium. These files have a lifetime of one hour, after which they get overwritten; at regular intervals they are stored into a central database that is kept at the traffic centre. As such, this database (called *MINDAT*) contains raw, unprocessed, and unvalidated data (note that no distinction is made between measurements coming from single inductive loop detectors and cameras) [VVC03]. With respect to the ranges of all stored measurements, we note that $q_{\text{c}_i}(t), q_{\text{t}_i}(t), \rho_i(t) \in \{0, \ldots, 127\}$, and $\overline{v}_{\text{t}_i}(t) \in \{0 \ldots, 255\}$. As an exception, a value of 127 for either the car flow, truck flow, or occupancy measurement is used as a sentinel value that is automatically placed in the database in case of a transmission failure. When such a failure occurs at the level of a measurement complex, all results stemming from the sensors corresponding to the measurement posts will have this value.

[5]The traffic centre in the Walloon region is located in Perex, Daussoulx.

From the above description, we know that each detector collects a measurement quadruple each minute. For one year, this corresponds to 60 minutes/hour $\times$ 24 minutes/day $\times$ 365 days/year, i.e., 525,600 measurements. Given the fact that each quadruple comprises 4 bytes, this corresponds to 2,102,400 bytes, or 2.01 MB. In the database, there are 1654 (for the year 2001), respectively 1800 (for the year 2003) sensors stored, resulting in a grand total of 869,342,400, respectively 946,080,000 measurements, corresponding to 3,477,369,600, respectively 3,784,320,000 bytes, or 3.24, respectively 3.52 GB. Compare this to another system, e.g., the famous *California Freeway Performance Measurement System* (PeMS), which has some 26,000 loop detectors, aggregating data at 30 second intervals into a database of 2 GB per day [Var05].

> For our study, we were able to obtain a database containing all the measurements of the SLDs and camera's in Flanders' motorway network (as depicted in Figure 6.1) for the years 2001 and 2003. It is interesting to note that it took nearly six months of extensive lobbying before the bureaucratic administration was able to handle and grant our request.

6.1.3 Visualising weekly patterns

Depending on the day of the week, the transportation demand will vary from location to location and time period to time period. In order to have a quick look at these phenomena, we have provided several three-dimensional charts in Figure 6.6 and Figure 6.7, plotting the total flows (cars plus trucks) for all days in 2001 and 2003, grouped together by the day of the week. As such, we pooled together all *similar weekdays* for both years, each time resulting in seven different data sets that contain all Mondays, Tuesdays, Wednesdays, Thursdays, Fridays, Saturdays, and Sundays, with holidays filtered from them [Chr00]. A blind-box approach would be to cluster these weekdays automatically based on the data itself, but in this case we know each day of the year *a priori*, so the clustering could be done manually[6].

In both figures, the flows during each day of a year are represented for all weekdays separately. For each of these days, the shown flows are an average taken over all detectors in Flanders' motorway road network; as such, they represent an average travel behaviour over the complete network. Looking at Figure 6.6, we can see that Mondays, Tuesdays, Wednesdays, and Thursdays have very similar patterns. Fridays however are a bit different, in that the evening rush hour starts earlier and has a broad peak. Considering the weekends, we note that Saturdays and Sundays are characterised by the absence of a morning rush hour. Furthermore, Saturdays have an evenly distributed afternoon peak, starting at approximately 10:00, lasting until 20:00; the peak for Sundays starts a bit later and is more intense towards the end. Comparing the weekly patterns of 2001 in Figure 6.6 and those of 2003 in Figure 6.7, we note that weekdays exhibit a slightly lower traffic demand, as opposed to the weekends where the traffic demand is more or less the same.

[6] Note that we did not take into account special events (e.g., concerts, sports, ...) or different weather conditions (e.g., summers versus winters).

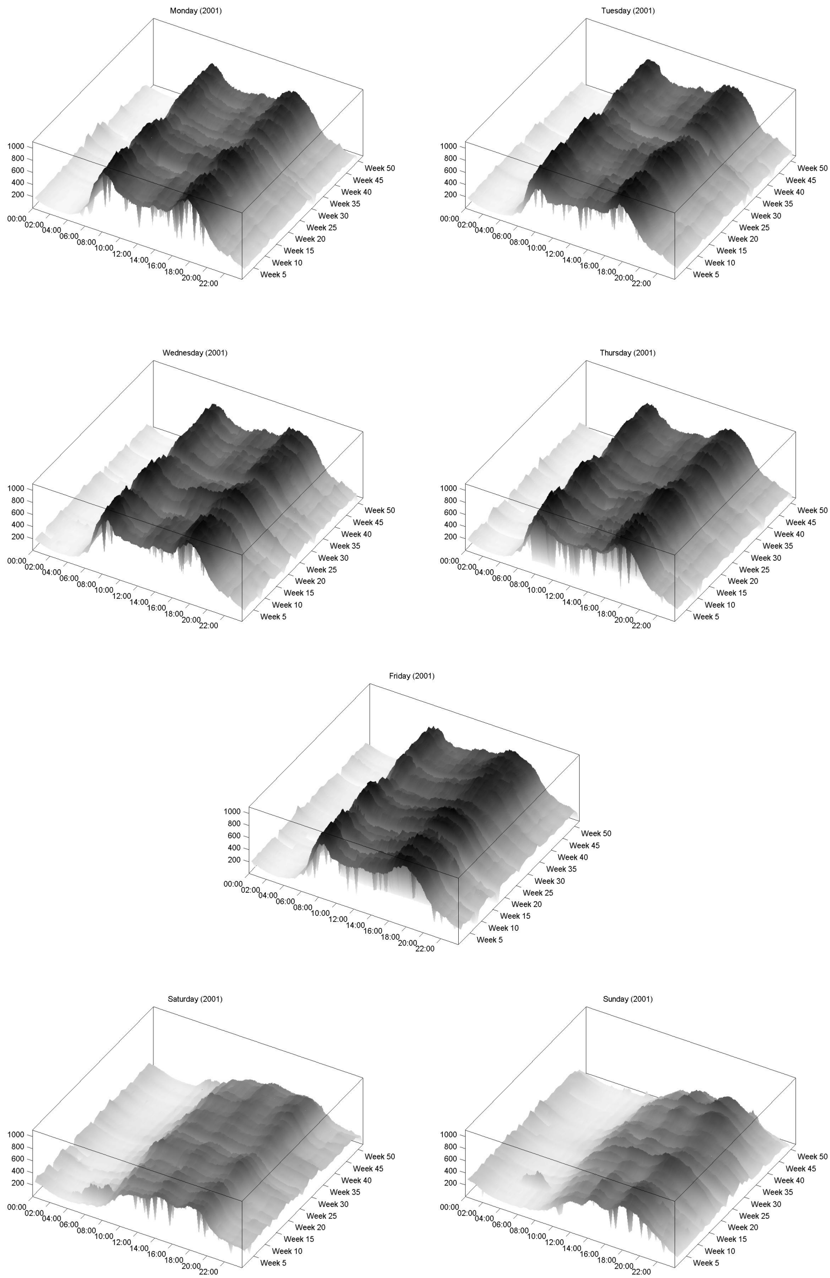


Figure 6.6: Three-dimensional charts that show the total flows (cars plus trucks) for all days in 2001; each day of the week is represented separately, with holidays filtered from them. For each of these days, the shown flows are an average taken over all detectors in Flanders' motorway road network, represent an average travel behaviour over the complete network.

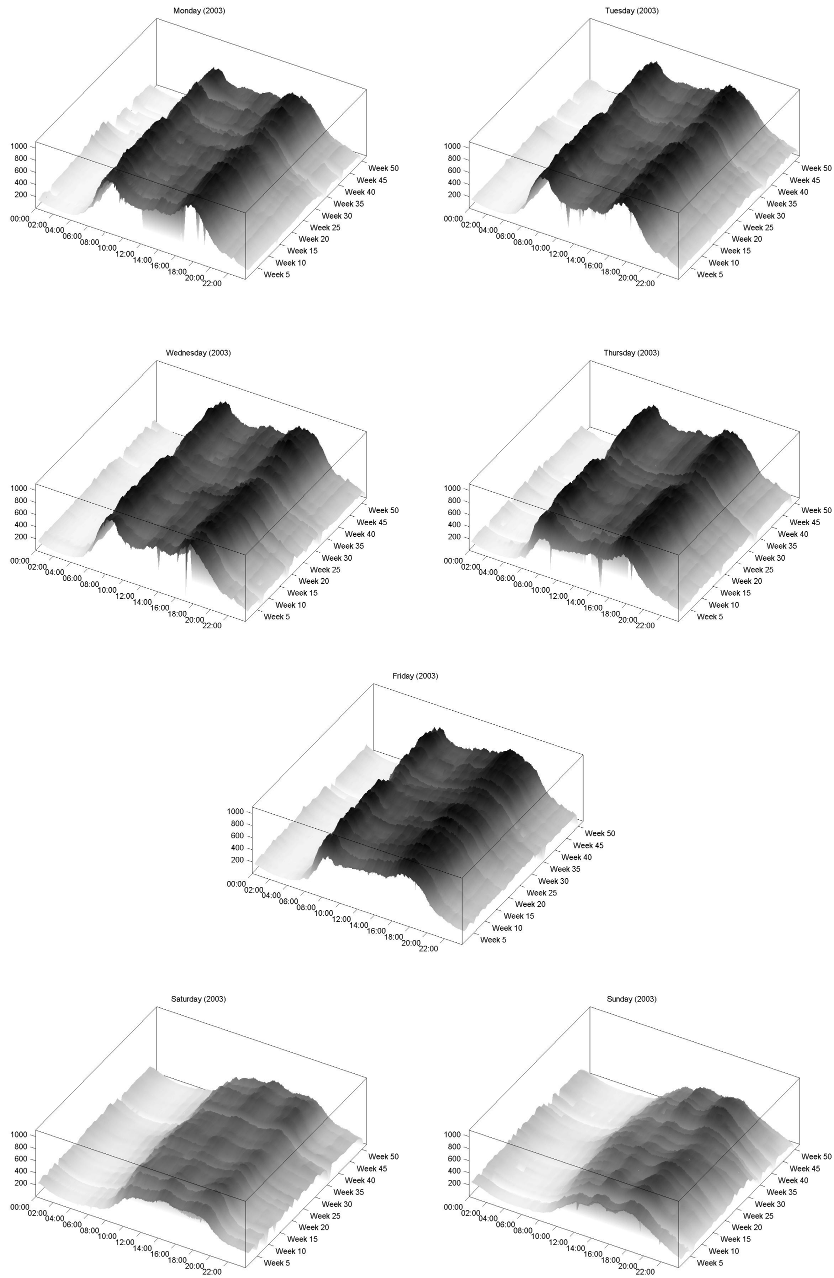


Figure 6.7: Three-dimensional charts that show the total flows (cars plus trucks) for all days in 2003; each day of the week is represented separately, with holidays filtered from them. For each of these days, the shown flows are an average taken over all detectors in Flanders' motorway road network, represent an average travel behaviour over the complete network.

In order to more rigourously assess the differences between the different days of the week, and the evolution from 2001 to 2003, we have provided contour plots of the standard deviations of the flows as they are averaged over all existing detectors in Flanders' motorway road network. In Figure 6.8 we show the results for 2001, whereas Figure 6.9 shows the results for 2003; each time, we plot the standard deviations for Mondays, Fridays, Saturdays, and Sundays (because Tuesdays, Wednesdays, and Thursdays are quite similar to the Mondays). Note that the white 'streaks' in the images are due to missing values, as we are working with the raw data.

Looking at the information contained in the top left (Mondays) and right parts (Fridays) of Figure 6.8 and Figure 6.9, we can see that the standard deviation is the highest during the morning and evening rush hours; it can also be seen that the peak of the evening rush hour is more spread. Furthermore, from the darker regions it can be seen that the standard deviation for 2003 lies higher than the one for 2001, implying more diversity in the mean flows recorded by the different detectors.

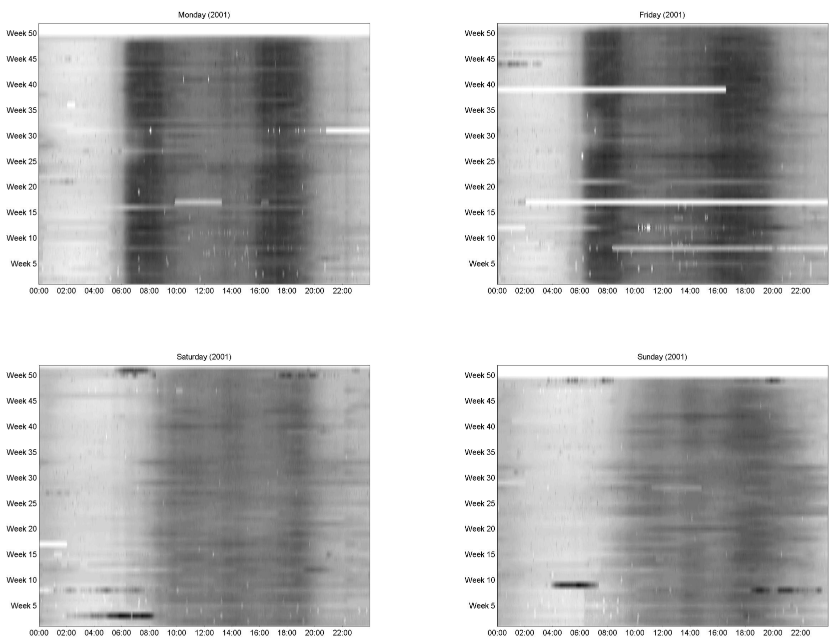


Figure 6.8: Contour plots of the standard deviations of the flows for 2001 as they are averaged over all existing detectors in Flanders' motorway road network (see also Figure 6.6). *Top-left:* Mondays. *Top-right:* Fridays. *Bottom-left:* Saturdays. *Bottom-right:* Sundays.

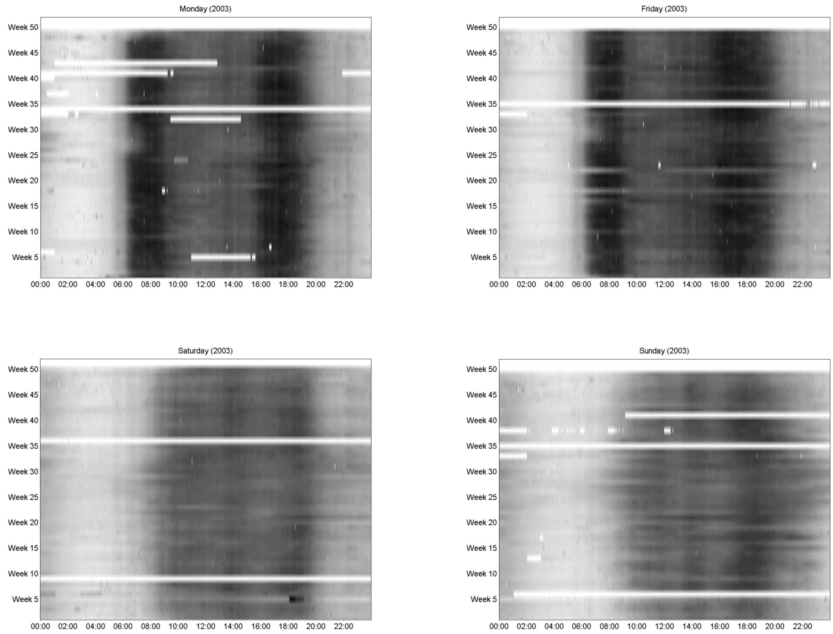


Figure 6.9: Contour plots of the standard deviations of the flows for 2003 as they are averaged over all existing detectors in Flanders' motorway road network (see also Figure 6.7). *Top-left:* Mondays. *Top-right:* Fridays. *Bottom-left:* Saturdays. *Bottom-right:* Sundays.

6.2 Quality assessment of the measurements

With respect to quality of the measurements, it is well known that single inductive loop detectors are notorious for their errors. These errors are caused by factors such as external influences by a change in the environmental temperature, faulty calibrations, detector cross talk, chattering, transmission failures, ... Some of them cause high values for the flow to be reported, and in some cases a detector blanks completely resulting in no measurements at all. When transmission errors to the front-end computers occur (see Section 6.1.2), typically the measurements of a whole detector station complex (grouping several SLDs) are lost, resulting in large gaps in the stored time series.

In this section, we first compare estimations of the mean speeds obtained from the algorithm explained in the previous section with those recorded by the detectors. We then take a look at the kinds of measurement errors that occur and the automatic detection of statistical outliers, after which we provide a methodology for quickly assessing area-wide detector malfunctioning.

6.2.1 Comparing estimations of mean speeds

During our investigation of the measurements stored in the database, we uncovered to our surprise a significant discrepancy between the mean speeds as estimated by the single inductive loop detectors and those explicitly calculated by the algorithm elaborated in Section 6.1.1.3; instead of what we expected, i.e., the same results, we obtained different estimations. In the remainder of this section, 'estimated mean speed' refers to the mean speed obtained by the SLD, whereas 'calculated mean speed' refers to the mean speed corresponding to the algorithm.

To illustrate this, we considered a sequence of SLDs (810 – 815), each of which belonged to a single measurement complex (at the E40 motorway near Ternat, two directions each consisting of three lanes). For each of these detectors, we calculated the mean speeds based on the algorithm from Section 6.1.1.3, and compared them with those as estimated by the SLDs. The results are shown in the scatter plots in Figure 6.10; all measurements were taken from the month November (30 days $\times$ 24 hours/days $\times$ 60 minutes/hour = 43,200 measurements), with black data points corresponding to the year 2001, and the gray data points to the year 2003.

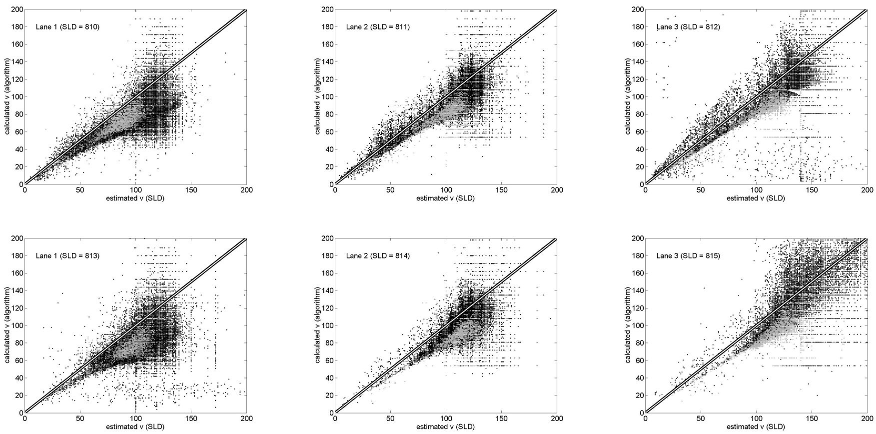


Figure 6.10: Scatter plots showing the differences between mean speeds as estimated by the single inductive loop detectors and those calculated by the algorithm from Section 6.1.1.3. The detectors belong to a single measurement complex (at the E40 motorway near Ternat, two directions each consisting of three lanes); all measurements were taken from the month November (30 days $\times$ 24 hours/days $\times$ 60 minutes/hour = 43,200 measurements), with black data points corresponding to the year 2001, and the gray data points to the year 2003 (the bissectrice is shown as the thick black/white line). As can be seen, there is good agreement for low speeds, but at moderately to high speeds the discrepancy between estimated and calculated speeds starts to grow. It is clear that the estimations in 2003 differ significantly from those of 2001, indicating a possible recalibration.

Considering the results in these scatter plots, we can see that there is good agreement for low speeds, corresponding to either low flows or high occupancies (i.e., congested conditions). However, at moderately high speeds (i.e., free-flow conditions), the discrepancy between estimated and calculated speeds starts to grow. Looking at the differences between the black and gray data points, it is clear that the estimations in 2003 differ significantly from those of 2001, indicating a possible recalibration; there is less scatter as the points are located in a more densely packed area. Furthermore, it would seem that the calculated speeds typically lie lower than the estimated speeds, especially in 2003.

In Figure 6.11, we show the same type of scatter plot, but now for detector 668 (which actually is a camera called CLOF, i.e., an acronym for 'Camera Linkeroever'; F stands for the hexadecimal numbering scheme used, i.e., the 15th camera), located at the E17 Gent-Antwerpen near Kruibeke. The scatter plot in the left part of the figure exhibits the same type of behaviour for low and moderately high speeds as explained in the previous paragraph. There is however one more visible artefact: at a relatively low estimated mean speed of 40 km/h, there is cluster of black data points (highlighted by the thick black ellipse). For these points, the estimated mean speed is fixed whereas the calculated mean speed differs significantly. After explicit investigation of the time series corresponding to these data points, it seems that the estimated mean speed fluctuates smoothly around the fixed value of 40 km/h, whereas the algorithm of Section 6.1.1.3 is better able to track the changes in occupancies (which carry more weight than the changes in car and truck flows).

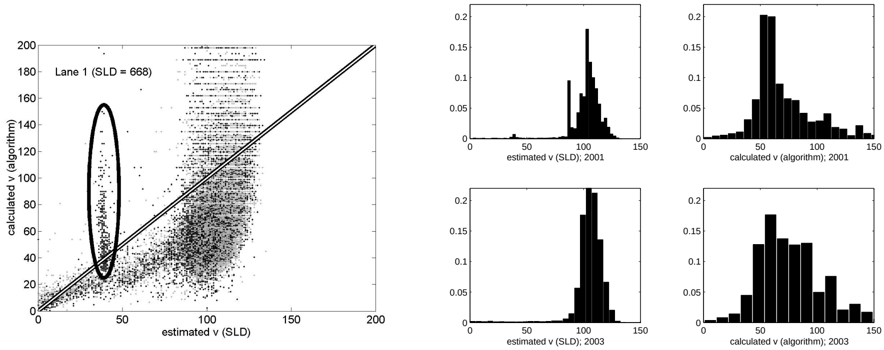


Figure 6.11: *Left:* a scatter plot showing the differences between mean speeds as estimated by the single inductive loop detectors and those calculated by the algorithm from Section 6.1.1.3. The shown detector is located at the E17 Gent-Antwerpen near Kruibeke; black data points correspond to the year 2001, the gray data points to the year 2003 (the bissectrice is shown as the thick black/white line). Note that at a relatively low estimated mean speed of 40 km/h, there is cluster of black data points (highlighted by the thick black ellipse); for these points, the estimated mean speed is fixed whereas the calculated mean speed differs significantly. *Right:* four histograms corresponding to the estimated and calculated mean speeds for 2001 and 2003. The former have a wider distribution than their latter counterparts.

For a more quantitative comparison, the right part of Figure 6.11 provides histograms of the estimated and calculated mean speeds. It can immediately be seen that for 2001, the calculated mean speeds have a wider distribution than their estimated counterparts (see also Table 6.2). As expected from the results of the scatter plots, the differences between all four distributions are more pronounced for higher than for lower mean speeds.

	Est. $\overline{v}$ *(2001)*	*Calc.* $\overline{v}$ *(2001)*	*Est.* $\overline{v}$ *(2003)*	*Calc.* $\overline{v}$ *(2003)*
Mean	102	74	114	80
Std.dev	15	33	38	34

Table 6.2: The means and standard deviations of the histograms from Figure 6.11, corresponding to the estimated and calculated mean speeds. For 2001, the calculated mean speeds have a wider distribution than their estimated counterparts. Furthermore, the differences between all four distributions are more pronounced for higher than for lower mean speeds.

6.2.2 Measurement errors and outlier detection

Faulty measurements and the like are a plague for single inductive loop detectors; as such, we take a look at what causes these errors, giving an automatic detection of statistical outliers. We first describe what is meant by these kinds of outliers, after which we explain our methodology, discuss the results and provide some pointers for dealing with missing values in the data sets.

6.2.2.1 Outliers in a statistical sense

When considering faulty measurements from detectors, we can in general distinguish between structural failures versus occasional errors. The former can be due to a miscalibration, resulting in consistently faulty data (e.g., over- and underestimations of flows, detectors that get stuck in an on-/off-position, ...). Spotting and correcting these failures is not a difficult task (it requires, e.g., a recalibration), in comparison with the latter class of occasional errors. These can have very different causes, such as detector cross talk, chattering, transmission failures, ... As a result, the detector logic can report incorrect data, for example, values that can easily be spotted are the sentinel values that get stored in the central database due to transmission failures.

Considering these ‘strange values’, we can look at them from a statistical perspective; as such, they are called *outliers*. From this point of view, *“outliers are observations that appear to be inconsistent with the remainder of the collected data”* according to Iglewicz and Hoaglin [Igl93]. The phrase *“being inconsistent with the remainder”* can be given a more mathematical characterisation by taking into account the distributions of the measurements. Values that fall outside these distributions, or those that occur in the tails of them, can then be considered as outliers. Note that from a statistical point of view, outliers are not necessarily bad values, as it is possible that these data points might come from another population/distribution [Ver05a].

6.2.2.2 Explanation of the methodology

As mentioned at the end of Section 6.2.2.1, we consider outliers to be values that are not conforming to the distribution of the measurements. In statistics, the process of automatically identifying outliers in univariate data is typically done based on the assumption that the measurements are *normally distributed*, with known mean and variance. The outliers are detected by comparing z-scores, which are measures that indicate how far a sample is located from the distribution's mean [Rou87]:

$$z_i = \frac{x_i - \mu}{\sigma}, \tag{6.11}$$

with x_i a sample taken from a distribution with mean μ and standard deviation σ. Outliers are then samples for which the z-score (expressed in units of the standard deviation) is greater than 3. Another method for assessing whether or not a sample is considered as an outlier, is by drawing a box-plot [Rou87].

The above methods might seem fine, but they are insufficient when dealing with multivariate data, consisting of n data points (*observations*) in p dimensions (*variables*): $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$. Each of these observations can be stored as a row in a $n \times p$ matrix $X = (\mathbf{x}_1, \ldots, \mathbf{x}_p)^{\mathrm{T}}$ with mean μ and covariance matrix Σ. So in order to tackle the problem of detecting outliers, we follow a similar methodology as with the z-score: for each point $\mathbf{x_i}$ in a multivariate data set, its so-called *Mahalanobis distance* (MD) is calculated [Mah36; Rou87]:

$$\mathrm{MD}_i = \sqrt{(\mathbf{x}_i - \mu)^{\mathrm{T}}\, \Sigma^{-1}\, (\mathbf{x}_i - \mu)}. \tag{6.12}$$

However, outliers contaminating a data set can introduce a severe bias of the mean and variance. To take care of this problem, we use a robust estimator, called the *minimum covariance determinant* (MCD) estimator, for which a computationally fast algorithm is available [Rou84; Rou99]. Note that it is assumed that $n > 2p$, i.e., low-dimensional data. Although the Mahalanobis distance measure explicitly takes into account the correlations of the data set, it still exhibits the bias attributed to the classical mean and variance. To this end, we now replace the standard mean μ and covariance matrix Σ by their robustly-estimated counterparts $\hat{\mu}_{\mathrm{MCD}}$ and $\hat{\Sigma}_{\mathrm{MCD}}$. The resulting *robust distance* (RD) is thus written as follows:

$$\mathrm{RD}_i = \sqrt{(\mathbf{x}_i - \hat{\mu}_{\mathrm{MCD}})^{\mathrm{T}}\, \hat{\Sigma}_{\mathrm{MCD}}^{-1}\, (\mathbf{x}_i - \hat{\mu}_{\mathrm{MCD}})} \tag{6.13}$$

Detection of outliers is now based on comparing this distance against some specified threshold. Under the assumption that the data is normally distributed, the Mahalanobis distance is χ^2 distributed; thus,we say that an observation $\mathbf{x}_i$ is considered to be an outlier when its robust distance exceeds a specified threshold, i.e., $\mathrm{RD}_i \geq \sqrt{\chi^2_{p,0.975}}$ (corresponding to a significance level α =5%).

Applying this methodology to the traffic flow measurements, we consider a data set that is bivariate ($p = 2$) in nature: each data point consists of the occupancy and the total flow, i.e., $\mathbf{x}_i = (\rho_i, q_i)$ with $q_i = q_{c_i} + q_{t_i}$. Note that even though we are working with the raw data, no correction for the number of trucks is needed as the total count is the most reliable measurement an SLD can give (it is the classification that gives problems, as mentioned at the end of Section 6.1.1.3). We used the following procedure to calculate the percentages of outliers in MATLAB:

1. For both years 2001 and 2003, we pooled together all *similar weekdays*, each time resulting in seven different data sets that contain all Mondays, Tuesdays, Wednesdays, Thursdays, Fridays, Saturdays, and Sundays, with holidays filtered from them [Chr00].

2. Out of these seven data sets, we constructed seven bivariate data matrices containing the occupancies and total flows $\mathbf{x}_i = (\rho_i, q_i)$. In order to reduce the data size, we removed all duplicate data points. Knowing that the detectors' speed estimations are the least reliable measurements, we furthermore explicitly calculated the space-mean speeds based on the detector's recorded occupancy and flow measurements using equation (6.10) from Section 6.1.1.4.

3. Because outliers in the free-flow traffic regime should not be compared to those in the congested traffic regime (due to the different distributions), we split all data sets into two non-overlapping parts. The criterion for discriminating between them was based on a combination of the free-flow speed and the critical occupancy; their respective threshold were set at $\overline{v}_{\text{ff}} = 85$ km/h and $\rho_{\text{c}} = 35\%$, respectively. Measurements below both threshold were classified as being in the free-flow traffic regime, all other measurements are assumed to belong to the congested traffic regime. Note that some part of the synchronised flow traffic regime (see Section 2.5.4.1) also belongs to our classification into a free-flow traffic regime, as this corresponds to a state of high flows at a relatively high speed.

4. For all these data sets, we now calculate a robust mean and covariance by means of the MCD estimator; to this end, we used the *Library for Robust Analysis* (LIBRA) [Ver05a]. Calculation of the MCD automatically gave us a classification for each data point as being either a regular observation or an outlier.

As an example of this methodology, we show the results for one detector in Figure 6.12. The small dots denote measurements belonging to the free-flow traffic regime; the small crosses belong to the congested traffic regime. For this particular example, the data set consisted of 49 Mondays, with 44 and 451 unique points in the free-flow and congested traffic regimes, respectively. The thick solid and dashed ellipses denote the 97.5% tolerance boundaries for both regimes, based on the results of the MCD estimator. This means that for the selected significance level, there is a probability of 5% that data points out of a large sample from a bivariate normal distribution, are misclassified as outliers outside the ellipse.

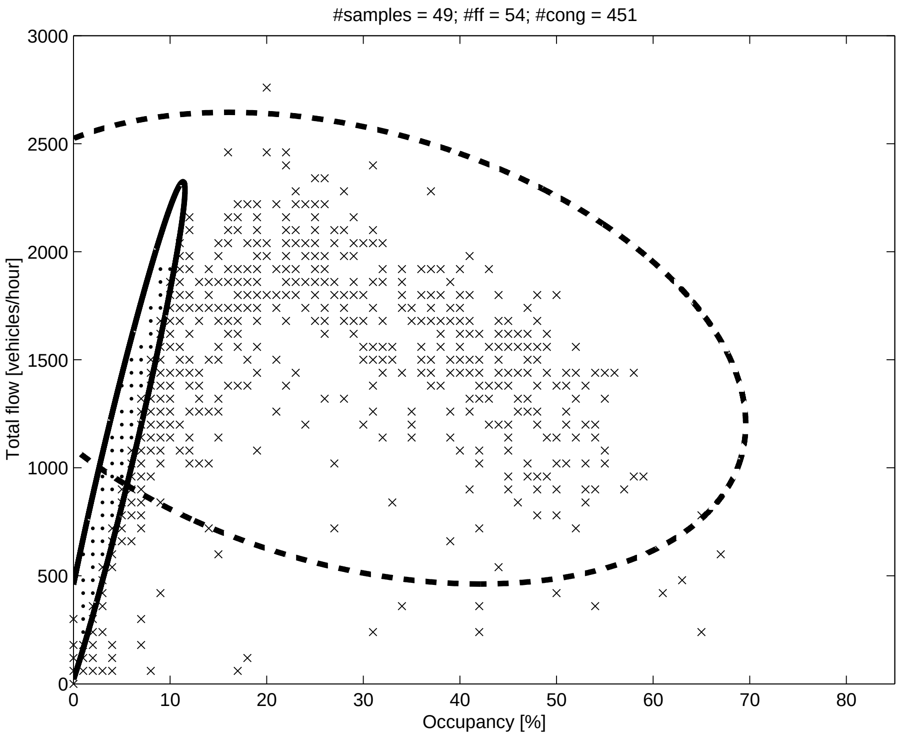


Figure 6.12: Detecting outliers in the bivariate data from the (ρ,q) diagram. The small dots denote measurements belonging to the free-flow traffic regime; the small crosses belong to the congested traffic regime. For this particular example, the data set consisted of 49 Mondays, with 44 and 451 unique points in the free-flow and congested traffic regimes, respectively. The thick solid and dashed ellipses denote the 97.5% tolerance boundaries for both regimes, based on the results of the MCD estimator.

6.2.2.3 Discussion of the results

In order to interpret our results, we constructed illustrative gray-scale images. Because there are over 1500 detectors in each year present, and only seven weekdays, the resulting images are very thin. In order to increase the visual clarity of the images, we enlarged them vertically using a rescaling factor. The results for the year 2001 can be seen in Figure 6.13, those for the year 2003 in Figure 6.14; each time, the top part shows the percentages in the free-flow traffic regime, the middle part shows the percentages in the congested traffic regime, and the bottom part shows the average of both regimes. Lighter colours denote lower percentages, whereas a complete black colour denotes an upper bound of 32% and 28.7% outliers for that detector at that weekday in the years 2001 and 2003, respectively. The detectors are arranged from left to right, with each ‘column’ in the image containing seven thin bars, one for each day of the week.

Looking at the resulting images, we can already spot errors that occurred when storing the measurements to the central database: in both Figure 6.13 and Figure 6.14 there exist ‘white vertical gaps’, denoting several detectors that malfunctioned at all weekdays, probably due to transmission failures. At these gaps, the percentage of outliers is zero, indicating that the corresponding detectors remained stuck in an on- or off-position during the entire measurement period. Furthermore, comparing the bottom parts of both figures, it seems that there were more outliers in 2003 than 2001; note that this can be an indication of an area-wide change in the calibration of the detectors. Another observation we can make is that for both years, the number of outliers in the congested traffic regime during Saturdays and Sundays is different from the other weekdays. This can be seen in the middle parts as the darker intensities in the two lower bands. There are also more outliers in the congested regime than in the free-flow regime, as can be expected.

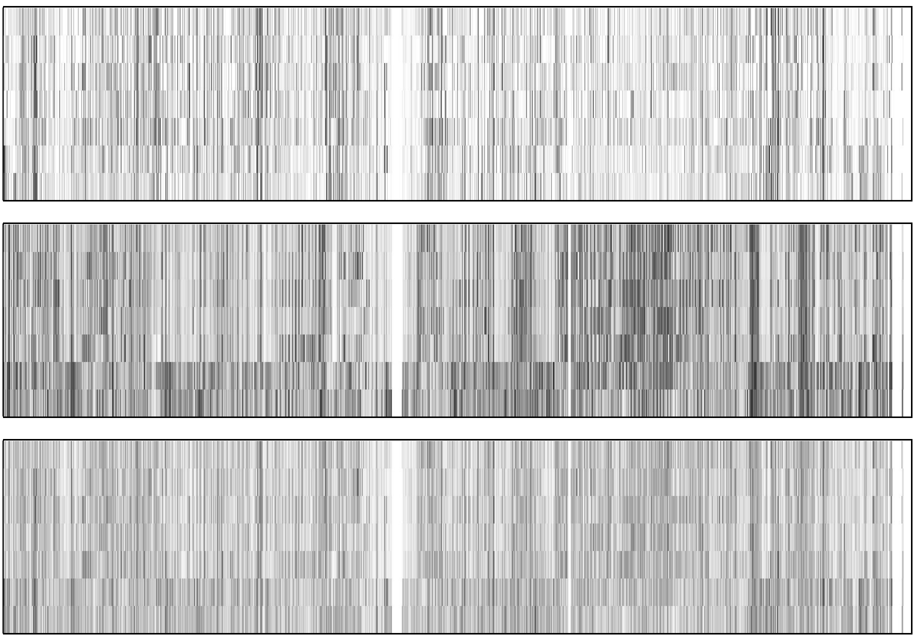


Figure 6.13: The percentages of outliers in the free-flow traffic regime (*top*), the congested traffic regime (*middle*), and the average of both regimes (*bottom*); the small thin rectangles correspond to the 1654 detectors for the year 2001, whereas the seven rows correspond to the different days of the week. Lighter colours denote lower percentages, whereas a complete black colour denotes an upper bound of 32% outliers for that detector at that weekday

6.2.2.4 Dealing with missing values

To conclude this section about measurement errors, we provide some pointers with respect to dealing with missing values. When working with contaminated data, a fre-

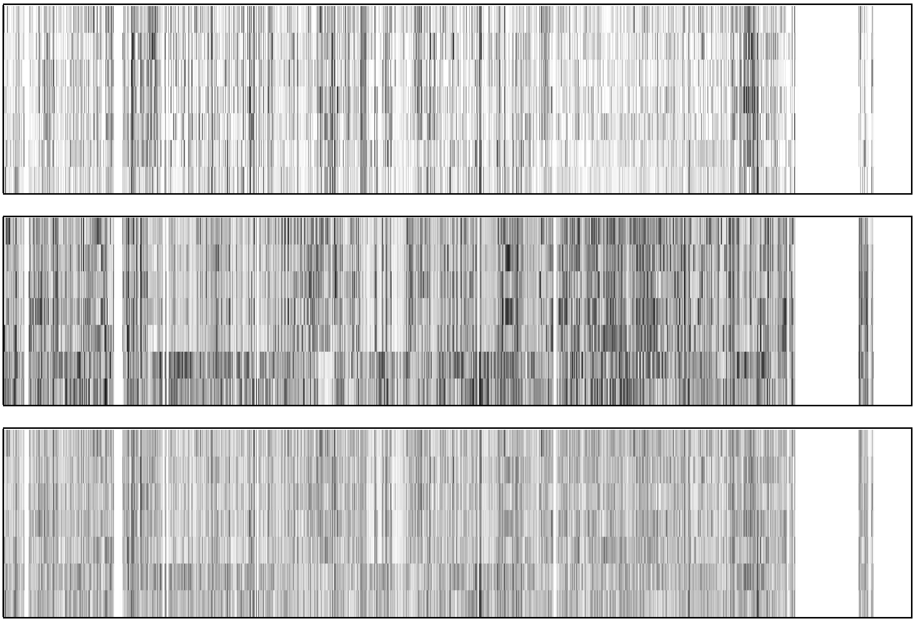


Figure 6.14: The percentages of outliers in the free-flow traffic regime (*top*), the congested traffic regime (*middle*), and the average of both regimes (*bottom*); the small thin rectangles correspond to the 1800 detectors for the year 2003, whereas the seven rows correspond to the different days of the week. Lighter colours denote lower percentages, whereas a complete black colour denotes an upper bound of 28.7% outliers for that detector at that weekday.

quently followed scheme is to first find all invalid data points (i.e., outlier detection), after which all these points are removed from the data set. As such, they are converted into missing values and the preprocessing problem now becomes one of filling in all these missing values. We highlight a few of the many possible approaches:

- **Using reference days**
  As opposed to the use of classical interpolation schemes (based on, e.g., linear or polynomial functions, splines, ...), Bellemans et al. proposed a method that is based on a *reference day*. In their work, they assumed the existence of an a priori known reference day that is representative of the day for which missing values have to be estimated. Based on the measurements $x(t-1)$ and $x_{\text{ref}}(t-1)$ at the previous time step, and the reference measurement $x_{\text{ref}}(t)$ at the current time step, the new measurement $x(t)$ is estimated as follows[7]

[7] Smith et al. later used a similar technique, called a *naive forecast* which served as a worst-case approach for predicting traffic flows [Smi02].

[Bel99; Bel00; Bel03]:

$$x(t) = \frac{x(t-1)}{x_{\text{ref}}(t-1)}\, x_{\text{ref}}(t). \tag{6.14}$$

The fraction in the previous equation plays the role for scaling the reference measurement such that it corresponds to the traffic dynamics of the day under study.

- **Multiple imputation**
  One popular way for dealing with missing values is by means of *imputation*, i.e., 'filling them in' based from samples drawn from a probability distribution [Lit87]. In principle, using Bayesian methods is a suited methodology for obtaining valid estimates for these missing values: once their distributions are known or estimated, the sought-after posterior probability can be calculated as the ratio between the likelihood times the prior and a normalising constant. In practice however, it is not always feasible to carry out such a full Bayesian analysis due to complexity issues, normality assumptions, …

  In short, *multiple imputation* (MI) can be summarised as follows: given an incomplete data set, the first step is to detect and fill in the missing values based on an imputation model that gives values drawn from a distribution. This is done not once but m times (hence the name 'multiple' imputation), resulting in m different complete data sets. Each data set is then analysed separately, after which the m results can easily be combined. A nice advantage of the MI method is that the value for m does not need to be large, e.g., $m = 10$ is typically sufficient [Rub87]. In the first step, the small number of imputed values can be drawn from predictive distributions by, e.g., a *Markov chain Monte Carlo*[8] (MCMC) method [Sch99b].

  Alternate approaches would be to use *maximum likelihood estimations* (MLE), which can be iteratively computed by a technique such as *expectation-maximisation* (EM) [Dem77; Myu03]. Advantages of using MI compared to MLE, are that it can work better with smaller sample sizes, and the fact that model used for analysing the results can be different than the imputing model that was used to obtain values. The main difference between both approaches, is that missing values are dealt with implicitly in the MLE method, whereas they are dealt with prior to the analysis in the case of MI [Sch02b].

- **Time series analysis**
  Probably the most employed methodology in classical time series analysis (TSA) is the approach towards forecasting known as *Box-Jenkins analysis*. Simply put, the analysis is based on what is known as a *autoregressive integrated moving average* (ARIMA) model. Box et al. provided a complete method for removing trends and seasonal effects, by means of differencing the time series. As such,

[8] In MCMC methods, the values are drawn from probability distributions based on Markov chains. These latter are discrete-time stochastic processes, in which the past is irrelevant for predicting the future.

an ARIMA(p,d,q) model expresses a time series as a combination of current and past observations, with p, d, and q the orders of the autoregression, integration (the differencing), and the moving average, respectively. Autoregression determines the relevance of previous values with respect to the current value, integration takes care of detrending the time series to make it stationary (i.e., the mean and variance remain constant over time), and the moving average allows smoothing of the time series [Box76; Box94].

The previously outlined methodology works well for finite-dimensional linear models; however, when considering, e.g., chaotic processes, the technique fails due to the inherent chaotic transitions and the presence of a continuous Fourier spectrum. To cope with this, we can look in another way at time series analysis, i.e., by means of *non-parametric models* that rely on the *state space* of the underlying dynamical system. Because a mathematical model of such a system is not always available, the state space can be constructed from a single time series by means of a process called *attractor reconstruction*. The principal method to this end is called *delay coordinate embedding* (DCE); it was derived by Packard et al. and put into a rigid mathematical formulation by Takens [Pac80; Tak81]. The idea behind DCE is that from the single time series, a set of new time series is constructed; each of these series is a *time-shifted version* of the original one[9]. If we assume that the time series is expressed as a sequence of observations $\mathbf{x}(t) = \{x_1(t), \ldots, x_n(t)\}$, belonging to an n-dimensional space, then the DCE method results in a vector $\mathbf{r}(t) = \{\mathbf{x}(t), \mathbf{x}(t-\tau_{\text{DCE}}), \ldots, \mathbf{x}(t-(m_{\text{DCE}}-1)\tau_{\text{DCE}})\}$. In this derivation, τ_{DCE} is called the *delay* and m_{DCE} the *embedding dimension*. The powerful result of Takens proves that if both the embedding dimension m_{DCE} and the delay τ_{DCE} are selected in an optimal fashion, then the dynamics of both the reconstructed state space and the system's original state space are topologically identical [Tak81; Sau93; Par98]. The search for the optimal values for both parameters is guided by techniques such as *average mutual information* (AMI) for the delay, and the *false nearest neighbours* (FNN) algorithm for the embedding dimension [Fra86; Ken92]. Practical implementations for this kind of analyses can be performed using, e.g., the *TISEAN* package [Heg99].

If the embedding dimension gets larger than 2 or 3, then a visualisation of this high-dimensional data becomes problematic. One way of dealing with this is by means of so-called *recurrence plots* (RP), invented by Eckmann et al. [Eck85; Eck87]. In this kind of plot, all information from the trajectory of the time series that was constructed by applying the DCE method is converted into a two-dimensional image: each point (i,j) in such a plot is then shaded according to the distance between two corresponding trajectory points $\mathbf{r}(i)$ and $\mathbf{r}(j)$. If each point in such a recurrence plot is compared to a predefined threshold, then the resulting black-and-white image is called a *thresholded recurrence plot* (TRP).

[9] It is also possible to use *derivative coordinates* instead of delay coordinates, as was done in the original work of Packard et al. However, these derivatives have proven to be susceptible to noise and are therefore not generally useful [Par98].

As an example, we show three RPs in Figure 6.15. The left part is obtained from a time series that essentially is generated from uniformly distributed noise; there are no clearly delineated structures present. The middle part shows a TRP of a sinusoidal time series; as can be seen, the image exhibits a large degree of periodicity in its structures. The right part shows the results for a time series that contained a drift due to slowly varying parameters.

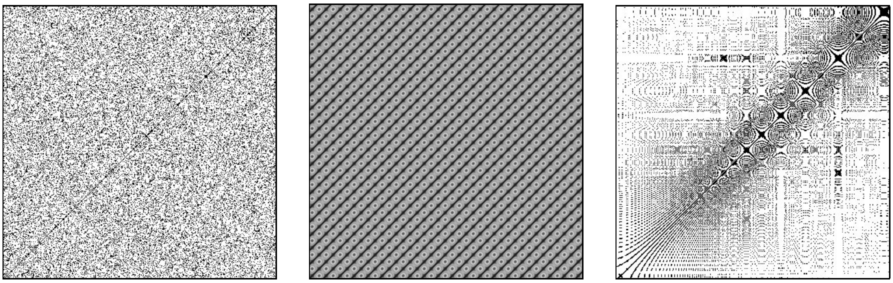


Figure 6.15: *Left:* a recurrence plot obtained from a time series that essentially is generated from uniformly distributed noise; there are no clearly delineated structures present. *Middle:* a thresholded recurrence plot obtained from a sinusoidal time series; as can be seen, the image exhibits a large degree of periodicity in its structures. *Right:* a recurrence plot obtained from a time series that contained a drift due to slowly varying parameters.

Assessing the structures in these RPs remains a somewhat 'visual discipline'; to cope with this, Zbilut and Webber extended a techniques called *recurrence quantification analysis* (RQA) that allowed a more quantitative treatment of RPs. Their technique is based on five statistics that describe phenomena such as recurrence, determinism, entropy, trend, and the largest positive Lyapunov exponent (which is a measure for the chaoticity of a system) [Zbi92; Web94]. Intuitively, these measures are related to visual features such as the percentage of lines to the main diagonal, the distribution of the lengths of diagonal lines, ... In light of the difficulties encountered by selecting the optimal embedding dimension, a promising result was obtained by Iwanski and Bradley, who state that it is possible to get the same RQA results without embedding [Iwa98].

With respect to the application of time series analysis to traffic flow data, we note the interesting result from Smith et al. In their work, they compared the use of classical ARIMA modelling to non-parametric modelling based on DCE. Their results indicate that the latter did not approach the performance of the former; this leads them to the belief that traffic data is rather stochastic as opposed to chaotic [Smi02].

6.2.3 Assessing detector malfunctioning

As already hinted at in the introduction of this section, most single inductive loop detectors exhibit a large degree of errors, missing and/or incorrect values, ... In order to provide a more qualitative assessment of these errors, we adopt a screening methodology that has been used in the PeMS project (see the end of Section 6.1.2 for more information); this allows us to provide clearly structured maps that allow a quick visual inspection of all detectors in Flanders' motorway network and their operations during the years 2001 and 2003. As a first part of this section, we explain the methodology behind the screening of the detector data, after which we provide and discuss the detector maps.

6.2.3.1 Explanation of the methodology

When a detector malfunctions (or even an entire measurement complex), its errors typically result in under- or overestimations of the flow, high occupancy values, blank data, ... Early methods for screening the measurements are based on acceptance and rejection regions in the scatter plots of a (k,q) diagram. For example, Payne et al. created tests on the bounds of minimum and maximum flows, occupancies, and mean speeds, in order to discriminate between good and bad 20-second and 5-minute samples of detector data [Pay76]. Another algorithm was constructed by Jacobson et al. at the University of Washington; their *Washington Algorithm* provides an explicit acceptance region within the (k,q) diagram [Jac90].

A more recent approach was followed by Chen et al., which resulted in the *Daily Statistics Algorithm* (DSA), currently used in the PeMS project [Che03; Bic04]. The idea behind this algorithm is to consider all measurement samples of a loop detector for one day, calculating four different scores based on these samples, and then, by comparison with some predefined thresholds, deciding whether or not the detector is considered to be malfunctioning. These scores check (1) for zero occupancy samples, (2) strictly positive occupancy samples with zero flow, (3) high occupancy samples, and (4) the entropy of these occupancy samples. The main strength of this algorithm is that it allows to test for detectors that continuously report faulty data, e.g., being stuck in the on/off position (although it is also possible that no vehicles crossed the detectors at all). For our study, we used the following scores:

$$
\begin{aligned}
S_1(i, T_{\text{DSA}}) &= \text{number of samples during } T_{\text{DSA}} \text{ with } \rho_i = 0, && (6.15)\\
S_2(i, T_{\text{DSA}}) &= \text{number of samples during } T_{\text{DSA}} \text{ with } \rho_i > \rho^*, && (6.16)\\
S_3(i, T_{\text{DSA}}) &= \text{entropy of the occupancy samples during } T_{\text{DSA}}, && (6.17)
\end{aligned}
$$

with T_{DSA} the time period over which the statistics are computed for detector i. The sentinel values in the database (typically denoting transmission failures), are automatically caught by $S_2(i, T_{\text{DSA}})$. The entropy $S_3(i, T_{\text{DSA}})$ is calculated as follows:

$$
S_3(i, T_{\text{DSA}}) = - \sum_{\rho_i \in p(\rho_i) > 0} p(\rho_i) \log(p(\rho_i)), \quad (6.18)
$$

with $p(\rho_i)$ the estimated probability density function, defined as the histogram of the occupancies ρ_i (we selected 100 bins for the estimation). The entropy provides a measure for the randomness of a stochastic variable, i.e., constant values will result in a zero entropy. In their work, Chen et al. also discuss several shortcomings of the DSA approach, most importantly the lack of exploiting spatial and temporal correlations between measurements stemming from neighbouring detectors [Che03]. As previously mentioned, in the original DSA, the daily decision on a detector being good or bad hinged on the scores that were compared to some predefined thresholds. Instead of adopting this methodology, we discard this binary classification and allow the complete range of results.

> Note that we are working with the raw unprocessed vehicle counts from the detectors, instead of converting them to passenger car equivalents. The reason is that the latter introduces an incorrect percentage of trucks, due to the problems with misclassification as mentioned at the end of Section 6.1.1.3.

6.2.3.2 Illustrative detector maps

Based on the scores S_1, S_2, and S_3 as explained in the previous section, we now provide charts of all detectors in the Flanders' region (as already mentioned in Section 6.1.2, the available data spans 1654 detectors for the year 2001, and 1800 detectors for the year 2003). To this end, we calculate these scores for each hour in both years; they are stored in matrices that have 24 hours $\times$ 365 days = 8760 columns (i.e., T_{DSA} = 60 minutes). All matrices are then normalised, after which they are converted to gray-scale images with each matrix element corresponding to one pixel in the image. We have chosen $\rho^* = 35\%$ as the predefined threshold for S_2, we chose, like done in the work of Chen et al. [Che03].

In Figure 6.16 we show the results after calculating the scores for all detectors during the entire year 2001, aggregated for each hour. The top, middle, and bottom row indicate the S_1, S_2, and S_3 scores, respectively. The darker a pixel is coloured, the higher the specific score is (black meaning that all samples during T_{DSA} contribute to the score). Figure 6.17 gives the same results, but for the year 2003.

Before we discuss both these detector maps, it is worthwhile to take a look at some general patterns that seem to occur. As these maps are highly detailed (i.e., spanning a width of some 8760 pixels), we provide two close-ups in Figure 6.18. As can be seen in the close-up to the left, there seem to be some slanted 'streaks'; these may indicate detector malfunctions at successive detectors at successive time periods. The close-up to the right reveals another more frequently occurring phenomenon, namely vertical and horizontal lines: a darker horizontal line may indicate detector failure during a certain time period; a darker vertical line may indicate several (probably neighbouring) detectors (i.e., at a measurement post or complex) that are failing. The wider a vertical line, the more extended the time period of failure. A vertical line that runs completely from top to bottom on the map, typically indicates a problem during transmission or archival of measurements to the central database; as it is highly unlikely that an area-wide malfunctioning seems to occur, it is more logical to assume that an error occurred at the database level. Note that the regular grouping of short lines in the right close-up is related to the fact that at night time, the occupancy is very low as few vehicles cross the detectors; as a result, a high number of zero occupancies is reported and shown as darker segments.

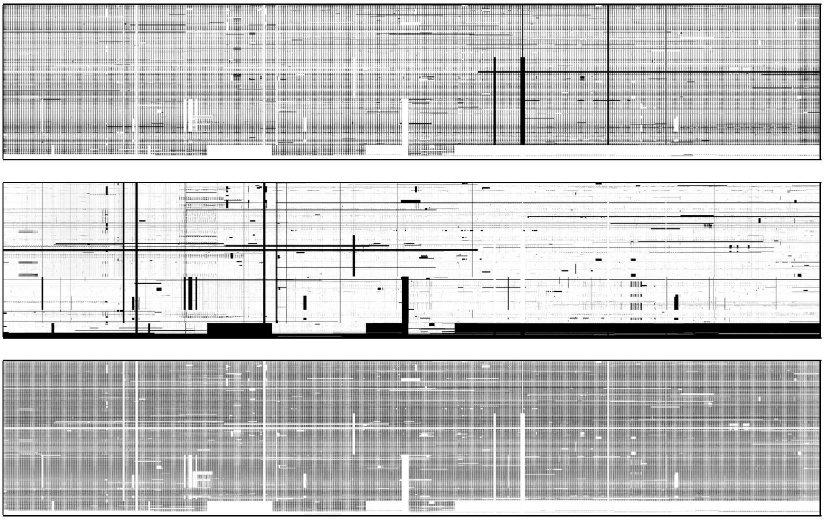


Figure 6.16: Illustrative detector maps of the S_1 (*top*), S_2 (*middle*), and S_3 (*bottom*) scores, for all 1654 detectors in the year 2001. The scores were calculated for each hour; the darker a pixel is coloured, the higher the specific score is (black meaning that all samples during T_{DSA} contribute to the score).

Returning to the detailed detector maps provided in Figure 6.16 and Figure 6.17, we can see that from 2001 to 2003, the number of detector malfunctions seems to have decreased, based on the occurrence of darker regions in score S_2 (i.e., high occupancy values). Still, as can be seen from the middle part of Figure 6.17, there are numerous detectors that seem to be malfunctioning during the entire year 2003, as is indicated by the frequent occurring of darker horizontal lines in the map. There were also several problems during transmission or archival to the central database, as is evidenced by the dark vertical lines. Finally, with respect to the entropy of the occupancy samples, we note that there seem to slightly less stuck detectors, as the white empty regions (indicating zero entropy) diminish from 2001 to 2003.

In order to consider these detector maps more quantitatively, Figure 6.19 presents histograms showing the distributions of all scores. The top row displays the results for the year 2001, the bottom row for the year 2003; the left, middle, and right histograms correspond to scores S_1, S_2, and S_3, respectively. The distinct bars in both left histograms correspond to an increasing number of zero occupancy samples, representative of traffic at night time. Furthermore, as already highlighted in the previous paragraph, the number of high occupancy values has decreased (see both middle parts). Finally, the probability of a low entropy (around 0.5) seems to have diminished from 2001 to 2003 (see both right parts).

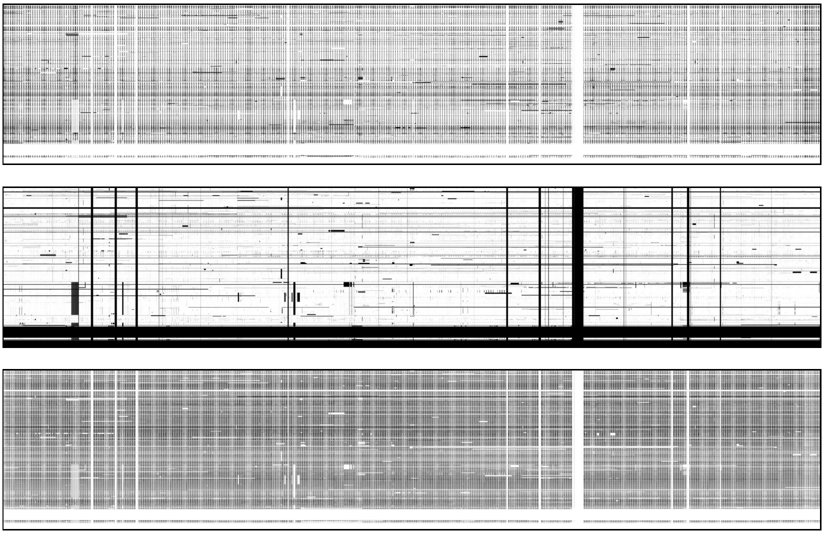


Figure 6.17: Illustrative detector maps of the S_1 (*top*), S_2 (*middle*), and S_3 (*bottom*) scores, for all 1800 detectors in the year 2003. The scores were calculated for each hour; the darker a pixel is coloured, the higher the specific score is (black meaning that all samples during T_{DSA} contribute to the score).

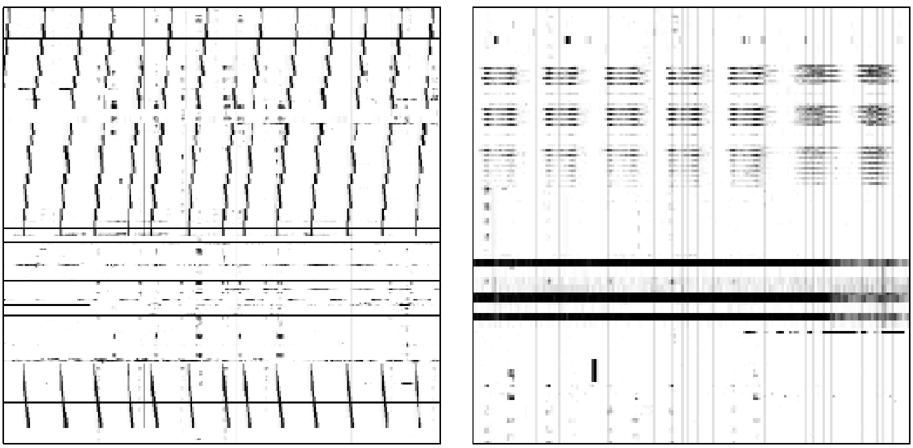


Figure 6.18: Some close-up examples of general patterns that occur in the detector maps. *Left:* the presence of slanted 'streaks' may indicate detector malfunctions at successive detectors at successive time periods. *Right:* a more frequently occurring phenomenon, namely vertical and horizontal lines: a darker horizontal line may indicate detector failure during a certain time period; a darker vertical line may indicate several (probably neighbouring) detectors (i.e., at a measurement post or complex) that are failing. The wider a vertical line, the more extended the time period of failure.

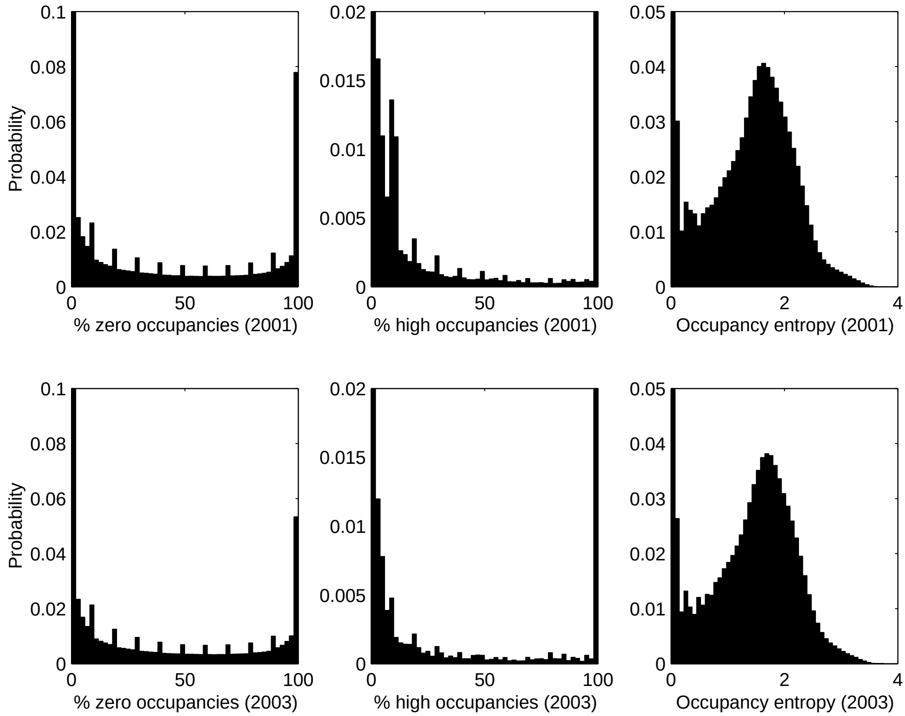


Figure 6.19: Histograms showing the distributions of all scores for the years 2001 (*top row*) and 2003 (*bottom row*). Each time the left, middle, and right histograms correspond to scores S_1, S_2, and S_3, respectively.

To conclude, we provide six more detector maps in Figure 6.20 for 2001 and Figure 6.21 for 2003. The difference between the previous maps, is that these ones show aggregated scores for whole days instead of every hour (i.e., T_{DSA} = 60 $\times$ 24 = 1440 minutes). As such, they are smaller in width, spanning only 365 pixels. In a sense, they convey the same information as presented in Figure 6.16 and Figure 6.17. Note the big black and white regions near the bottom of all six maps; they are most likely indicative of place holders in the central database for new detectors, resulting in default values for the flow and occupancies and correspondingly giving high scores.

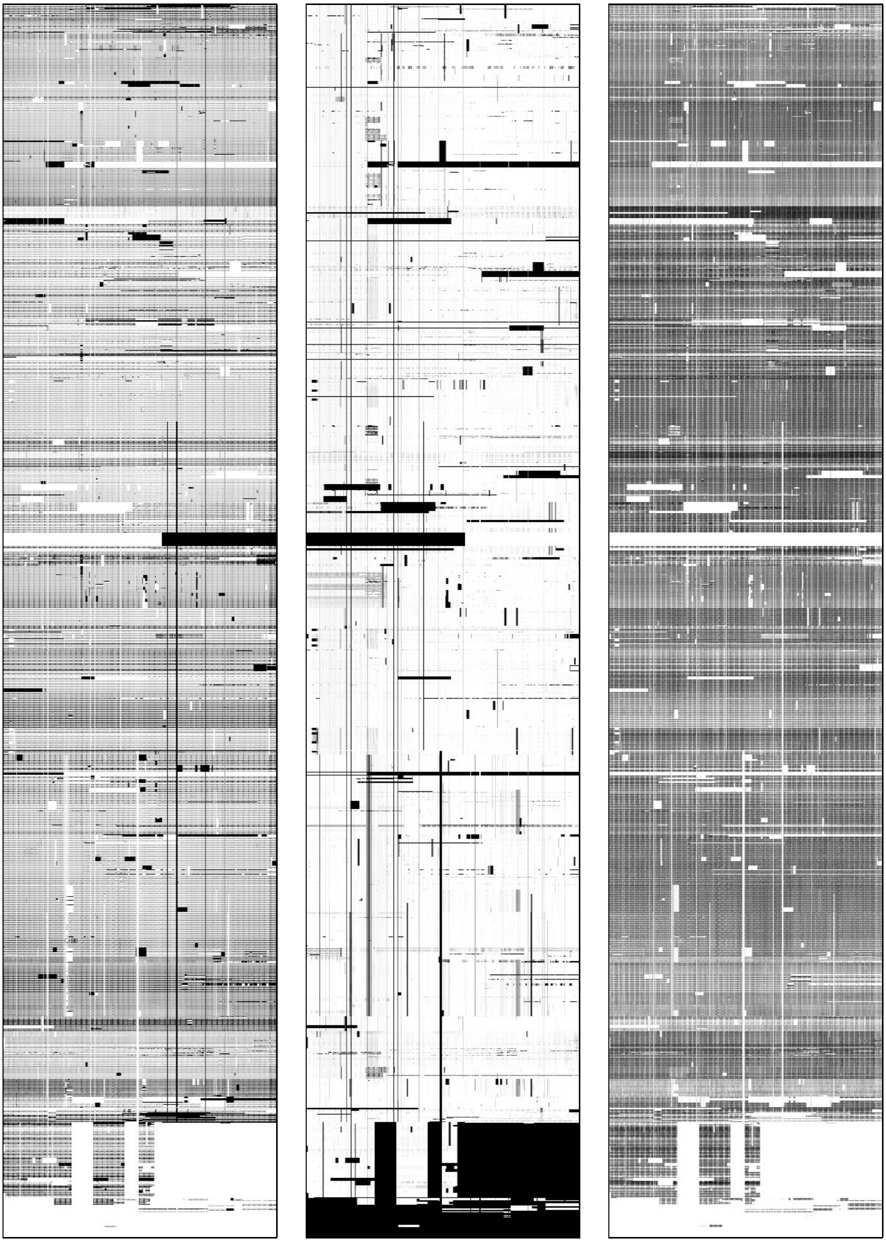


Figure 6.20: Illustrative detector maps of the S_1 (*top*), S_2 (*middle*), and S_3 (*bottom*) scores, for all 1654 detectors in the year 2001. The scores were calculated for each day; the darker a pixel is coloured, the higher the specific score is (black meaning that all samples during T_{DSA} contribute to the score). Note the big black and white regions near the bottom of all six maps; they are most likely indicative of place holders in the central database for new detectors.

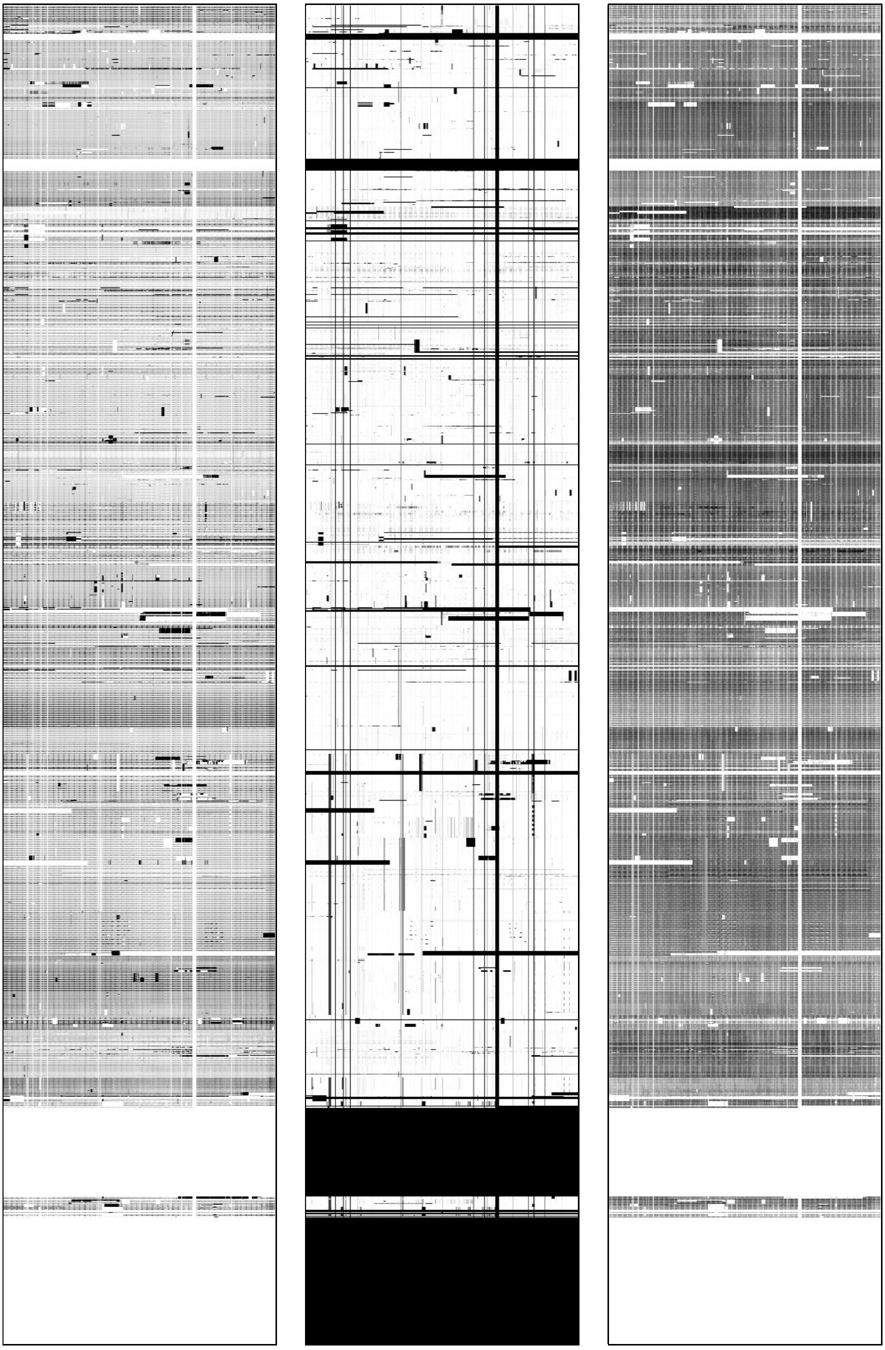


Figure 6.21: Illustrative detector maps of the S_1 (*top*), S_2 (*middle*), and S_3 (*bottom*) scores, for all 1800 detectors in the year 2003. The scores were calculated for each day; the darker a pixel is coloured, the higher the specific score is (black meaning that all samples during T_{DSA} contribute to the score). Note the big black and white regions near the bottom of all six maps; they are most likely indicative of place holders in the central database for new detectors.

6.3 Off-line travel time estimation and reliability indicators

The final section of this chapter is concerned with reliable *estimations* of the travel time in an off-line setting. As such, we are not predicting the travel times for unknown future traffic conditions. The methodology elaborated in this section can be classified according to Van Lint's taxonomy as a *data-driven modelling approach using a short-term prediction horizon; it constitutes indirect pre-trip estimation using a flow-based technique, having individual motorway sections as its spatial scope* [Lin04].

The first part of this section discusses some of the techniques that are commonly applied when estimating travel times based on historical traffic flow measurements. In the subsequent part we present our methodology for estimating travel times between measurement posts, after which we present an application to real-world measurements stemming from single inductive loop detectors. The final part discusses some reliability and robustness properties related to travel times and traffic flow dynamics.

> Note that we exclude a large part of the literature dealing with predictions; for an example of this approach we refer the reader to, e.g., the work done by Tampère et al. in the *PredicTime* project that mainly focusses on the use of Kalman filtering [Klu03; Ver03b], and the doctoral dissertation of Van Lint who employs state-space neural networks (SSNN) for reliable on-line predictions of travel times [Lin04].

6.3.1 Common approaches towards travel time estimation

The most prominent technique employed, is by deriving the travel time based on the inverse of the measured mean speed[10]. Briefly recapitulating Section 2.4.3.1, we make the distinction between the *experienced dynamic travel time*, starting at a certain time t_0, over a road section of length K [Bov00]:

$$T(t_0) = \int_0^K \frac{1}{v(t,x)} dx \qquad \forall\, t \geq t_0. \tag{6.19}$$

If not all local instantaneous vehicle speeds $v(t,x)$ are known at all points along the route, a simplification can be used, resulting in the *experienced instantaneous travel time*:

$$\widetilde{T}(t_0) = \int_0^K \frac{1}{v(t_0,x)} dx. \tag{6.20}$$

[10] Note that we do not cover another popular approach, i.e., using floating car data (FCD) to directly measure travel times; see Section 2.3.5 for more information.

As an example, the following relation gives the experienced instantaneous travel time for a section, based on the mean speed measurements from an upstream and downstream detector post:

$$\widetilde{T}_{\mathrm{D_{up}}\rightarrow\mathrm{D_{down}}}(t_0) = \frac{K}{2}\left(\frac{1}{\overline{v}_{\mathrm{s}_{\mathrm{D_{up}}}}(t_0)} + \frac{1}{\overline{v}_{\mathrm{s}_{\mathrm{D_{down}}}}(t_0)}\right). \tag{6.21}$$

In the previous equation, it is assumed that the measured mean speeds at both detector posts correspond to the traffic situation within the section; during congested conditions, this assumption can lead to an underestimation of the travel time if both detectors are still registering free-flow traffic.

In practice, the travel times are based on the consideration of all vehicle trajectories (see, e.g., Figure 2.3); to this end, it is necessary to estimate the mean speed of each trajectory. When performing this step, a possible approach is to estimate the trajectories' speeds as piece-wise linear functions. This latter can be done by sampling them at discrete distances using, e.g., single inductive loop detectors [Bov00].

Some examples of this type of travel time estimation are the work of Dailey who uses volume and occupancy measurements from SLDs, incorporating the stochastic nature of these quantities in order to derive the travel time; a Kalman filter is included to address the variability of the observations [Dai97], Petty et al. who adopt a similar methodology by using a stochastic model to accurately estimate the distribution of travel times, based on the measured upstream and downstream arrivals at SLDs [Pet98], Coifman who provides a methodology for estimating vehicle trajectories, and hence also link travel times, from the measurements at DLDs, based on a triangular (k,q) fundamental diagram [Coi02a], and Coifman and Cassidy who adopt a vehicle reidentification algorithm that matches vehicle measurements made at a downstream detector to those made at an upstream detector; as such, it is possible to derive the link travel time [Coi00; Coi02b].

6.3.2 Estimating travel times based on flow measurements

As previously noted, many travel time estimation techniques are based on the inverse of the measured mean speed. In our work, we are dealing with single inductive loop detectors, of which the mean speed estimations are notoriously known to be unreliable. However, as mentioned at the end of Section 6.1.1.4, the vehicle counts measured by these SLDs are quite reliable (for which we ignore the misclassification issues between cars and trucks, as only the total vehicle counts matter). As such, our methodology is based on the use of flows, i.e., the *cumulative curves* as explained in Section 2.3.2.2.

Briefly recapitulating, these cumulative curves represent the cumulative number of passing vehicles (denoted by N) with respect to time at different locations. Consider now a closed section of the road that conserves the number of vehicles (i.e., no on- or off-ramps); this section is demarcated by two measurement posts which measure its inflow, respectively outflow of vehicles. The result is a pair of monotonically increasing functions $N_{\text{up}}(t)$ and $N_{\text{down}}(t)$, which increase each time a vehicle passes by (see the left part of Figure 2.4 on page 24 for an example of these curves).

> The time needed to travel from one post to another can now easily be measured as the horizontal distance between the respective cumulative curves. Similarly, the vertical distance between these curves allows us to derive the accumulation of vehicles on the road section, which gives an excellent indication of growing and dissipating queues (i.e., congestion). Finally, at each time instant t the slope of this function corresponds to the flow $q(t)$.

In the remainder of this section we outline our methodology by first discussing how we extract the cumulative curves from the data, then how we deal with synchronisation issues and systematic errors, and finally how we estimate the distribution of the travel time.

6.3.2.1 Constructing the cumulative curves

The first step in the estimation of travel times, is to create the cumulative curves, by extracting the total flow measurements $q_{\text{up}}(t)$ and $q_{\text{down}}(t)$ from both upstream and downstream measurement posts, respectively. The underlying assumption we make is that all vehicles in the total inflow upstream to and outflow downstream from the section are accounted for by both measurement posts. Note that because we are dealing with multi-lane traffic, vehicles are allowed to overtake each other, and as such the *first-in, first-out* (FIFO) condition (see the end of Section 2.2.2) no longer holds [Dag95a]. However, this is not a significant problem because we are measuring the total flows at both detector posts.

Once the total flows are extracted, the cumulative curves $N_{\text{up}}(t)$ and $N_{\text{down}}(t)$ are constructed as follows[11]:

$$N(t) = \sum_{t'=t_0}^{t} q(t') = N(t-1) + q(t). \tag{6.22}$$

In this equation, $N(t)$ thus represents the number of the last vehicle that passed the measurement post at time t. Because the original flow measurements were aggregated at periods of one minute, the resulting $N(t)$ increases each time from $t \longrightarrow t+1$ with the total number of vehicles that passed the measurement post.

6.3.2.2 Dealing with synchronisation issues and systematic errors

When plotting both cumulative curves $N_{\text{up}}(t)$ and $N_{\text{down}}(t)$ in a single diagram, it is necessary to make them comparable. This entails the following two tasks:

[11] Note that we dropped the functional dependency on space, as for each cumulative curve its location is explicitly mentioned by the 'up' and 'down' subscripts.

- For starters, when both curves are reporting the same number of counted vehicles, they need to be *synchronised* with each other. To understand this, suppose a reference vehicle passes the upstream measurement post at a certain time instant t_{up}; after a certain time period Δt, the vehicle reaches the downstream measurement post at a later time instant t_{down}. The amount $\Delta t = t_{\text{down}} - t_{\text{up}}$ is the time it takes to cross the distance between both measurement posts, allowing the synchronisation mechanism to shift the respective cumulative curves over this time period (i.e., initialising them with the passing of the reference vehicle). As a result, both curves are synchronised such that $N_{\text{down}}(t_{\text{down}}) = N_{\text{up}}(t_{\text{up}} + \Delta t)$ for the reference vehicle (with Δt being the horizontal separation of both curves at time instant t_{up}).

- Secondly, all lane detectors in both measurement posts can be *differentially tuned*, or their measurements can be obtained by SLDs at one post and cameras at the other post. As a result, they can be systematically biased towards lower or higher vehicle counts (note that this is not related to the internal vehicle classification by the local controller).

Synchronising both cumulative curves
The first method we tried for synchronising both cumulative curves and obtaining Δt, was to estimate the experienced instantaneous travel time $\widetilde{T}_{\text{ff}}(t_{\text{up}})$ under free-flow conditions, according to equation (6.20):

$$\Delta t = \widetilde{T}_{\text{ff}}(t_{\text{up}}) = \frac{K}{\overline{v}_{\text{ff}}}, \tag{6.23}$$

with K the distance between both upstream and downstream measurement posts and $\overline{v}_{\text{ff}}$ the mean speed of all vehicles in free-flow traffic. However, due to the nature of the original flow measurements that are aggregated at periods of one minute, it is necessary that $\widetilde{T}_{\text{ff}}(t_{\text{up}}) \geq 1$ minute. A further complication is the fact that this travel time should be an integral value because the flow measurements are equidistant in time. As a consequence, the synchronisation methodology can only be applied to detector posts for which the travel time under free-flow conditions is at least one minute; e.g., when travelling at $\overline{v}_{\text{ff}} = 100$ km/h, then the minimal distance between both measurement posts should be (100 $\div$ 3.6) m/s $\times$ 60 seconds $\approx$ 1.6 km. Further problems arise due to the rounding of the travel time towards the nearest whole minute. As a result, it is possible that both cumulative curves intersect each other, which is physically impossible as this would mean that the total number of vehicles contained in the section becomes negative, which can even lead to negative travel times.

In order to tackle the previous problem, and to be able to work with measurement posts that are separated at arbitrary distances, we derived a new methodology. The key idea is to 'undo' the aggregation of the flow measurements. To this end, we convert the $N(t)$ curves to equivalent $t(N)$ curves. The conversion is done by constructing a new curve, such that all N vehicles at time instant t are distributed over the time interval $t \longrightarrow t + 1$. Note that the vehicles' detector passing times are now expressed in seconds (which are allowed to be fractional values). The original $N(t)$ curves were equidistant in time t, the corresponding $t(N)$ curves are now equidistant in the vehicle number N. Furthermore, whereas we assumed that $N_{\text{up}}(t) > N_{\text{down}}(t)$, we now assume that $t_{\text{up}}(N) < t_{\text{down}}(N)$.

At this stage, we have for each individual vehicle that passed both measurement posts, its corresponding t_{up} and t_{down} time instants. With this knowledge, we can now again compute $\widetilde{T}_{\text{ff}}(t_{\text{up}})$. Instead of using equation (6.23) which is explicitly based on the mean speed $\overline{v}_{\text{ff}}$ of all vehicles in free-flow traffic, we adopt another strategy. Abandoning the mean speed in free-flow traffic, we directly work with the 'history of the traffic stream', i.e., its characteristic properties that travel downstream. These undulations in both the $t_{\text{up}}(N)$ and $t_{\text{down}}(N)$ curves propagate freely and allow us to achieve a better and more stable synchronisation.

> One way to achieve this, is by looking at the respective shapes of both curves during light traffic conditions (e.g., the early morning period when free-flow conditions are prevailing). *The idea now is to shift one curve such that the difference between the two curves' shapes is minimal.* This approach for finding the correct time lag $\widetilde{T}_{\text{ff}}(t_{\text{up}})$ loosely corresponds to the concept of *platoon matching* where the characteristic features of groups of vehicles are compared and matched, as described in the Travel Time Data Collection Handbook of Turner et al. [Tur98]. Our method also corresponds to what Bovy and Thijs call *travel time estimation based on mass-balance*, where the inflow of a section is compared to its outflow [Bov00]. And finally, as explained by Muñoz and Daganzo, finding the minimal difference in shapes between the two cumulative curves is also analogous to the maximisation of the *cross-correlation* of both curves [Muñ03a].

In practice, we adjust the $t(N)$ curves, not by looking for the correct time difference Δt, but instead by finding the equivalent difference ΔN:

1. Choose a morning time t_{morning} with prevailing free-flow traffic.
2. Find the corresponding number of the first vehicle that passes the upstream measurement post during this morning time period, i.e., $N_{\text{up,morningBegin}}$.
3. Choose the length $N_{\text{morningLength}}$ (expressed in vehicles) of the morning time period (e.g., 400 vehicles, which corresponds to some 10 minutes on a three-lane motorway); this allows us to retrieve the number of the last vehicle that passes the upstream measurement post, i.e., $N_{\text{up,morningEnd}} = N_{\text{up,morningBegin}} + N_{\text{morningLength}} - 1$.
4. Setup a region in which we will synchronise the cumulative curves:

$$N_{\text{syncRegionBegin}} = N_{\text{up,morningBegin}} - \frac{N_{\text{syncRegionLength}}}{2}, \tag{6.24}$$

$$N_{\text{syncRegionEnd}} = N_{\text{up,morningEnd}} + \frac{N_{\text{syncRegionLength}}}{2}. \tag{6.25}$$

If we set $N_{\text{morningLength}} = 400$ vehicles, then we choose the length of the synchronisation region slightly higher, e.g., 500 vehicles. Note that these values should not be taken too large, as this will introduce a systematic error due to the possibly differential tuning of both upstream and downstream measurement posts.

5. Define the following objective function that represents the total squared difference of the travel time and the average time difference between both curves $t_{\text{up}}(N)$ and $t_{\text{down}}(N)$:

$$f(n) = \int_{i=1}^{N_{\text{morningLength}}} \left[t_{\text{down}}(n+i) - t_{\text{up}}(i) - \overline{\Delta t}(n)\right]^2 dN, \tag{6.26}$$

with the $\overline{\Delta t}(n)$ being the average time difference, defined as follows:

$$\overline{\Delta t}(n) = \frac{1}{N_{\text{morningLength}}} \sum_{i=1}^{N_{\text{morningLength}}} \left[t_{\text{down}}(n+i) - t_{\text{up}}(i)\right]. \tag{6.27}$$

Within the region $[N_{\text{syncRegionBegin}}, N_{\text{syncRegionEnd}}]$, we now search for the value of n_{min} at which the objective function is minimal (note that during the optimisation, we approximate the integral with a discrete sum over N):

$$\min_{n \in [N_{\text{syncRegionBegin}}, N_{\text{syncRegionEnd}}]} f(n). \tag{6.28}$$

In Figure 6.22 we graphically illustrate how the synchronisation mechanism works: the $t_{\text{down}}(N)$ curve is shifted with respect to the $t_{\text{up}}(N)$ curve within the synchronisation region until the objective function becomes minimal.

Finally, if we assume that all vehicles are driving with the same mean speed (i.e., the free-flow traffic regime), then the shapes of the curves are qualitatively the same. As such, the difference ΔN for which the two shapes differ the least is defined as follows:

$$\Delta N = N_{\text{up,morningBegin}} - (N_{\text{syncRegionBegin}} + n_{\text{min}} - 1). \tag{6.29}$$

The two curves are now synchronised, because $t_{\text{down}}(N) = t_{\text{up}}(N - \Delta N)$ holds for the selected morning period.

Correcting for systematic errors

When comparing cumulative curves from both upstream and downstream measurement posts, it is possible that they exhibit a bias towards under- or overcounting of vehicles, e.g., the downstream post consistently counts more vehicles than the upstream post. In order to compensate for these systematic errors, we apply the same synchronisation mechanism, but this time for an evening period with free-flow traffic.

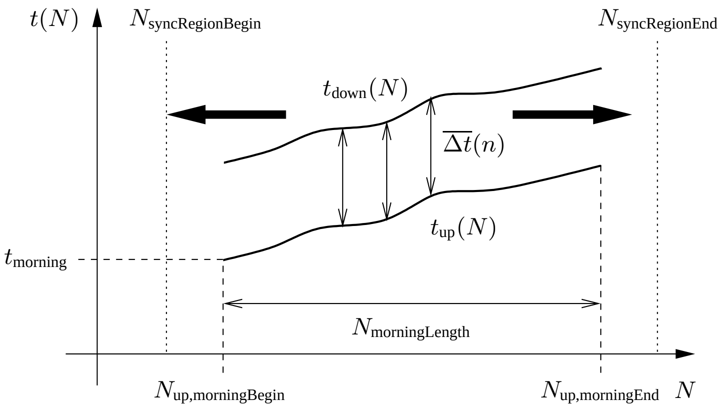


Figure 6.22: A graphical illustration of the synchronisation mechanism: the $t_{\text{down}}(N)$ curve is shifted with respect to the $t_{\text{up}}(N)$ curve within the synchronisation region until the objective function becomes minimal; $\overline{\Delta t}(n)$ represents the average time difference between the two curves. After the minimisation, the shapes of both curves are qualitatively the same.

After this step, we now have $\Delta N_{\text{morning}}$ and $\Delta N_{\text{evening}}$ available. In the case of bias, $\Delta N_{\text{morning}} \neq \Delta N_{\text{evening}}$, so the idea is to adjust the counts of $t_{\text{down}}(N)$ so that it counts the same number of vehicles as $t_{\text{up}}(N)$ does. There are two possibilities:

- $|t_{\text{down}}(N)| > |t_{\text{up}}(N)| \Longrightarrow$ we shorten $t_{\text{down}}(N)$ as follows:

$$\forall n \in |t_{\text{up}}(N)| : t_{\text{down}}(n) \leftarrow t_{\text{down}}(\left\lceil n \frac{|t_{\text{down}}(N)|}{|t_{\text{up}}(N)|} \right\rceil) \tag{6.30}$$

- $|t_{\text{down}}(N)| < |t_{\text{up}}(N)| \Longrightarrow$ we enlarge $t_{\text{down}}(N)$ by means of interpolation. Let us define the following index:

$$n_{\text{new}} = n \frac{|t_{\text{down}}(N)|}{|t_{\text{up}}(N)|}, \tag{6.31}$$

then the interpolation assigns the new indices as follows:

$$\begin{aligned} \forall n \in |t_{\text{up}}(N)| : t_{\text{down}}(n) \leftarrow t_{\text{down}}(\lfloor n_{\text{new}} \rfloor) + \\ (t_{\text{down}}(\lceil n_{\text{new}} \rceil) - t_{\text{down}}(\lfloor n_{\text{new}} \rfloor)) \cdot |\lceil n_{\text{new}} \rceil - n_{\text{new}}|. \end{aligned} \tag{6.32}$$

An underlying assumption in the previous derivation is that the measurement error of the posts occurs uniform in time, and is not related to the current traffic regime (e.g., free-flow versus congested traffic). Note that in a similar spirit, Muñoz and Daganzo correct for bias by explicitly multiplying the cumulative curves of each measurement post with a detector-specific correction factor, such that both posts count the same number of vehicles [Muñ03a].

6.3.2.3 Estimating the distribution of the travel time

After synchronising both upstream and downstream cumulative curves and correcting for systematic errors, the travel time is now equal to:

$$T(N) = t_{\text{down}}(N) - t_{\text{up}}(N). \tag{6.33}$$

It is also possible to derive the space-mean speed, which is in accordance with equation (2.27) equal to the ratio of the distance K between both measurement posts and the average experienced dynamic travel time $\overline{T}$ of all the vehicles:

$$\overline{v}_{\text{s}} = \frac{K}{\overline{\overline{T}}}. \tag{6.34}$$

Once all travel times are known for all vehicles during the measurement period, we can estimate the empirical distribution of the travel time by creating histograms (for more information, see the example in Section 2.4.3.3 at page 38).

6.3.3 Indicators of reliability

Considering the methodology elaborated upon in the previous section, we now give some examples of estimating the travel time for a real-world example. To this end, we first give an overview of the area that encompasses our case study. We then estimate the travel time for a motorway within this case study, indicating what the expected travel time will be, as well as its deviation. Finally, we construct several reliability maps that show for a complete motorway the locations of recurrent congestion, as well as the encountered fluctuations which give indications of stop-and-go traffic.

6.3.3.1 Overview of the case study area

The area of our case study encompasses what we call *"Flanders' Triangle"* (see also Figure 6.23); its corners are formed by three major cities, called Antwerpen, Gent, and Brussel (which is the capital of Belgium). Three bidirectional motorways are connecting these cities, i.e., the **E17** (having a length of approximately 2 $\times$ 54 km between Antwerpen and Gent), the **E40** (having a length of approximately 2 $\times$ 43 km between Gent and Brussel), and the **E19** (having a length of approximately 2 $\times$ 34 km between Brussel and Antwerpen). These motorways connect to the cities via ring roads; the **R0** around Brussel, the **R1** around Antwerpen, and the **R4** around Gent.

All motorways are composed of three lanes in each direction, with some exceptions such as the E19 which has a part where there are four lanes (separated in two) in each direction. The small grey filled circles represent the locations of the single loop detectors (a minority of the detectors are cameras). We collect traffic flow measurements for the year 2003, based on data stemming from 288 detectors (clockwise around the Triangle) and 264 detectors (counter-clockwise around the Triangle), which gives a grand total of 552 detectors.

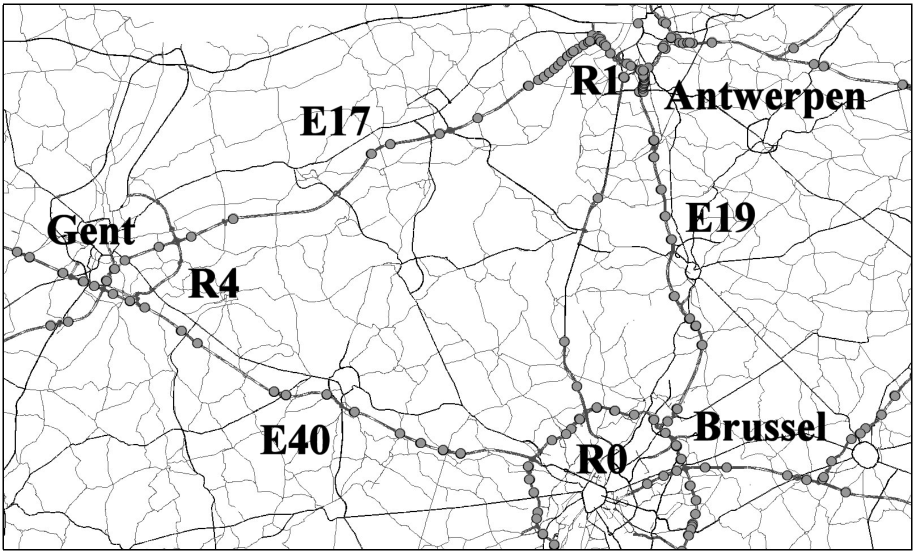


Figure 6.23: The area of our case study, encompassing what we call *"Flanders' Triangle"*; its corners are formed by the three major cities Antwerpen, Gent, and Brussel, which are connected by the three bidirectional motorways, i.e., the E17 between Antwerpen and Gent, the E40 between Gent and Brussel, and the E19 between Brussel and Antwerpen. These motorways connect to the cities via ring roads; the R0 around Brussel, the R1 around Antwerpen, and the R4 around Gent. Within our case study, we collect traffic flow measurements for the year 2003, based on data stemming from 552 detectors in total (represented by the grey filled circles).

6.3.3.2 Travel time reliability

We now apply our methodology for off-line travel time estimation as explained in Section 6.3.2, to a small real-world example. To this end, we take the following steps:

1. We chose two measurement posts that demarcate a road section with no intermediate on- or off-ramps (hence we have conservation of the number of vehicles). For convenience, we set all sentinel values of the flows, occupancies, and mean speeds equal to zero (see Section 6.1.2 for an explanation of these sentinel values in the database).

2. Next, the flows, occupancies, and mean speeds of all individual lanes at all measurement posts are aggregated together according to equations (2.13), (2.25), and (2.37), respectively. Note that we normalise the occupancy to obtain the average occupancy across all lanes, as described in Section 2.3.3.

3. We then collect all groups of similar weekdays (e.g., all Mondays, Tuesdays, ...), as described in Section 6.1.3.

4. For each individual day in each group of these weekdays, we consider all 60 $\times$ 24 = 1440 measurements for each lane detector. Using equation (6.22), we then construct the cumulative curves $N_{\text{up}}(t)$ and $N_{\text{down}}(t)$ for both measurement posts, convert them to equivalent $t_{\text{up}}(N)$ and $t_{\text{down}}(N)$ curves while synchronising them using the methodology explained in Section 6.3.2.2. Finally, we derive the travel times according to equation (6.33).

5. Based on these travel times, we construct histograms that represent the distributions of the travel times for the road section on the specified days, including the means and standard deviations of these travel times, as well as the 90% percentile. We also consider the medians and the median absolute deviations (MAD)[12]. Note that we include these latter robust estimators, in order to eliminate outliers such as special occasions, accidents and incidents, ...

6. Finally, all histograms belonging to the same day of the week (e.g., all Mondays) are averaged to obtain an estimate of the distribution of travel times on a typical day.

As a real-world example, we consider a three-lane section of the E40 motorway between Brussel and Gent (in the direction of Gent), located between the detector posts at Erpe-Mere and Wetteren; the distance K between these two measurement posts is 8.1 kilometres. For this section, we estimate the travel times on a typical Monday and a typical Friday in 2003, each time during the morning (08:00 – 10:00) and evening rush hours (17:00 – 20:00). The resulting histograms are shown in Figure 6.24; Table 6.3 summarises the resulting statistics.

Whereas the mean and the median tell us what the expected travel time on the section will be, the standard deviation and the median absolute deviation give a clue about the expected fluctuations around this travel time [War05]. The more reliable a road section is, the smaller the fluctuations will be. The histograms show that for a typical Friday, the distribution of the travel time is in general narrower than on a Monday. From the statistics, we can see that the mean and median values of the travel times lie close to each other; their deviations however show a difference. The reason for this is that our methodology is quite capable of estimating a qualitatively good travel time. Furthermore, there is no real difference between the expected travel time during the evening rush hours on a Monday and a Friday; there is however an increase in the deviation for a Monday.

With respect to the reliability of the travel time, we also note the evolution of the 90% and 95% percentiles; they indicate that a traveller can expect, with a probability of 90, respectively 95 percent, a travel time less than the percentiles. As can be seen in Table 6.3 and the histograms in Figure 6.24, the section has more robustness against congestion (both recurrent and non-recurrent due to incidents) on a Friday than on

[12] The median absolute deviation is defined as: MAD $=$ med $|x_i - \overline{x}_{\text{med}}|$, with $\overline{x}_{\text{med}}$ the median of the data. Note that we can scale the MAD to make it unbiased with respect to the normal distribution: $\text{MAD}_{\text{E}} = 1.4826 \times \text{MAD}$. The factor 1.4826 corresponds to the inverse of the 3rd quartile for the cumulative distribution function of the normal distribution, i.e., $1/\Phi^{-1}(\frac{3}{4})$.

a Monday. Furthermore, we can also spot an asymmetry between the morning and evening rush hours on a Friday. Further investigation of this phenomenon revealed that it was caused by some outliers (they could be caused by accidents, but this was not confirmed), which led to an increase in the calculated travel times. These outliers also cause the small the peaks in the histograms near higher travel times.

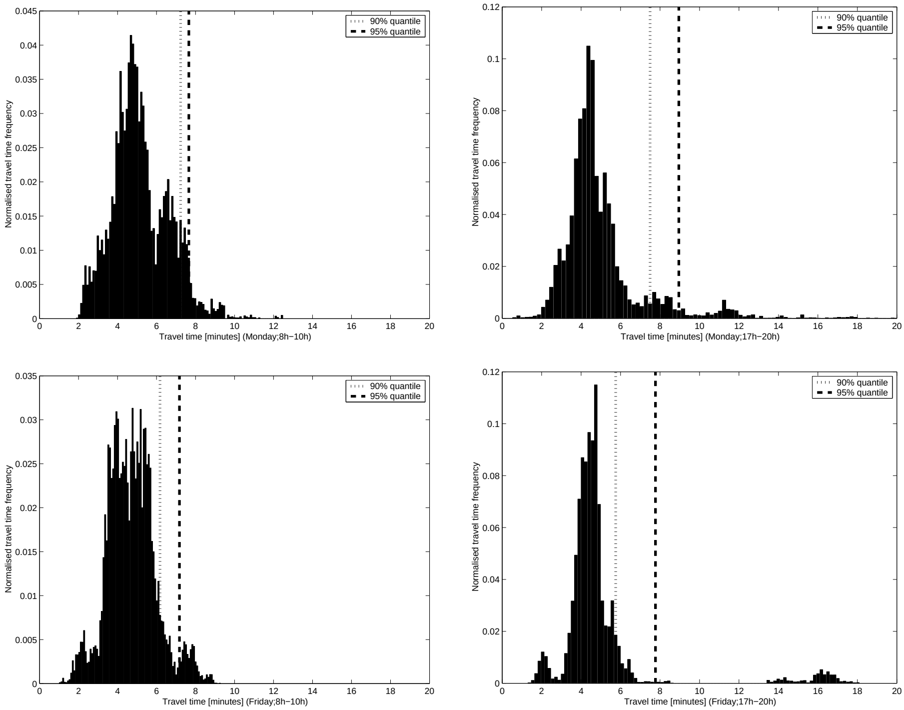


Figure 6.24: Histograms of the estimated travel time for the E40 motorway example from the case study for a typical Monday (*top*) and a typical Friday (*bottom*) in 2003, each time for the morning (*left*) and evening (*right*) periods. The dotted line indicates the 90% percentile, whereas the dashed line indicates the 95% percentile. Note how the distribution of the travel time is more narrow for a typical Friday then a typical Monday.

Period	*Mean*	*Median*	*Std. Dev.*	*MAD*	*90%*	*95%*
Monday morning	5.1734	5.1661	0.5390	0.4295	7.2335	7.6579
Monday evening	4.9795	4.9712	0.5631	0.4546	7.4946	8.9535
Friday morning	4.7165	4.6630	0.4815	0.4029	6.1798	7.1738
Friday evening	4.9540	4.9743	0.4037	0.3281	5.7519	5.7519

Table 6.3: The means and medians of the travel times, the corresponding standard deviations and median absolute deviations (MAD), as well as the 90% and 95% percentiles for the E40 motorway example from the case study.

6.3.3.3 Constructing reliability maps

Another aspect related to reliability of a transportation system, is expressed by on the one hand the expected mean speed a traveller will encounter, and on the other hand how the mean speed varies from one day of the week to another. To this end, we provide a methodology that allows us to construct time-space diagrams of all similar weekdays. As such, these containing 'typical days' clearly show the locations where *recurrent congestion* occurs. After explaining our technique, we apply it to some of our roads in the case study, i.e., the E19 between Antwerpen and Brussel and the R0 ring road around Brussel.

Our methodology encompasses the following steps:

1. For all motorways and ring roads in the case study, we consider all lane detectors at all measurement posts. For convenience, we set all sentinel values of the flows, occupancies, and mean speeds equal to zero (see Section 6.1.2 for an explanation of these sentinel values in the database).

2. We then collect all groups of similar weekdays (e.g., all Mondays, Tuesdays, ...), as described in Section 6.1.3.

3. For each individual day in each group of these weekdays, we consider all $60 \times 24 = 1440$ measurements for each lane detector; we then calculate the median over all days for each measurement separately, as well as the median absolute deviation (MAD). See also Figure 6.25 for a graphical depiction of this process. Note that we use these robust estimators instead of the classical mean, in order to eliminate outliers such as special occasions, accidents and incidents, ...

4. Next, the flows, occupancies, and mean speeds of all individual lanes at all measurement posts are aggregated together according to equations (2.13), (2.25), and (2.37), respectively. Note that we normalise the occupancy to obtain the average occupancy across all lanes, as described in Section 2.3.3.

5. Once all aggregated measurements are available, we collect them in time-space diagrams that each contain the medians and MADs of either the flows, the occupancies, or the speeds.

6. Finally, we apply Treiber and Helbing's tempo-spatial filter to the previously constructed time-space diagrams containing the medians; this filter reduces the noise in the measurements as it essentially is a low-pass filter (LPF), whilst interpolating locations and times between measurement posts in the time-space plane. The filter is also adaptive, in the sense that it keeps track of the current traffic conditions surrounding a measurement point: instead of just considering a point's nearest Euclidean neighbours, the filter takes into account the direction of information flow (i.e., forward moving in free-flow traffic and backward moving in congested traffic) [Tre02]. As a result, the traffic flow characteristics are more clearly pronounced, increasing the visibility of the resulting diagrams.

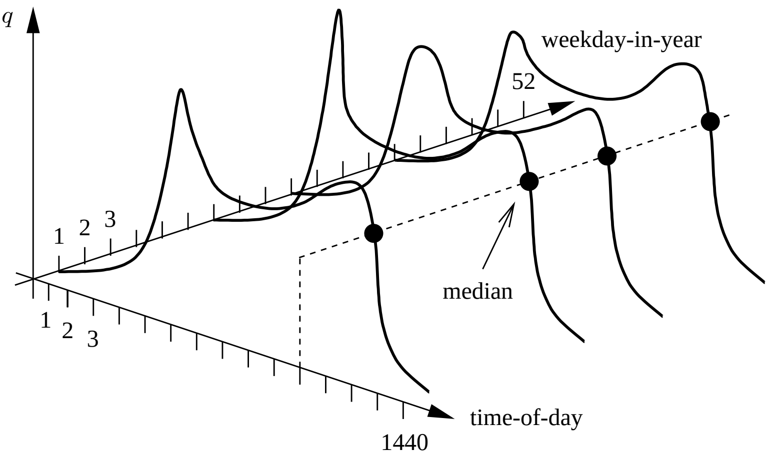


Figure 6.25: For each individual day in each group of similar weekdays (e.g., all Mondays), we consider all 60 × 24 = 1440 measurements for each lane detector; we then calculate the median over all days for each measurement separately, as well as the median absolute deviation (MAD). Note that the graph shows the calculation for one detector; we for purposes of illustration, we plot the daily profiles of four distinct weekdays.

Applying the previously explained methodology to some of our roads in the case study, we obtain the filtered time-space diagrams in Figure 6.26 and Figure 6.28 for the E19 between Antwerpen and Brussel and the R0 ring road around Brussel (extended with a part of the E19 between Brussel and Mons), respectively. In all maps, the driving direction is upwards, while time advances to the right. Note that due to the spatial sparseness of the detectors in the Flanders' region (they are only located right before and right after an on- and off-ramp), it is possible that we miss certain jams that originate and dissolve in a motorway section without ever being recorded by an upstream or downstream detector.

> In practice, a good indicator for congestion can be found in the occupancy. However, because the mean speed is tied uniquely to the density, and hence also to the occupancy, we use its median and variance as the main macroscopic characteristics in our reliability maps.

In these maps, dark spots represent regions where the mean speed is rather low, which is indicative of congestion: jams start at the top part of these structures, they grow downwards and recede upwards. Because the mean speed is considered over all similar weekdays, the maps of these 'typical days' clearly show the locations where *recurrent congestion* occurs. The information contained in these charts thus gives an indication of the expected mean speed a traveller will encounter at a certain moment and location in the time-space diagram of a motorway. Anyone who wants to avoid traffic congestion, should therefore try to stay clear of the dark spots; this can be accomplished by not entering the motorway and thus choosing an alternative route, and/or by selecting another departure time.

Another interesting quantity to consider is the variance, as plotted in the maps in Figure 6.27 and Figure 6.29. These maps represent how the mean speed varies from one day of the week to another; dark spots now represent regions where a traveller can expect strong fluctuations in the traffic pattern (note that no filtering was applied to these maps).

We now discuss the results of our case study; in Figures 6.26, 6.27, 6.28, and 6.29, the left parts each time relate to a typical Monday, whereas the right parts relate to a typical Friday.

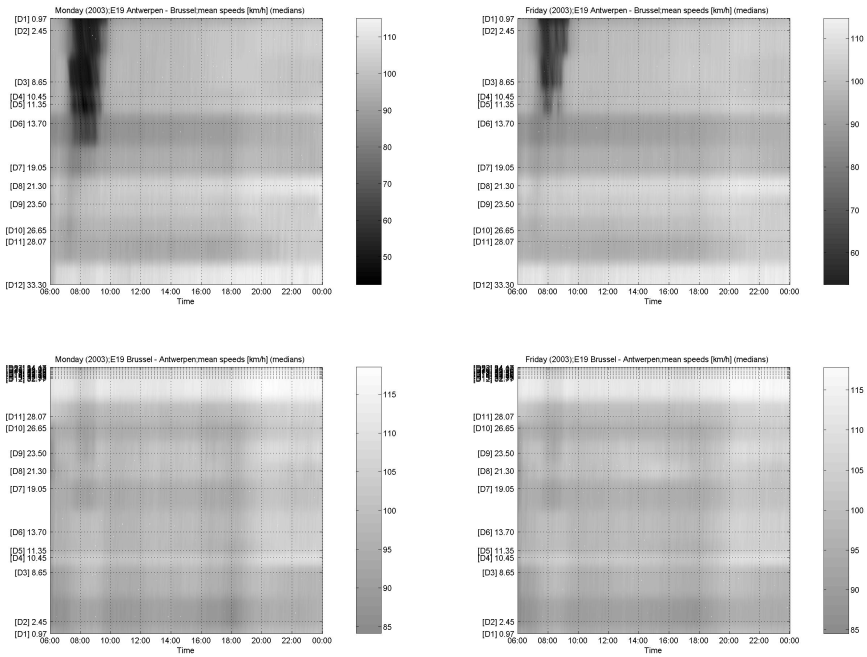


Figure 6.26: Time-space diagrams showing the evolution of the mean speed and recurrent congestion on a typical Monday (*left*) and a typical Friday (*right*) in 2003, for the E19 motorway in the direction of Brussel (*top*) and the direction of Antwerpen (*bottom*). Note the region of severe congestion during the Monday morning rush hour near Brussel (top parts) between posts D1 and D6; in contrast, the Friday morning rush hour is less pronounced. There is also a band of slower traffic between posts D6 and D7, where the complex near Mechelen, a major city located right next to the motorway, is located.

E19 motorway between Antwerpen and Brussel

- The morning rush hour when entering Brussel on a Monday (top-left part of Figure 6.26), exhibits severe congestion over a length of some 16 kilometres (ending between posts D6 and D7, located at the complex near Mechelen-Zuid and Mechelen-Noord), during from 07:00 until 09:30. For a Friday (top-right part of Figure 6.26), the length is reduced to some 13 kilometres (ending at post D6, located at Mechelen-Zuid). In both cases, the congestion starts at the connection of the E19 with the R0 ring road near Vilvoorde (post D1). Note that there is no evening rush hour for traffic in the direction of Brussel.

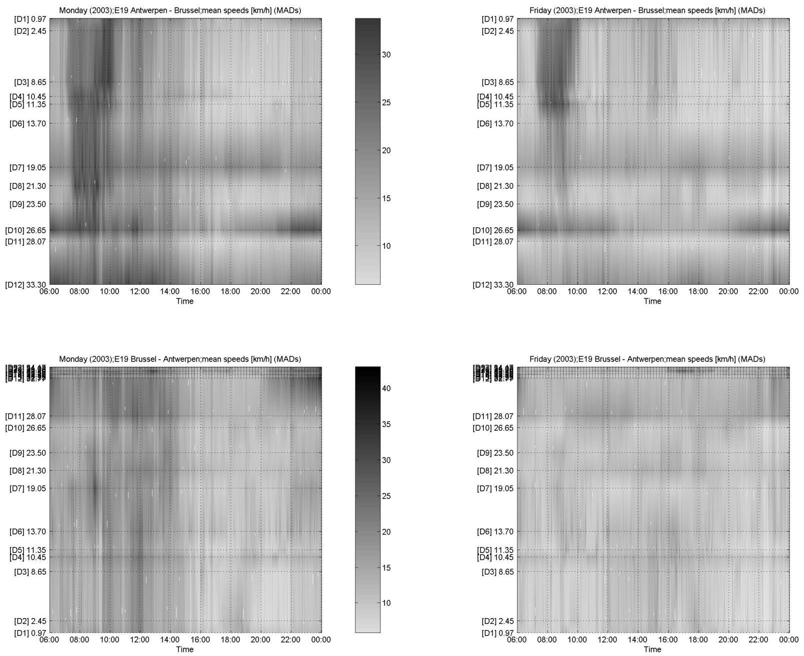


Figure 6.27: Time-space diagrams showing the evolution of the median absolute deviation (MAD) of the mean speed on a typical Monday (*left*) and a typical Friday (*right*) in 2003, for the E19 motorway in the direction of Brussel (*top*) and the direction of Antwerpen (*bottom*). Note how the fluctuations in the top parts have a magnitude of approximately 30 to 35 km/h, probably indicating heavy stop-and-go traffic during the morning rush hour.

Within this congested region, the mean speed drops to some 50 km/h; the fluctuations in the top-left and top-right parts of Figure 6.27 have a magnitude of approximately 30 to 35 km/h, which leads us to the conclusion that there probably is heavy stop-and-go traffic during the morning rush hour. Note the increased fluctuations near post D10, located at Kontich: right before this point, the E19 was split up into 2×2 lanes, separated by a verge. Post D10 lies right after an on- and off-ramp where the 2×2 lanes join together again.

Outside the congested region, the mean speed of traffic is approximately 90 km/h, with a small speed-up to some 100 km/h between posts D7 (Mechelen-Noord) and D10 (Kontich). The darker region between posts D6 and D7 represents traffic at the complex near Mechelen, which is a major city located right next to the motorway.

- Considering the E19 in the direction of Antwerpen (bottom-left and bottom-right parts of Figure 6.26), we note that there is no associated morning or evening rush hour present. Instead, we observe that the mean speed approximately lies between 90 and 100 km/h. A mild slowdown towards some 95 km/h is spotted from post D7 (Mechelen-Noord) on when entering Antwerpen.

  An interestingly observation is the light band in both parts: this band represents an increase in the mean speed to some 110 km/h. At this point, traffic enters the Craeybeckx tunnel which has four lanes; in this tunnel, traffic is neatly split up for the three outgoing directions towards the city centre of Antwerpen, the R1 ring road around Antwerpen in the direction towards Gent, and the R1 in the direction towards The Netherlands.

  Considering the bottom-left and bottom-right maps in Figure 6.27, we note that there are no clearly visible large fluctuations, except for a small increase from post D7 (Mechelen-Noord) on when entering Antwerpen, corresponding to the previously mentioned area where mild congestion occurs in the morning.

R0 ring road around Brussel

- During the morning rush hour, we can spot a region of severe congestion on the inner-ring road (i.e., the top-left and top-right parts of Figure 6.28) between posts D15 and D21; this corresponds to the region at the viaduct over Vilvoorde (see the satellite image in the top-right part of Figure 6.30), spanning the R0 near Strombeek-Bever until the detectors at Wemmel and Merchtem where the A12[13] between Antwerpen and Brussel joins the R0 (see the satellite image in the top-left part of Figure 6.30). The mean speed here lies around 30 km/h. The congested area covers some 8 kilometres, during a period between 07:00 and 10:00.

  Right above and below this congested location, there are two regions in which the mean speed is approximately 80 km/h; the upper region is located near post D14 (Machelen) where the E19 motorway between Antwerpen and Brussel joins the R0 (see the satellite image in the bottom-left part of Figure 6.30). The lower region is located near post D22 (Merchtem) where the E40 motorway between Gent and Brussel joins the R0. There is also a small region of slower traffic (60 km/h) between posts D31 (Beersel) and D34 (Huizingen), where the R0 blends into the E19 between Brussel and Mons in the Walloon region of Belgium.

[13]Note that the A12 road is considered as an alternative that lies parallel to the E19 motorway between Antwerpen and Brussel. However, it does not constitute a full-fledged motorway, as it contains traffic lights and road crossings at the part located in Antwerpen.

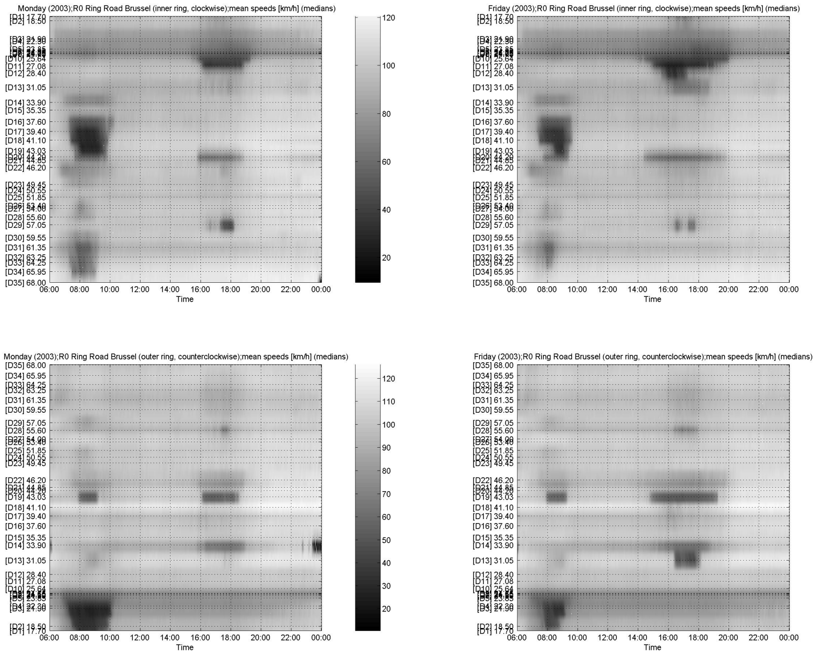


Figure 6.28: Time-space diagrams showing the evolution of the mean speed and recurrent congestion on a typical Monday (*left*) and a typical Friday (*right*) in 2003, for the R0 clockwise inner ring road around Brussel (*top*) and the counter-clockwise outer ring road (*bottom*). Note the area of severe congestion during the morning rush hour between posts D15 and D21 (i.e., the viaduct over Vilvoorde, spanning the R0 near Strombeek-Bever until the detectors at Wemmel and Merchtem where the A12 joins the R0), as well as the congestion during the evening rush hour between posts D3 and D12 (i.e., Vierarmenkruispunt, Tervuren, Wezembeek-Oppem near the junction with the E40 motorway in the direction of Luik).

The evening rush hour is clearly visible between 15:30 and 19:00; within it, the mean speed drops to some 40 km/h between posts D3 and D4 (Vierarmenkruispunt, see the satellite image in the bottom-right part of Figure 6.30), D5 and D10 (Tervuren), and D11 and D12 (Wezembeek-Oppem near the junction with the E40 motorway in the direction of Luik). Note that we can see a darker band in the maps between posts D2 and D9 (Vierarmenkruispunt and Tervuren) during the entire day, indicating slower traffic. Finally, there are also two small areas of congestion near posts D29 (Anderlecht) and D20 (Wemmel), where the mean speed drops to some 60 km/h.

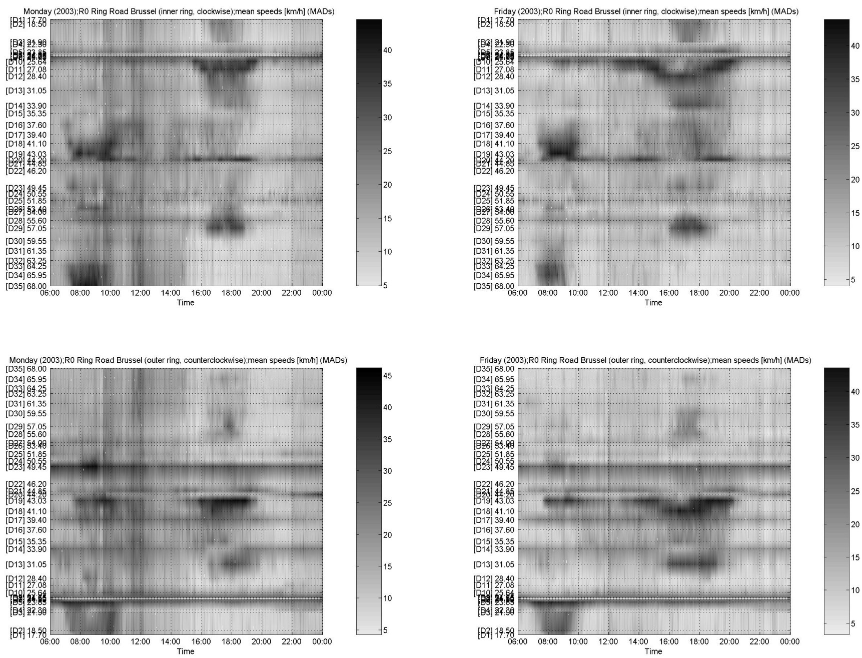


Figure 6.29: Time-space diagrams showing the evolution of the median absolute deviation (MAD) of the mean speed on a typical Monday (*left*) and a typical Friday (*right*) in 2003, for the R0 clockwise inner ring road around Brussel (*top*) and the counter-clockwise outer ring road (*bottom*). Note the fluctuations in the top parts near post D10; for Mondays, there appear to be more fluctuations in the morning period than for Fridays; conversely, the evening rush hour is more sensitive to disturbances on a Friday then on a Monday.

Comparing the left and right parts of the maps, we can see that in general the congestion is worse on a Friday then on a Monday; the evening rush hour previously discussed now already starts at the early time of 14:00, lasting until 20:00. Note that with respect to the mean speed, the congestion also results in low measurements of some 20 km/h. Furthermore, during the entire day, the darker band in the maps between posts D2 and D9 (Vierarmenkruispunt and Tervuren) is more pronounced, indicative of slower traffic (60 km/h).

Looking at the top-left and top-right parts of Figure 6.29, we note an important observation with respect to the reliability of the mean speed: on a Friday, there occur a lot of fluctuations near post D10 (Tervuren), starting already at 12:00, lasting until 15:00. They reprise at 18:00, lasting until 20:30. Considering both parts, we can also see that for Mondays, there appear to be more fluctuations in the morning period than for Fridays; conversely, the evening rush hour is more sensitive to disturbances on a Friday then on a Monday.

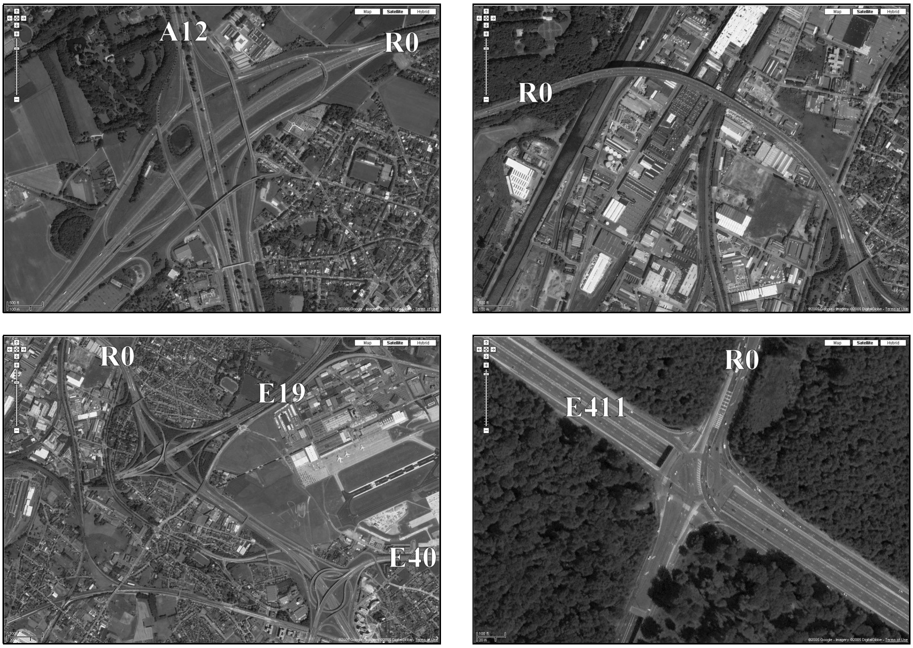


Figure 6.30: Some satellite images of the more severely congested locations on the R0 ring road around Brussel. *Top-left:* the junction between the A12 (between Antwerpen and Brussel) and the R0. *Top-Right:* the R0 on the large bridge over the city Vilvoorde and the Renault factory. *Bottom-left:* the junctions between the E19 (between Antwerpen and Brussel), the E40 (between Brussel and Luik) and the R0. *Bottom-Right:* the junction Vierarmenkruispunt (‘4-arms crossing’) between the E411 (between Brussel and Namen) and the R0 (all images reproduced after [Goo06]).

- Considering the outer-ring road (i.e., the bottom-left and bottom-right parts of Figure 6.28), we note that most congestion during the morning rush hour between 07:00 and 10:00 occurs between posts D1 (Groenendaal) and D9 (Tervuren), where the junction of the R0 and the E411 (between Brussel and Namen) at the Vierarmenkruispunt (‘4-arms crossing’) is located. Here, the mean speed drops to some 30 km/h, indicating severe congestion. The outer ring also exhibits a slower speed of some 50 km/h near post D19 (Strombeek-Bever). The same darker band of slower traffic is also visible between posts D2 and D9 (Vierarmenkruispunt and Tervuren) during the entire day.

  With respect to the evening rush hour, we can see some congestion (60 km/h) at post D19 (Strombeek-Bever), between 16:00 and 18:00 approximately. Also note the same level of congestion near post D14 (Machelen) where the E19 motorway between Antwerpen and Brussel joins the R0. Looking at the difference between Monday and Fridays, we can see that on a Friday, the morning rush hours has a smaller duration than on a Monday. The opposite is true for the evening rush hour, which now lasts from 15:00 until 19:00.

Looking at the fluctuations of the mean speed in the bottom-left and bottom-right parts of Figure 6.29, we can make the same remarks as for the inner-ring road.

Note that with respect to the locations of congestion, we find a close match between those discussed in our research, and the ones reported in the *"Belgium's Congestion Top-25"* of Logghe and Vanhove [Log04].

6.4 Conclusions

This chapter provided several techniques which can assist in the analysis of traffic flow measurements gathered on Flanders' motorways. We described how all these measurements are obtained and how they are stored in a central database. We then discussed the quality of the measurements, from a statistical point of view by giving a technique that tracks outliers. We also provided a methodology for quickly assessing structural and incidental detector malfunctioning, by means of creating maps that give a clear visual indication of when and where the problems occurred. We also gave clues as to which methods are suitable for dealing with missing values. Subsequently, we elaborated on a methodology for the off-line estimation of travel times, based on flow measurements (as opposed to the much used technique based on speed measurements). Finally, we gave some reliability and robustness properties related to travel times and traffic flow dynamics, which gives us an extra instrument for the analysis of recurrent congestion.

With respect to the tackling of congestion on Flanders' road network, the current policy is to keep the congestion on the motorways, avoiding its spreading towards the underlying secondary road network. At the moment, Flanders' government has a plan of action to investigate 25 'missing links' and bottlenecks in the road network (see the left part of Figure 6.31 for an overview). At the current rate of progress, the government hopes by 2009 to have accomplished one third of its goal [AWV06]. One important remark that needs to be made here, is the fact that the underlying principles for the identification of these 25 points of interest are based on not-so-rigidly defined measures such as accessibility, throughput, safety, viability, … [Des01]. A better method for selecting these points, is by applying a structured methodology that assesses their properties with respect to their reliability and response to incidents. A recent example in this latter direction is the work of Tampère et al.; they first create a list of candidate links that have a high probability of an incidident occurring on them, after which this list is shortened using criteria based on expert knowledge. In a final stage, they simulate incidents on the links in the list and compute the impact of each incident on, e.g., the travel time. An example of their selection of 12 vulnerable links for a case study on the E40 motorway corridor between Gent and Brussel can be seen in the right part of Figure 6.31 [Tam06].

In light of the *socio-economic use* of traffic data in Flanders, we note that at the moment, many information with respect to incidents is distributed by radio, mobile

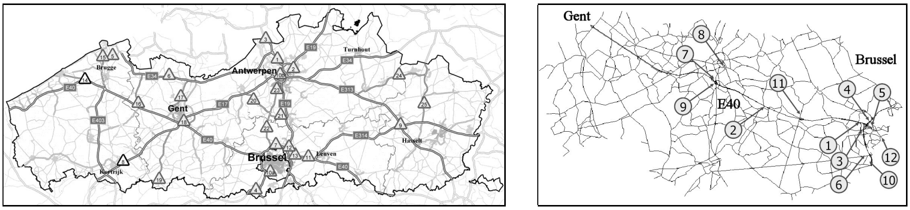


Figure 6.31: *Left:* an overview of the plan of action to investigate 25 'missing links' and bottlenecks in Flanders' road network; at the current rate of progress, the Flemish government hopes by 2009 to have accomplished one third of its goal. *Right:* a selection of 12 vulnerable links for a case study on the E40 motorway corridor between Gent and Brussel, based on the structured methodology of Tampère et al. (images reproduced after [AWV06] and [Tam06]).

phone, Internet, ... to the individual road traveller. However, most of the information has a low quality, especially when it comes to structural congestion: everybody already *knows* where the congestion occurs (and in the case of incidental congestion, there is often a large time lag involved). In this respect, there is a clear market available for supplying qualitative traffic information towards the road travellers; this is especially useful for individual people that have a high value-of-time.

> Providing useful information to the road traveller is of paramount importance. With respect to important quantities such as travel times that get advertised on VMS/DRIPS systems, the agencies and corporations engaged in this business, should therefore not neglect the fact that the human interpretation of probability most likely fails. People tend to assign too high probabilities to extreme events (e.g., incidents). Furthermore, it is also important to mention the cause of a delay, as humans want to know *why* they are entering congestion; this fact should not be neglected as it is important for a driver's perception and peace of mind. Given a good enough reason, people get less annoyed by non-recurrent congestion; especially when they are supposed to take an alternative route in order to reach their destination.

Part IV

Integrated Dynamic Traffic Assignment

Chapter 7

Dynamic traffic assignment based on cellular automata

Within the framework of transportation demand modelling, we are left with three major approaches, being trip-based, activity-based, and equilibrium-based as put forward by Boyce (see also Sections 3.1.2 and 3.1.3). This diversification in the scientific field is a clear sign that different techniques are considered, based on distinct ideas. All techniques nevertheless borrow certain elements from one another, implying some generality between the models. As such, a travel forecasting model will most certainly be a give-and-take between requirements/desires and the current state-of-the-art [Boy98].

Looking at the structure behind these methodologies, it is known that a core component in each of them is the concept of traffic assignment [Boy04a]. In this part of the dissertation, we propose a method for performing dynamic traffic assignment, whereby we integrate departure time choice (leading to the phenomenon of peak spreading) and dynamic route choice, coupled with a dynamic network loading model. The method is built around a traffic flow model that is represented as a computationally efficient cellular automaton. The chapter ends with a brief overview of some possible applications.

7.1 Integrated dynamic traffic assignment

As already mentioned in Section 3.1.2.2.IV, the four step model of transportation demand modelling, contains a step called traffic assignment, in which all traffic demand is assigned to the network: the routes vehicles will follow are calculated, such that the load on the road network is evenly distributed over all links. This distribution is governed by Wardrop's criterion, i.e., the user equilibrium W1 which states that *"the* *journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route".*

With respect to the assignment of vehicles to the available routes, there are mainly two different approaches that can be followed: static (STA) or dynamic traffic assignment (DTA), as explained in Sections 3.1.2.3 and 3.1.2.4, respectively. When discussing the DTA approach, we hitherto mainly focused on the route choice and dynamic network loading components. In this section, we introduce *integrated dynamic traffic assignment*, by which we mean that our modelling framework consists of three components:

- departure time choice (DTC),
- dynamic route choice (DRC),
- and dynamic network loading (DNL).

The first part of this section discusses two approaches towards DTA, namely analytical DTA and simulation-based DTA. In a second part, we explain our modelling framework that encapsulates the previously mentioned three components. The DNL component is expanded upon separately in Section 7.2.

7.1.1 Approaches to dynamic traffic assignment

As noted in Section 3.1.2.4, it is important to capture the temporal character of congestion: the buildup and dissolution play an important role. As such, travel times depend on the history of the system, which should not be neglected. The 'dynamic' part in DTA refers to this dependency, implying that two fundamental components are considered, i.e., *route choice* and *dynamic network loading*. The former calculates the routes that vehicles take, by assuming that an equilibrium condition holds; the latter is responsible for loading all the vehicles onto the network, by explicitly simulating the physical propagation of the time-varying traffic flows[1]. The last two decades, a third fundamental component is considered, i.e., *departure time choice*, which is typically associated with the route choice behaviour: the choice of which route to take now becomes time-dependent.

In this section, we briefly describe the two main approaches towards DTA, being analytical DTA and simulation-based DTA. Note that some of the models described in this overview constitute more than a simple DTA procedure; they are embodied as complete travel forecasting models.

[1]The DNL component in a DTA approach has also received special attention, e.g., in the work of Wu et al. who discuss it as a system of functional equations [Wu98]. In this sense, the DNL model can be considered as a within-day model, as explained by Cascetta and Cantarella [Cas91].

7.1.1.1 Analytical dynamic traffic assignment

The attractiveness of analytical DTA is that it can build on a long line of previous research, starting in the 1950s with the work of Beckmann et al. [Bec55] (see also the introduction of Section 3.1.2, as well as Section 3.1.2.3 for more information regarding the implications of their work). According to the overview of Joksimovic and Bliemer, we can distinguish between three categories of analytical DTA models[2] [Jok02]:

- **Mathematical programming**
  The root of this technique can be traced back to the seminal work of Beckmann et al. in 1955, who reformulated the Wardrop equilibrium as a convex optimisation problem [Bec55] (see also Section 3.1.2.3 for more information). Closely related to this work, is the approach taken in static traffic assignment; more information can be found in the work of Patriksson [Pat94]. A major drawback of mathematical programming, is that it is not able to fully capture the time-varying interactions of traffic flows and travel times.

- **Optimal control theory**
  Around 1989, a number of researchers proposed to describe the temporal evolution of traffic flows in the framework of optimal control theory (OCT), which deals with dynamic systems. Due to some of the severe limitations of this approach, the OCT research in the context of traffic flow modelling has not received much attention nowadays. For an overview of the OTC approach, we refer the reader to the work of Peeta and Ziliaskopoulos [Pee01].

- **Variational inequalities**
  The most promising approach to tackle analytical DTA models, is based on a variational inequality (VI), which allows a natural incorporation of flow propagation constraints that explicitly contain link travel times. From the early nineties on, the VI technique has been applied to the DTA problem, fuelled by its ability to include certain measures in the models, e.g., the integration of road pricing policies. More information can be found in the works of Nagurney [Nag93c] and Patriksson [Pat99].

  Some recent examples of this class of analytical DTA models include the work of Lo and Szeto who propose a formulation that is based on a dynamic user equilibrium and an encapsulation of the cell-transmission model[3] (CTM) [Lo99; Lo02], the work of Bliemer et al. who propose a multi-class DTA model (i.e., different vehicle types) called *INteractive DYnamic traffic assignment* (INDY) [Bli01; Bli; Mal03; Bli04], and the work of Liu et al. who introduce a probabilistic framework for travel times with random perception errors [Liu02].

[2] More information can be found in the overviews of Bliemer [Bli01], Boyce et al. [Boy01], and Peeta and Ziliaskopoulos [Pee01].

[3] See Section 3.2.1.4 for more details regarding the LWR first-order macroscopic model and one of its numerical schemes, i.e., the CTM.

Three of the more appealing aspects of this modelling class, are (i) the fact that due to their analytical nature, the solutions can be readily evaluated, (ii) the existence of well-defined algorithms that can lead to stable convergence, and (iii) their construction implies that computational complexity does not increase with a raise in traffic demand. A drawback however, is the fact that with respect to the current state-of-the-art, these analytical DTA models are not capable of handling large-scale road networks due to the computational complexity of the models themselves. Notwithstanding this critique, the aforementioned work by Bliemer et al. seems promising in this direction [Bli01; Bli].

7.1.1.2 Simulation-based dynamic traffic assignment

In order to address the problems associated with the application of analytical DTA techniques to large-scale road networks, simulation-based DTA can provide a solution. The iterative approach taken by this model class can largely be stated as follows:

1. Traffic demand is specified to a dynamic route choice model.
2. Based on the current travel times, the dynamic route choice model is executed, specifying the routes for all vehicles.
3. The dynamic network loading model is executed (note that the DNL model can be micro-, meso-, or macroscopic in nature).
4. The experienced travel times on all routes are extracted from the last simulation, and fed back to the route choice model in step 2. The iterations terminate when the algorithm converges.

Whereas the analytical expression of a DNL model is in general too cumbersome to include detailed traffic operations such as traffic lights, certain control measures, . . . , the simulation-based DTA methodology can easily incorporate these effects (e.g., in a microscopic model). And although it can deal with large networks, there exists a subtle but major caveat, i.e., it uses a heuristic approach with respect to convergence. To state it more clearly, *convergence is not guaranteed*. An often employed method is a *relaxation procedure* that performs iterations until all travel times in the road network are stationary [Gaw97; Gaw98a; Gaw98b; Bar02b]. In this context, because the concept of mathematical convergence is too strict, researchers then refer to the terminology of ‘relaxation’. Note that the question as to whether or not the simulation relaxes, or when it does, if it then relaxes to a unique equilibrium or instead exhibits oscillations, still remains an open debate; in some worst cases, even a gridlock situation can occur [Ric97; Nag98a; Nag98b].

Some examples of simulation-based DTA models include the following:

- Microscopic models: AIMSUN2 [Bar02a; Bar02b], MITSIM [Yan97], Paramics [Cam94; Lim00], TRANSIMS [Nag98c], VISSIM [PTV05]; see also Section 3.2.3.6.
- Mesoscopic models: CONTRAM [Tay03], DynaMIT [BA96; BA98; Sun02], DynaSMART [DYN03], METROPOLIS [Pal02; Mar03]; see also Section 3.2.2.1.
- Macroscopic models: METANET [Mes90]; see also the beginning of Section 3.2.1.7.

7.1.2 Integrated dynamic traffic assignment

As already highlighted in the introduction of Section 7.1, our transportation demand modelling framework encapsulates three components, i.e., departure time choice (DTC), dynamic route choice (DRC), and dynamic network loading (DNL). Our framework is constructed with the following points in mind: we would like to be able to incorporate a given synthetic population on a rather low level (i.e., represented as individual drivers or agents). Furthermore, traffic is considered unimodal, in the sense that each agent only uses one mode of travel, and does not change modes (i.e., an agent's itinery does not constitute trip legs by means of different transportation types). The agents themselves are not strictly considered as intelligent particles, by which we mean that an agent in general does not change his/her route choice in an en-route fashion (although we allow for the possibility of en-route choice, it is not explicitly taken into account in this dissertation). The goal of our DTA framework is to obtain an equilibrium on the road network, with respect to the choice of the departure time and the choice of the routes, and this over one time period that encompasses either the morning or evening rush hour.

In the first part of this section, we give an overview of the framework, the second part then goes into more detail with respect to the modelling of traffic demand, after which the following two parts describe the DTC and DRC components. The section ends with some remarks on convergence criteria.

With respect to the issue of solving the combined DTC-DRC problem, we briefly mention some of the existing research approaches in literature: Yang and Meng consider optimal pricing strategies (i.e., a congestion toll) by solving a system optimisation problem that combines the DTC and DRC models, as well as the optimal tolls of bottlenecks [Yan98b], Stada et al. use the existing concept of ‘shadow networks’ to combine on-peak and off-peak travel (the latter occurs in the shadow network) by assuming a fixed transfer penalty between both time periods [Sta01], de Palma and Marchal present the METROPOLIS toolbox who directly use the DTC model to assign a static OD matrix to a commuter period [Pal02; Mar03], Ettema et al. describe a microscopic traffic flow model in which drivers base their departure time and route choice decisions on a mental model of the traffic conditions (i.e., the mean and variances of the experienced travel times) [Ett03], Lago and Daganzo describe a departure-time user equilibrium model, combining Vickrey’s DTC work (see also Section 3.1.4.2) and the LWR first-order macroscopic model (see also Section 3.2.1.2) [Lag03a; Lag03b], Szeto and Lo propose a formulation based on the cell-transmission model through a variational inequality problem (see also Section 7.1.1.1) [Sze04], Lim and Heydecker identify a new equilibrium condition for the combined DTC-DRC problem, providing a computationally efficient solution algorithm that is based on the concept of reasonable paths [Lim05], and finally Yperman et al. construct a dynamic network loading model based on cumulative curves; they couple this model with Vickrey’s bottleneck model to determine an optimal road pricing policy by analytically solving the combined DTC-DRC problem [Ype05a; Ype05b].

7.1.2.1 Overview of the framework

Our approach towards dynamic traffic assignment is simulation-based; it combines a departure time choice (DTC) model with a dynamic route choice (DRC) model, and incorporates a dynamic network loading (DNL) model based on an efficient traffic cellular automaton. A flow chart of the framework is given in Figure 7.1. For simulation-based DTA, integrating them into a single iterative loop is quite difficult, as the departure times are typically considered for known routes, and these routes are in turn based on known departure times. Consequently, we combine the DTC and DRC models in a *sequential* manner, as can be seen from the chart.

In general, we can describe our modelling technique as follows:

1. Based on the information contained in a static OD matrix that is valid for the time period under consideration (e.g., the morning rush hour), we create a population of N agents by disaggregating the entire OD matrix.

2. Based on the description of the road network, and on the information from the static OD matrix, we generate a set of routes that connect the origins and the destinations. This set actually denotes *feasible routes*, as not all paths between an origin and a destination should be considered (doing this selection beforehand avoids extra complications in the DRC model).

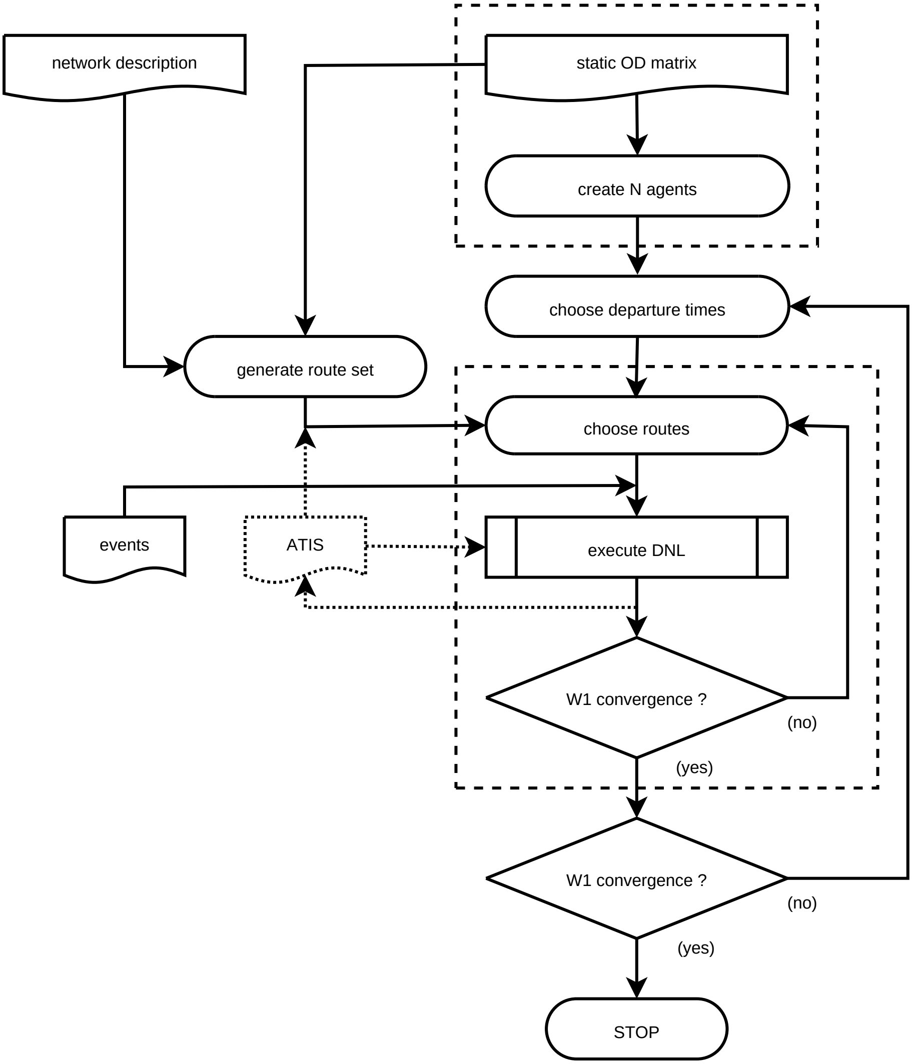


Figure 7.1: A flow chart of our framework for simulation-based dynamic traffic assignment (DTA); it combines a departure time choice (DTC) model with a dynamic route choice (DRC) model, and incorporates a dynamic network loading (DNL) model based on an efficient traffic cellular automaton. The part containing the module for advanced traveller information systems (ATIS) is optional. The upper dashed block represents the disaggregation of the traffic demand into individual agents (i.e., commuters), the lower dashed block denotes the DRC model.

3. All agents' departure times are re-calculated according to the DTC model (they are uniformly distributed over the entire time period, before the first iteration, thus corresponding to the assumption of a constant flow).

4. Using the information contained in the route set, and the current experienced travel times on the links in the road network, all agents get a route assigned, corresponding to their current shortest path in the network (shortest in the sense of minimal cost expressed as the travel time).

5. All agents are then put into queues located at the origins of the road network, after which the DNL model is executed until all of them have reached their destination (events, such as lane closures to simulate an accident, are modelled as well). Note that advanced traveller information systems (ATIS) can influence the route choice model, as well as the DNL model; ATIS can also draw information from the last run of the DNL model.

6. A convergence criterion is checked, corresponding to a *W1 user equilibrium* as explained in Section 3.1.2.2.IV; the associated cost for each agent is its total travel time. If convergence is not reached, we return to step 4 and recalculate the new shortest routes, based on the last experienced total travel times after execution of the DNL (their departure times remain fixed during subsequent iterations).

7. If the previous step resulted in sufficient convergence with respect to the routing of the vehicles, another convergence criterion is checked, but this time with respect to the agents' departure times. In this step, schedule delay costs are taken into account for all agents, thereby including effects of arriving too early or too late at their destinations. If the algorithm does not converge, we return to step 3.

Note that as opposed to analytical DTA models for the combined DTC and DRC problem (see Section 7.1.2), we adopt two different convergence criteria in this framework. In the former they are simultaneously combined, whereas we work with a *sequential procedure*[4]. The modified *W1 user equilibrium* (see Section 3.1.2.2.IV) as used in step 7 when checking the convergence, is based on a generalised travel cost. It corresponds to the following formulation: "*The generalised travel costs for all agents are equal and less than those which would be experienced by a single agent departing at a different time.*"

In the following sections, we discuss the modelling of traffic demand, after which we describe the DTC and DRC components, concluding with some remarks on the convergence of simulation-based DTA.

[4]In our framework, a traveller wishing to undertake a journey is faced with two options: at what time does he/she depart, and which route will he/she take ? Different possibilities exist for tackling this problem; one example is the approach taken by the DynaSMART model [DYN03], in which a commuter can select between changing his/her departure time, changing the route, changing both, or remaining with his/her current choice; the choice is based on a multinomial logit model.

7.1.2.2 Traffic demand generation

In most travel forecasting models, traffic demand is typically expressed as one or more origin-destination (OD) matrices (see also Section 3.1.2.2.III for more information). Within the classical four step model, it is assumed that steps (I) to (III) result in an OD matrix that is fed to the fourth step, i.e., the traffic assignment which calculates the routes vehicles will take.

The estimation of these OD matrices, forms a theme on its own; a typically encountered problem is the fact that due to the large amount of unknown variables (it is a considerably underdetermined system of equations), additional constraints need to be introduced. Besides the OD matrix estimation techniques explained in Section 3.1.2.2.III, we also mention some other methodologies, e.g., the doctoral dissertation of Bierlaire, which provides a nice overview of several different OD matrix estimation techniques [Bie95], Abrahamsson who gives a detailed literature survey of OD estimation based on traffic counts [Abr98], and the work of Balakrishna who provides a detailed methodology for the joint calibration of OD matrix estimation and route choice models within the DynaSMART model [Bal02; DYN03].

Another approach for traffic demand specification builds upon the activity-based modelling approach (see also Section 3.1.3 for more details). Examples in this direction include the work of Balmer et al. who develop a methodology for creating individual demand (i.e., tailored towards individual commuters) out of a general specification; they allow for the construction of agents' plans out of general OD matrix data [Bal05; Bal06], the work of McNally who integrated household activities, land-use distributions, regional demographics, ... into a microscopic model for traffic demand forecasting [McN96]. A recent approach closely related to ours, is the one by Kemper who disaggregates one static OD matrix into individual commuters, whilst taking into account traffic flow profiles over time. However, his model currently does not support departure time choice [Kem; Kem05].

Within our framework, we start from a known *single static OD matrix* that captures an entire time period (e.g., the morning rush hour). We furthermore assume *unimodal traffic* in the network, but we allow for a distinction between cars and trucks. From this single OD matrix we create a population of N agents by disaggregation (note that by 'plans' we mean an agent's route end points, but not the actual path of the route taken). As can be seen in Figure 7.2, our approach also borrows ideas from the activity-based modelling approach, such that — as opposed to the paradigm of a single static OD matrix — it is flexible enough to incorporate historical OD matrices, an explicit synthetic population of households, land-use data, ... Note that if necessary, it is possible to recreate a single OD matrix from the generated agents.

7.1.2.3 Departure time choice (DTC)

Within our framework we explicitly provide *time-of-day modelling* in the form of a departure time choice model. We assume the single static OD matrix specifies the

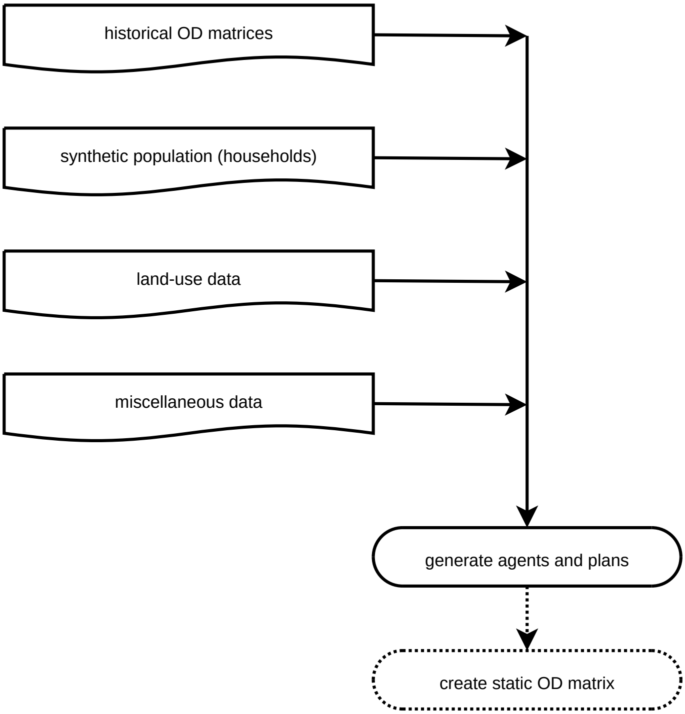


Figure 7.2: An overview of our flexible approach that borrows ideas from activity-based modelling. As opposed to the paradigm of a single static OD matrix, it can incorporate historical OD matrices, an explicit synthetic population of households, land-use data, … If necessary, it is possible to recreate a single OD matrix from the generated agents.

traffic demand over a given time period $[t_{\text{demand,start}}, t_{\text{demand,end}}]$, e.g., between 07:00 and 09:00. As such, it represents the total number of commuters that *want to depart* during the specified time period. For this latter time period, it is important that it completely encapsulates the period of heavy congestion (otherwise, some exogeneously given boundary conditions are needed); it is also preferred that all commuters can reach their destinations within the time period.

At the first iteration of the DTC model's execution, all agents' departure times are uniformly distributed over the entire time period (we thus assume a constant flow); the departure time for the i^{th} agent is given as:

$$t_{\text{departure}_i} = t_{\text{demand,start}} + (i-1)\,\frac{t_{\text{demand,end}} - t_{\text{demand,start}}}{N}. \tag{7.1}$$

Once the DNL model has been executed and the agents have reached their destinations (assuming an equilibrium situation was achieved according to step 6 of the framework in Section 7.1.2.1), we can calculate the general costs experienced by each individual traveller. At this point, we know the departure times $t_{\text{departure}_i}$ of all the agents, as well as their waiting times $\mu_i(t_{\text{departure}_i})$ in the queues at the origins, and their travel times $T_i(t_{\text{departure}_i})$. With the latter two of these quantities, we associate certain costs, resulting in C_{μ_i} and C_{T_i}. With respect to the departure time, we incorporate so-called *schedule delay costs* C_{sd_i}, as used in Vickrey's bottleneck model mentioned in Section 3.1.4.2. In this model[5], each agent is assumed to have a certain *preferred arrival time* (PAT) t_{PAT_i}; extra costs are incurred if the agent arrives too early or too late [Vic69]. Combining all these costs, the *generalised travel cost* for each agent is then defined as:

$$
\begin{aligned}
C_{\text{total}_i}(t_{\text{departure}_i}) \quad = \quad & C_{\mu_i}(\mu_i(t_{\text{departure}_i})) + \\
& C_{\text{T}_i}(T_i(t_{\text{departure}_i})) + \\
& \max\{C_{\beta_i}(t_{\text{PAT}_i} - (t_{\text{departure}_i} + T_i(t_{\text{departure}_i}))), 0\} + \\
& \max\{C_{\gamma_i}(t_{\text{departure}_i} + T_i(t_{\text{departure}_i}) - t_{\text{PAT}_i}), 0\}, \qquad (7.2)
\end{aligned}
$$

with $C_{\beta_i} < C_{\text{T}_i} < C_{\gamma_i}$ the costs for arriving too early or too late, respectively[6]; as such, arriving too late carries a higher weight than arriving too early.

Note that it is also possible to specify a cost associated with an agent's departure time; this cost can for instance be represented as a *utility function* $U_{\text{departure}}(t)$, which assigns a 'score' to each possible time instant for departure (e.g., some people do not want to leave too early). As such, the range of the time period in which a commuter wishes to depart is demarcated, such that utility maximisation of the individual results in a chosen departure time [Bat01].

Our DTC model then selects the 10% agents with the highest costs and shifts their departure times towards those associated with the lowest costs (we reserve an *indifference band* for small shifts on the order of five minutes). In terms of Wardrop, this corresponds to a *W1 user equilibrium* as explained in Section 3.1.2.2.IV, but now with respect to the agents' departure times: "*The generalised travel costs for all agents are equal and less than those which would be experienced by a single agent departuring at a different time.*" At this step, it is important to take into account certain constraints, e.g., placing a restriction on the maximum number of agents that can depart in a certain time period (as the absolute inflow into the network is bounded by the capacity at that point).

Let us finally mention that there are other approaches towards integrated DTC modelling, e.g., the model of Levinson and Kumar, which constitutes an implementation of the four step model. They extend it with a DTC model that is based on a binomial logit model, whereby commuters can choose between travelling in the peak hour or

[5] In the period after Vickrey, Hendrickson and Kocur were among the first to consider DTC in combination with a deterministic queueing model in a user equilibrium setting [Hen81]. A while later, Small developed a DTC procedure that is based on a generalisation of the multinomial logit model [Sma82; Sma87].

[6] The symbols β and γ stem from Vickrey's original formulation; in his work, he denoted the cost associated with the travel time as α.

in the shoulder hours of the peak period (they use a static user equilibrium method for solving the DRC component) [Lev93]. Discrete choice modelling is a popular approach for dealing with these kinds of decisions in a DTA context [BA85; BA99]. All techniques have one thing in common: when the projected demand exceeds the capacity, this will lead to the phenomenon of *peak spreading*. Commuters change their departure times in order to accommodate for the excess demand during the assignment procedure.

> Note that the previously discussed algorithm is based on a *single static OD matrix* that encapsulates, e.g., the morning rush hour. Another approach would be to use multiple exogeneously given *time-dependent/dynamic OD matrices* (see also Section 3.1.2.2.II). Each of these latter non-overlapping matrices covers, e.g., one half hour. As the DNL simulation is executed, all traffic demand matrices are to be consecutively assigned to the network. When performing this step, agents' departure times can be shifted *within* one matrix's time period, but also *between* subsequent matrices' time periods. In this case, it is important to put a constraint on the maximum number of vehicles transferred between such OD periods; e.g., it is impossible to allow a million agents to start at the same time instant at a single origin.

7.1.2.4 Dynamic route choice (DRC)

With respect to the routing of vehicles through the network, there are two approaches possible, i.e.:

- **Pre-route assignment**
  This method is also known as *equilibrium assignment*, because each driver now tends to minimise his/her travel cost; it is built on the assumption that travellers have *perfect information* of the *experienced travel times* (corresponding to *day-to-day learning* of the traffic pattern on a typical day, given the current traffic demand [Cas91]), such that their travel time costs are fixed and known at the time of departure. Due to its equilibrium nature, iterations are required until convergence is met.

- **En-route assignment**
  As opposed to the former method, this one is based on the *instantaneous travel times*, and allows vehicles to change their routes as they proceed through the road network (this is what is meant by 'dynamic' route choice). Note that this approach requires only one simulation run of the DNL, whereby vehicles can change their route at each junction they encounter in the road network. Within this method, it is quite straightforward to incorporate route guidance; the caveat however is that each vehicle should have a default route, i.e., pre-trip assignment, in case no attention is given to route guidance.

Because in general there exists more than one route between an origin and destination pair, commuters have to select which route they will take. We use the following approach towards route choice:

1. At the first iteration of the DRC model's execution, we calculate an initial travel time for each link in the road network; this travel time is defined as the length of the link divided by its free-flow speed. In subsequent iterations, we define the travel time of a link as the arithmetic average of the travel times of all the vehicles traversing the link.

2. Based on the current travel times of all the links, shortest paths are calculated for each OD-pair, using, e.g., Dijkstra's algorithm [Dij59], the heuristic shortest-path algorithms of Jacob et al. [Jac98], or the efficient algorithm of Rosswog et al. which is based on tree heuristics that uses the hierarchical information of a road network [Ros01], and taking into account the set of predefined feasible routes $\mathcal{R}_{\text{pre}}$ (see step 2 in Section 7.1.2.1).

3. All vehicles now select a route from this set, according to the following multinomial logit model[7], which gives the probability of a given alternative route i at departure time $t_{\text{departure}}$ as:

$$p_i(t_{\text{departure}}) = \frac{e^{\mu(-U_i(t_{\text{departure}})+\epsilon_i)}}{\sum_{j\in\mathcal{R}} e^{\mu(-U_j(t_{\text{departure}})+\epsilon_j)}}, \tag{7.3}$$

in which $U_i(t_{\text{departure}})$ is the utility of route i at the specified departure time, $\mathcal{R} \subseteq \mathcal{R}_{\text{pre}}$ is the set of all available routes between a vehicle's origin and destination pair, μ is a dispersion factor and ϵ_i is a stochastic error term.

7.1.2.5 Some remarks on the convergence of simulation-based DTA

As already mentioned in Section 7.1.1.2, convergence is in general not a guarantee in simulation-based dynamic traffic assignment. Issues such as existence and uniqueness of a well-defined equilibrium, as well as stability of the solution, still remain an open question. Due to the stochastic nature of the simulations, convergence can be hindered in the case of oscillations, or even due to a gridlock situation.

The primary technique adopted in this case, is called *heuristic relaxation*: a predefined performance measure is compared for subsequent iterations, until the difference is no longer significant. Two possible performance measures are the following:

- the total travel time in the system, i.e., the sum of the travel times of all the vehicles at the end of the simulation,

[7] See the report of Batley et al. for a thorough overview of different logit models for route choice [Bat].

- or the combined variances of the travel times of all links separately, or the variance of the total travel time in the system.

In a more system-oriented setting, we can use the so-called *relative duality gap* (RDG), for which we refer to the work of Carlier et al. [Car05].

Note that it is advisable to set a pre-defined upper limit on the number of feedback iterations in the framework depicted in Section 7.1.2.1 (this might indicate a failure to converge). Finally note that this construction *assumes* the existence of an equilibrium; it is however *not a necessity*, as the algorithm is terminated anyway after a finite number of iterations.

> With respect to the framework proposed in Section 7.1.2, we implemented a small case study. The road network consisted of one origin and one destination, connected via two single-lane links. The traffic flow model was represented as the single-cell STCA traffic cellular automaton model (see also Section 4.3.2.1), with $\Delta X = 7.5$ m and $\Delta T = 1$ s; for the first link, the slowdown probability was set at 0.25, the maximum speed at 5 cells per time step, and a total link length of 1000 cells. For the second link, these values were set at 0.75, 3 cells per time step, and 500 cells, respectively. We loaded 5000 agents onto the network over a simulation period of three hours (their preferred arrival times were all set at halfway this period), with vehicles put in a waiting queue at the origin if they could not enter a link (as such, we kept a FIFO discipline whereby one link is able to block the other one).
>
> As the vehicles were driving in the network, their respective travel times were used for determining the average travel time on each link (only completed journeys were taken into account). These were then converted into utilities, whereby the route choice component employed a binary logit (this was justified because both routes are independent of each other). Departure times were chosen in accordance with the methodology explained in Section 7.1.2.3. The execution of the route choice model required not many iterations (e.g., generally less than 10), as there were only two routes and the model converged easily in this setting. For the departure time choice model however, we needed 20 iterations or more, before the generalised travel costs of all the agents became balanced. Note that the computational complexity of the simulation increases as there are more agents in the system.

7.2 An efficient dynamic network loading model (DNL)

Today, a main challenge is the construction of macroscopic and microscopic models that lend themselves to a faithful representation of road traffic, as these models are used in several key aspects in the control of traffic flows.

Within this context, our research is aimed at assisting traffic engineers who wish to evaluate what-if scenarios and/or perform real-time control of traffic flows. Whereas the former requires a sufficiently detailed model, the latter calls for an efficient implementation that allows fast simulations. The challenge thus consists of the development of a flexible testbed environment that is capable of providing us with a detailed simulation model of a real-world road network, not containing too many parameters that require extensive calibration: the TCA models described in Chapter 4 fit this description nicely.

Although these TCA models allow for fast computations, they are nevertheless computationally very expensive because they are based on behavioural models that need to be applied to each vehicle at each time step (i.e., the car-following and lane-changing models). We thus need to find the most optimal solution in terms of time and space complexity. A logical step in this direction, is an efficient parallellisation scheme that lowers the computational overhead involved. This can be accomplished by using distributed computing, where we partition the road network in several distinct geographical regions that are assigned to different machines which run in parallel.

We automatically gain platform independency using Java™ . The challenge now is to get reliable and efficient (i.e., faster than real-time) operation of a very heterogeneous computing environment. To this end, the simulator consists of one master, controlling several different workers that efficiently simulate local traffic flows.

In this section, we first give an introduction that describes traffic flow simulation from a historical perspective, paying attention to the role of open-source software development. We then give a functional description of our DNL model, called *Cellular Automata Traffic SIMulation* (CATSIM). This is followed by some implementation details of the code, after which we discuss our approach towards an efficiency increase through the paradigm of distributed computing [Mae03b; Mae04a].

7.2.1 Development of traffic flow simulators

In this section, we first give a brief overview of the development of traffic flow simulators, looked at from the perspective of the programming languages involved and the computational complexity of the models. The second part of this section considers the effects of developing software under an open-source flag. It is worthwhile to take a look at this aspect, as most of the traffic flow models tend to be developed in an accessible academic setting, but once commercialisation 'kicks in', the model's internals tend to get shrouded in legalese. In our discussion, we contrast open-source development with the classical approach of non-disclosure of the software's internals, give pointers to some of the existing licences that can regulate the commercial and non-commercial use of this type of software, and finally conclude with a note on legal issues related to intellectual property rights, the patenting of ideas, inventions, and algorithms.

7.2.1.1 Traffic simulation from a historical perspective

Traffic flow simulators have come a long way since their inception in the early fifties (see, e.g., the *TRAffic Network Simulator* – TRANS, which shows a remarkable parallel with early traffic cellular automata models [Kat63]). In those days, computers operated in both an *analogue* and *digital* fashion. However, as the former became more expensive when larger systems were simulated, the latter gained a strong foothold in the simulation community [Ger64]. Nowadays, as desktop computers get smaller and more powerful, the traffic flow simulation software has undergone a drastic evolution. It is implemented in either procedural languages (e.g., C, FORTRAN, ...) or object-oriented ones (mainly C++ and Java™), with the latest trend to employ script-based languages (e.g., Ruby) *within* the environment of a simulator itself (see for example the OmniTRANS project [Ver03a]). The simulators are applied to moderately sized transportation networks, whilst still allowing a rather detailed view on traffic operations.

Another evolution that is noticeable, is the upcoming market of complete travel forecasting models that are based on dynamic traffic flow models. This class of software applications has features such as complete GUIs, fully integrated travel demand modelling, calculation of measures of effectiveness (e.g., noise and pollutant emissions), ... Examples of such full-fledged models are TransCAD [Cal01], OmniTRANS [Ver03a], DynaSMART [DYN03], ... In many cases, the DNL core of these models is formed by a mesoscopic or macroscopic model, but there is an evolving trend towards more realistically microscopic models. However, noting the current state-of-the-practice in the field of traffic flow engineering, we note that it is becoming more and more appealing to move from static paradigms towards the use of fully integrated DTA models on a commercial basis [Mae04b].

> In our view, it is not necessary to get all the dynamics correct on a detailed microscopic level. As such, TCA models (see Chapter 4) can offer a certain degree of detail, while retaining computational performance and remaining comparable to their mesoscopic/macroscopic counterparts. One of the main advantages of the TCA modelling paradigm is that it does not require many parameters, as opposed to other microscopic traffic flow models in which the plethora of parameters and features clouds a clear understanding of the models' dynamic properties (see also Sections 3.2.3.4 and 3.2.3.6).

7.2.1.2 The benefits of software development under an open-source flag

Despite the fact that most of the traffic flow models tend to be developed in an accessible academic setting, the traditional approach towards the creation of the majority of ready-to-use software is mainly oriented towards its commercialisation. As a consequence, it is beneficial from a marketing perspective to provide prospective customers with complete packages that integrate transportation planning models, e.g., the four step model (see Section 3.1.2 and some of the microscopic simulators mentioned in Section 3.2.3.6).

In many cases, the main stream company policy is aimed towards the non-disclosure of the models' internals, effectively reducing these commercial packages to advanced versions of black-box models. When such software starts to grow more mature and complex, it becomes increasingly difficulty to answer the question *"What is really under the hood ?"* The importance of this statement should not be underestimated, as it is vital for transportation engineers to be acquainted with a model's inner workings, features, and limitations, when interpreting results for, e.g., policy decisions.

This lack of openness, can be remedied by developing the software under an open-source flag. From this point on, the complete underlying model structure remains revealed at all times, as it is now possible for many programmers to read, redistribute, and modify the source code. When a company exhibits this sagacity, the unlocked potential of open source can be fully brought into play. One of the main benefits of this paradigm is that there are effectively 'many eyes looking at one single problem'. As a direct result, the debugging, maintenance, and support life cycles of such software become more transparent, as opposed to the monolithic approach typically encountered in propriety software [Ray00]. If such an open-source project is properly managed (which implicitly requires skilled people), it can receive an increased gain from the feedback of its user base. Already, several successful examples of this type of software development can be found in real life, e.g., the Linux operating system, the Netscape and Mozilla web browsers, the StarOffice suite and OpenOffice.org project, ... Within the traffic community, the open-source approach is slowly starting to pick up, for instance with a prime example such as the *Simulation of Urban MObility*[8] (SUMO) [Kra04].

When releasing open-source software, there literally exists a myriad of licences that regulate the commercial and non-commercial use of this type of software, as well as its incorporation in third-party software. Archetypical examples are the Free Software Foundation's GNU General Public Licence[9] (GPL) with the popular catch phrase *"free as in free speech, not as in free beer"*, the Open-Source Initiative[10] (OSI) which provides a marketing vessel for 'selling' free software, the Creative Commons Licences[11] (CCL) that offer a flexible copyright for creative work, ...

Finally, note that in our discussion, we did not state anything about legal issues such as the management of intellectual property rights, issues related to the patenting of ideas, inventions, and algorithms, et cetera. Indeed, most licences undoubtedly steer clear of these topics, allowing their interpretation to remain up to the developer and/or the company. However, the central core that forms the business model for open-source software, is to freely share the software, whilst selling support. With respect to academic institutions and their management of intellectual property, dissemination of algorithms by means of publications in journals might be discouraged. In these cases,

[8] http://sumo.sourceforge.net
[9] http://www.gnu.org/licenses
[10] http://www.opensource.org
[11] http://creativecommons.org/licenses

we still deem it appropriate to publish the results, as we believe that the money remains in the selling of the software. Another less-commercial track that can be followed, is to release the software as a web service, thus effectively hiding the underlying code of an algorithm's implementation when confidentiality issues and ownership of intellectual property rights are at stake.

7.2.2 Functional description of the simulator

Considering the CATSIM DNL model from a functional point of view, this section first gives a description of the topological and geographical structure of the road network, after which we explain some vehicle-related information, ending with what kinds of statistical data can be collected during a simulation run.

7.2.2.1 Topological and geographical structure of the road network

In CATSIM we opt for an intuitive structure, whereby the network is topologically decomposed into *nodes* and *edges*. Geographically, these graph characteristics correspond to *nodes* and *links*. Different link types may exist (such as on-ramps, off-ramps, merging areas, ...). For reasons of efficiency, we define each link to consist of one or more undivisible road segments. All links are connected by special *junction nodes*, where vehicles are transferred from one link to another (more than one link can enter or exit a junction node). The intermediate nodes connecting the different segments of a link are called *bend nodes* (note that the entire road section containing all the segments is represented with just one single CA lattice). They allow for a more realistic modelling of the road network. Note that the specification of a node requires *X*, *Y* and *Z coordinates*, thus we take road gradients (e.g., elevations, tunnels, ...) explicitly into account (although it is up to the car-following model to actually use this information). With these elementary building blocks, the motorway network can easily be constructed using data provided by satellite images and/or geographical information systems (GIS).

It should be stated that in our current specification, there is no definition of what the actual underlying low-level TCA models are. They may even vary from link to link if necessary, giving a flexible and open architecture. Notwithstanding this freedom, they do have to agree 'functionally', e.g., lane changes should either be mandatory or discretionary (see Section 3.2.3.1), time steps should be comparable, ...

7.2.2.2 Vehicle-related information

As heterogeneity of a traffic stream is a necessity for a rich dynamical behaviour, we allow for different classes of vehicles. This includes cars and trucks, with trucks occupying extra cells as described in the multi-cell TCA models of Section 4.4. A convenient method for representing this difference, is using passenger car units (see

Section 2.3.1.2). Note that the rule sets do not model a vehicle as explicitly occupying more than one cell, but instead adjust the safe space headway to account for the difference in vehicle length.

Furthermore, as opposed to most other implementations, our cells do not just contain a number indicating the presence and/or speed of a vehicle. Instead, we allow for complete objects to be contained in the cells, e.g., a vehicle with a commuter's personal routing plan. Because most interactions of the vehicles are based on local information, we add another subtle refinement: information such as link travel times for example, can be put in a *central data storage* that is available to the network simulator. This means that some vehicles can be considered as *'informed drivers'* having access to this data storage, and are thereby able to reroute their trip in order to avoid encountering network congestion.

7.2.2.3 Collecting statistical data

The simulator's road network can be equipped with artificial loop detectors (see Sections 4.2.3.1 and 4.2.3.3 for an overview of different types of detectors in a TCA setting). They accurately compute various statistics from the passing traffic flow, continuously storing all results in the central data storage. Even travel times recorded by probe vehicles can be contained, such that this information becomes available to some of the vehicles (i.e., the informed drivers) as they travel through the network.

7.2.3 Code implementation details

Beside the functional description of the previous section, we now shed some light on our proposed choice of programming language for implementing a traffic flow simulator. Afterwards, we illustrate some technical aspects related to the implementation of CA lattices, as well as some details regarding the implementation of links in a road network.

7.2.3.1 Choice of programming language

Whereas earlier designs of traffic flow simulators were based on procedural languages (e.g., pure C code), we nowadays observe a trend towards the adoption of object-oriented programming languages. In this spirit, we propose to use the Java™ language, as it has been designed around a *"write once, run anywhere"* (WORE) philosophy, implying cross-platform portability without needing any recompilation of the code base.

7.2.3.2 Some technical aspects related to the implementation of CAs

Classical implementations of CA models were typically aimed at obtaining a high computational speed. This led to the use of techniques, e.g., single-bit coding schemes, typically targeted towards specific hardware platforms. The coming of popular object-oriented programming languages such as C++ and Java™ , coupled with the steady increase of computational power in average desktop computers, makes the original line of work a bit outdated.

As with respect to the implementation of a CA's grid itself, there are two approaches possible:

- **site oriented**: this is typically based on *an array of cells*, which is more suited for links having *high densities*,
- **particle oriented**: this is typically based on *a linked list of vehicles*, which is more suited for links having *low densities*.

In practice, it is best to consider the best of both worlds, i.e., only relevant sites are updated. In this view, a site corresponds to a lateral section of a multi-lane link (i.e., all cells located at the same longitudinal position).

When exchanging vehicles between consecutive links at junction nodes, we use a *lane connectivity table* that contains the numbers of all outgoing and incoming lanes, each time in the local numbering scheme (the same holds for all intersection logic). The following table gives the connectivity for the example of a main road and on-ramp towards a merge section as depicted in Figure 7.3:

$$\begin{pmatrix} 1a.1 & \longrightarrow & 2.2 \\ 1a.2 & \longrightarrow & 2.3 \\ 1b.1 & \longrightarrow & 2.1 \end{pmatrix}$$

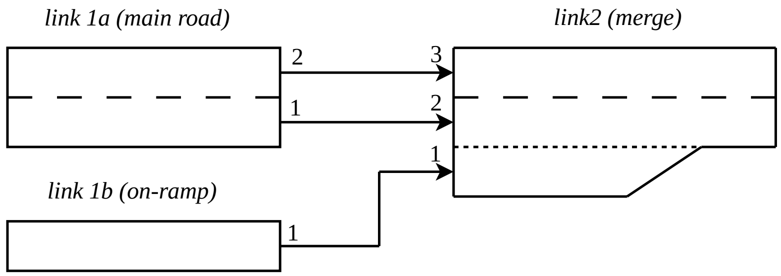


Figure 7.3: A graphical sketch of the lane connectivity for a main road and on-ramp towards a merge section; vehicles are exchanged between consecutive links based on a lane connectivity table corresponding to the diagram.

Furthermore, in CATSIM, each link has both a car-following and lane-changing TCA model with corresponding parameter vectors. Separating the parameters from the models allows us to keep the latter while performing on-line adjustments to the former.

Finally, the slowdown probabilities are a property of the links, not of individual vehicles. We justify this on the basis that (density,flow) measurements are more easily calibrated for a complete road section, than for each vehicle individually. This latter would give rise to a distinct fundamental diagram for each vehicle, whereby the combination of them would result in an average fundamental diagram, depending on the vehicles' locations and surrounding traffic conditions. This clearly encompasses a cumbersome approach.

7.2.4 Increasing efficiency through distributed computing

As already stated, using microscopic traffic simulators places a large computational burden on the employed machine architecture. Many existing simulators were initially designed to run on a single CPU. Only afterwards were they converted for parallel operations (e.g., AIMSUN2 [Bar02a; Bar02b], TRANSIMS [Nag98c; Nag01], ...), with some exceptions such as PARAMICS which was designed in a parallel fashion from the ground up [Cam94; Lim00].

The same train of thought holds for most of the traffic cellular automata models. In the beginning, when they were built using parallel implementations, the parallellisation scheme was strongly reflected in their code base, relying heavily on the underlying machine architecture. Examples are models whose computations were performed on a large number of CPUs (e.g., 1024), all contained in one shared memory architecture, employing special techniques such as single-bit encoding et cetera [Nag95c].

In the recent past, we already developed a microscopic traffic simulator in Java™ , called *Mitrasim 2000*. Instead of being a true parallel implementation, it was based on a *client-server architecture* (CSA), in the sense that the simulator ran on one machine (the server); several different other machines (the clients) showed an animation of the traffic evolution on the motorway network [Mae01b]. A major problem was that, mainly due to the single CPU architecture, the simulator did not achieve real-time speed at all. However, our past experiences allow us to build a more efficient and scalable simulator, in which parallelism can be implemented through *distributed computing*.

In our framework, the concept of distributed computing implies that we no longer use a homogeneous environment of CPUs working in a shared memory architecture. Instead, a very *heterogeneous computing environment* is provided, like for example a Beowulf cluster [Nag01]. Whereas supercomputers performed intensive tasks in the past, we can nowadays observe a shift towards grid-based computing [Bak02]. For us, the challenge now is to get reliable and efficient (i.e., faster than real-time) operation of this latest networked architecture. In Figure 7.4, we can see an example of distributing the load of the motorway network over a group of computing units.

The flexible functional description set out in Section 7.2.2, allows several implementations. In this part, we present such a possible approach, in which parallelism is achieved using distributed computing. In the following sections, we first shed some light on the difference between high-throughput and high-performance computing. We then give a description of the technologies used with respect to direct communication between different processing units, as well as a method that provides us with

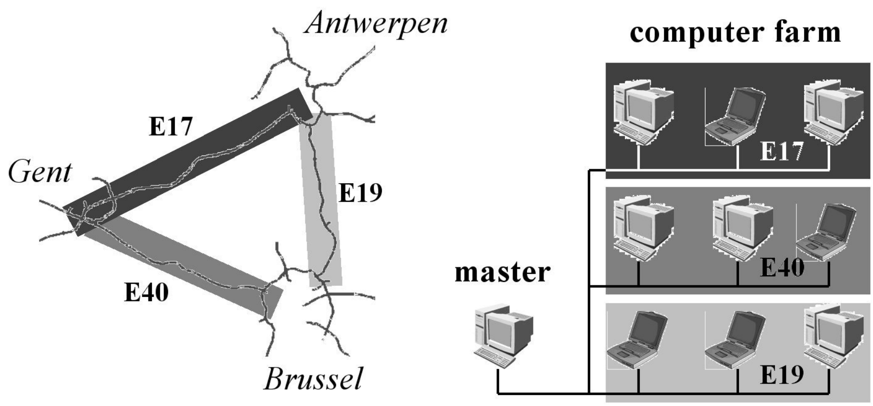


Figure 7.4: The idea behind distributed computing in the CATSIM dynamic network loading model: one computer (the master) controls several workers in a heterogeneous computing environment. All these computing units work together, whereby the load of the entire road network is distributed. In the shown example, three major motorways are modelled whereby the responsibility of each motorway is assigned to several grouped workers.

a shared memory. In a subsequent section, we explain the adopted parallellisation scheme from a programmatorical and technical point of view. We end with a brief consideration of some issues related to synchronisation, graph cycles, and data sharing.

7.2.4.1 High-throughput versus high-performance computing

As desktop computers got increasingly more powerful during the last decade, the paradigm of distributed computing has started to gain serious importance. Within this concept, a distinction is made between two radically different methodologies:

- **High-throughput computing (HTC)**
  In this setting, software is installed in a heterogeneous computing environment, thereby distributing the processing power over different computing nodes (e.g., all desktop computers in a university's research group). Users can submit tasks, which are then optimally assigned to these nodes, taking into account priorities, waiting queues, performance, … At each time, attention is given to the fact that a user can regain and keep control over all the processing power of his/her own machine, at which point the running task is scheduled and resumed at another free computing node. As a result, HTC offers a large degree of fault-tolerant computing power, available of long periods of time. Two examples of these kinds of environments are the *Condor* [Tha04] and the *H2O* projects [Kur03].

- **High-performance computing (HPC)**
  Another important aspect of distributed systems, is their ability to quickly execute certain tasks. As opposed to HTC, for which the speed of the requested computation is not per se a strict constraint but the availability of computing power on a large spatial scale is, HTC is centred around a close tie between computing power in space and time. Examples of applications in this direction are aspects such as computer graphics (e.g., ray tracing [Mae01a]), and microscopic traffic flow simulation (see, e.g., the introduction of Section 7.2.4).

7.2.4.2 Technologies used

As the whole simulator will be constructed for the Java™ Virtual Machine (JVM), we automatically achieve *cross-platform portability*; this is a necessity in order to efficiently address the heterogeneous computing environment.

Reliably controlling such a networked architecture requires a strict scheduling scheme: all processing nodes in the computing network are tightly coupled with each other. In this case, we have opted for a mixture of the *'master/worker'* and *'command' application patterns* [Fre99]. This means that we have one master computer that controls N distinct worker computers who execute the different tasks.

Currently, most distributed implementations of traffic simulators use classical communication techniques such as a *Message Passing Interface* (MPI) [Lus03] and/or *Parallel Virtual Machines* (PVM) [Gei94]. In spite of this, we strongly believe that this is not sufficient for our envisioned architecture. There's a trade-off involved, between on the one hand efficient direct communication and on the other hand an accessible shared memory architecture:

- **Direct communication**
  It is very important that the computation/communication ratio remains as high as possible, after parallelising[12] the simulator [Fre99]. Translating this to our Java™ implementation, we establish *dedicated communication channels* between the master and the workers and between the workers themselves; *sockets* provide a suitable and efficient method for this type of communication. Furthermore, all communication done between the workers, should be scheduled simultaneously with the computations they perform, so delay times can be minimised.

- **Shared memory**
  Because we are working in a heterogeneous computing environment, no standard shared memory architecture is directly available. There is however a nice solution to this: *JavaSpaces* provide a generic environment that can be accessed

[12] As Balmer et al. noted, *"With 100 Mbit Ethernet, the best possible real-time ratio of a parallel traffic simulation with a one-second time step is approximately 170."* [Bal04b]. In these systems, the bottleneck is not the communication speed but the *network latency*, i.e., the time needed to initiate a message; the only way to tackle this is to use different hardware with a lower communication overhead, e.g., Myrinet (`http://www.myri.com`), increasing the real-time ratio towards 800 [Cet03].

by any worker located anywhere in the network. It is based on the concept of *Linda Spaces* [Gel96], implemented using Java™ 's *Remote Method Invocation* (RMI) facility [Gro02], and provided as a service of the *Jini Network Technology* [Fre99; Sun03]. A major advantage is that when using a JavaSpace, no explicit network addresses (e.g., IP addresses and TCP port numbers) need to be known when communicating. All communication can be done *anonymously* (loose coupling in space), and even *asynchronously* (loose coupling in time). Despite its flexibility, the JavaSpaces service is — as stated before — in essence a medium for loosely coupled communication, and thus not well suited for performant dedicated communications. Because of this, we only use it for setup purposes and infrequently accessed resources and information.

Furthermore, it is important to keep the following two points in mind:

- the underlying Jini technology is a network based protocol, that does not offer any real-time guarantees,
- and JavaSpaces is semi-scalable: it runs in a single JVM, implying that the service itself might become a bottleneck for communication, or worse, it might fail. A solution to this problem can be to use multiple JavaSpaces that are clustered together.

7.2.4.3 Programmatorical and technical aspects

After discussing the communication aspects and shared memory setup mentioned in the previous section, we now explain the adopted parallellisation scheme from both a programmatorical and technical point of view. We describe how the workers' tasks are set up and distributed. We then give details on how the dedicated communication channels are constructed, ending with some comments on the execution of a simulation step.

- **Setting up the workers' tasks**
  We assume that, at the beginning of the simulation, the master has all the information available about the road network infrastructure, the travel demand and routing plans, ... It then performs a *domain decomposition* on a geographical basis, dividing the network in exactly N partitions (the splitting of the links is preferably done far away enough from any junction nodes, such that we can avoid the complexities of intersection logic). Note that this encompasses static load balancing (see Figure 7.4 for an example of such a decomposition); it is also possible to opt for another scheme, thereby providing us with some means to perform dynamic load balancing. This can, for example, be accomplished by keeping track of the workers' computation times and redividing and reassigning partitions.

- **Distributing the tasks**
  The next step consists of the master distributing the tasks (i.e., the different motorway stretches in each partition) into the JavaSpace. All the workers then check this JavaSpace and each worker picks one task. Note that we assume at this point that each task comprises more or less an equal amount of workload.

- **Setting up dedicated communication channels**
  Once all tasks are distributed among the workers, they proceed to create direct channels for communication with their nearest-reachable neighbours (all the workers' IP addresses are broadcasted in the JavaSpace, together with information on the neighbouring partitions). This is necessary, as all workers need to exchange information of the traffic flows at their respective boundaries. The master also advertises its location in the JavaSpace, after which each worker requests a private communication channel to the master. Figure 7.5 shows the relations between the computing units in the heterogeneous network architecture. Note that, because of its reliability, the TCP/IP protocol remains the communication mechanism of the underlying network transport layer.

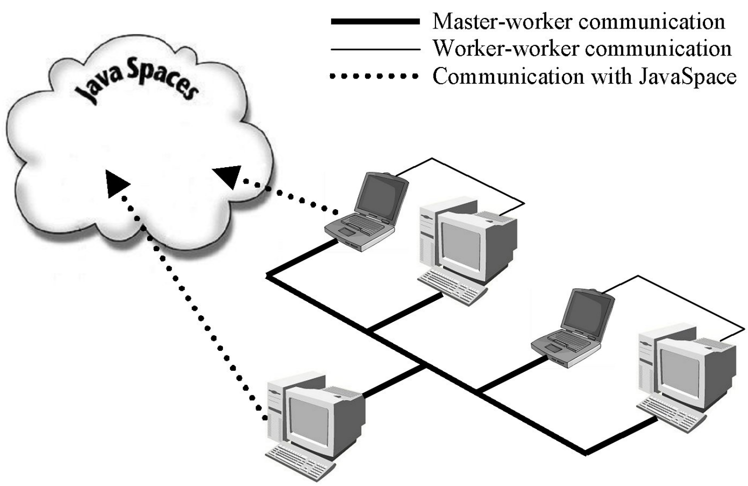


Figure 7.5: A schematic overview of the three different kinds of dedicated communication channels; communication between the master and the workers (shown as thick lines), between workers' neighbours (shown as thin lines) and with the JavaSpace (shown as dotted arrows).

- **A simulation step**
  The master then initiates communication with all the workers, instructing them to advance to the next time step of the master clock. At this stage, several intricate aspects need to be dealt with:

  – each worker knows its neighbours, and communicates with them in order to transfer vehicles that are crossing zones,

- communication should only be performed when there are vehicles to transfer; consecutive links have small overlapping regions such that vehicles transfers only occur within these regions,
- for reasons of computational efficiency, we propose a hybrid cell/vehicle oriented approach: when simulating, only active cells (i.e., containing vehicles) are updated,
- because the JavaSpace itself is not efficient enough yet and because it assumes loose coupling, we only use it for information that is not frequently accessed (e.g., link travel times that are conventionally broadcasted on a radio station or displayed above a certain road section), as mentioned in Section 7.2.4.2.

Note that with respect to the real-time simulation of traffic flows, several promising projects were carried out. Examples of this are the simulations of Duisburg, Germany [Bar99] and those of the German Autobahn network of the North-Rhine-Westphalia (NRW) region [Ric96c; Ric96b; Ric97; Wah02]. This latter example also provides the user with a prediction of the traffic state up to one hour in the future. It can be consulted on-line via a website `http://www.autobahn.nrw.de` (see Figure 7.6). The prediction is based on the *On-Line SIMulator* (OLSIM), which is an implementation of the brake-light BL-TCA model (see Section 4.4.2.2) [Chr04; Pot04]. The tuning of the simulation to the current state of the real-world road network, is done by comparing measurements from virtual detectors in the model and real-world loop detectors from the motorways at certain checkpoints (containing sources and sinks). Whenever a mismatch is found, vehicles are either added or removed, taking into account to avoid severe disturbances of the current traffic flow [Bar99; Wah02].

7.2.4.4 Issues related to synchronisation, graph cycles, and data sharing

Because the workers in the computing environment need to exchange information at their boundaries, *deadlocks* may occur in which some workers are mutually waiting for each other. However, in our implementation as described earlier, this can not happen because all the workers are directed by the master computer. This kind of *arbitration by an external third party*, is frequently done in systems needing robust synchronisation.

In a previous attempt at describing traffic in a road network, all links were initially topologically sorted after which they could be processed [Mae01b]. This excluded the presence of cycles in the graph describing the road network, which we now consider to be a major flaw of any simulator that exhibits this phenomenon. Using traffic cellular automata models solves this problem seamlessly, because all vehicle updates are now being executed simultaneously (see Section 4.1.4 for more details). In general, vehicles are tagged for lane changes (taking care of side effects such as ping-pong traffic, as described in Section 4.5.1.3), then they execute their respective lane

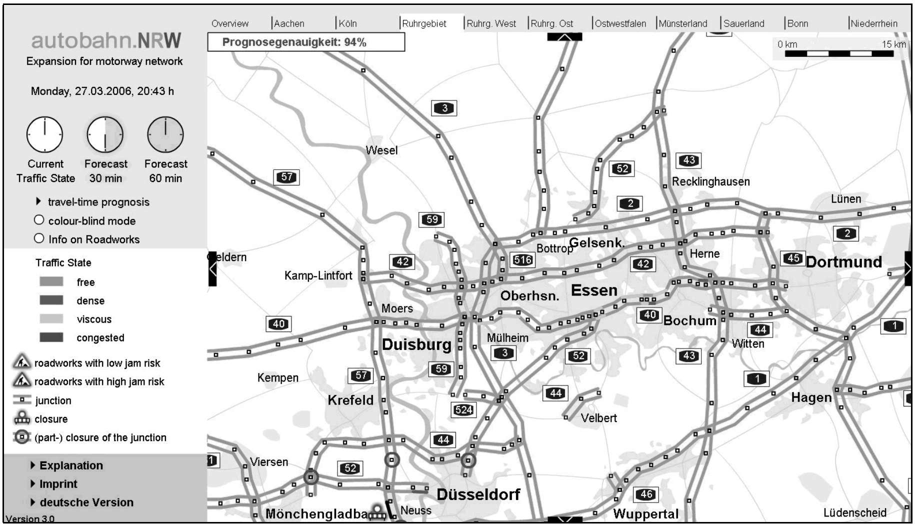


Figure 7.6: A visualisation of the traffic in the Autobahn network of the North- Rhine-Westphalia (NRW) region in Germany. The figure shows the traffic state in the Ruhr area, predicted one half hour in the future, as broadcasted on an on-line website (image reproduced after [Stü06]).

changes completing the lane-change model. Each vehicle's speed is then re-evaluated, after which all vehicles are moved to their new positions, completing the car-following model.

We conclude this section by mentioning that any data that should be shared among the workers, can be kept by a dedicated data server. In our implementation, we choose this data server to be the JavaSpace itself, providing an anonymous service of which the network address no longer explicitly needs to be known. The fact that the JavaSpace service is not yet efficient enough, is no problem because all time critical operations are done using dedicated communication channels, whereas the JavaSpace is only used for distributing tasks, sharing infrequently accessed information, . . .

7.3 Some example applications

Let us briefly consider some of the target applications of our framework (i.e., the integrated DTA methodology from Section 7.1 and the DNL model described in Section 7.2), these encompass traffic state estimation, sustainability effects of traffic management systems, and assessing the impacts of traffic control measures.

7.3.1 Reliable state estimation of the road network

As implied at the end of Section 7.2.4.3, it is possible to estimate the collective state of the traffic on the entire road network, based on information from the real-world (e.g., measurements stemming from single inductive loop detectors). As such, the framework can either simulate traffic in an off-line setting, based on historical data captured, e.g., in a single static OD matrix. It is then possible to derive information for a typical day, whereby the following aspects can be studied:

- the lengths of jams in both time and space,
- travel time losses and robustness properties,
- indicators for high-risk zones that contain recurrent congestion,
- and assessing the impact of an incident, leading to, e.g., lane closures.

In an on-line setting, the framework needs to be fed with real-time data, after which the simulation is ran to get a global updated view of the traffic state. Coupled with a prediction step, this leads to a powerful methodology that can be used to steer traffic, e.g., by advertising travel times on variable message signs (VMS).

7.3.2 Sustainability effects of traffic management systems

When thinking in a sustainable mobility framework, one approach could be to limit the traffic demand and to balance this demand over different traffic modes. As a complementary approach, one could also try to optimise the use of the existing infrastructure. With respect to the latter approach, we carried out a project, funded by the Belgian Federal Science Policy (DWTC) [Mae04e; Mae06].

One of the central components within the project, is a method to assess the 'quality' of a simulated traffic situation. To this end, we need to define goals we would like to achieve; stated in control terms, this corresponds to a cost function, called the *sustainable cost function* (SCF). In the scope of this project on sustainable mobility, a definition of the cost function includes penalisations for *pollutant emissions* (environmental costs), *congestion* (socio-economic costs), *noise emissions*, *dangerous situations* (like shock waves), ... The cost function is expressed in terms of the states of the model and can be evaluated during simulation (all these costs are expressed in monetary terms). Within the project, we controlled traffic flows with respect to this cost function. If we use the SCF in steps 4 and 6 of the framework described in Section 7.1.2.1, then this will lead to a social optimum as explained in Section 3.1.4.2 on the concept of road pricing policies. The framework can then be used as a mirror of the real world.

7.3.3 Assessing the impacts of traffic control measures

As we believe all political decisions should hinge on advice from studies, these require an a posteriori interpretation with a good dose of common sense. Most of the time, such studies try to assess the impact of policy decisions that are implemented by means of local and global control measures. Typical decisions and measures include the following:

- rerouting effects, requiring a study of day-to-day and within-day replanning of commuters,
- ATMS effects (e.g., ramp metering, speed harmonisation, platoon driving, . . .),
- policy decisions (e.g., overtaking prohibitions for trucks, road pricing strategies, . . .),
- and changes in the road infrastructure, possibly leading to induced traffic, which we believe requires a more activity-based approach.

7.4 Conclusions

In this chapter, we constructed a framework that allows us to perform dynamic traffic assignment (DTA), integrating departure time choice (DTC) and dynamic route choice (DRC), coupled with a dynamic network loading (DNL) model. The method is built around a traffic flow model that is represented as a computationally efficient cellular automaton. After explaining two of the mainstream DTA approaches, i.e., analytical and simulation-based, we gave an overview of each of the framework's components. In a separate section, we payed explicit attention to the DNL model, considering traffic flow simulation from a historical perspective, and discussing the benefits of open-source software development. After a functional description of the simulator, some code implementation details were given, ending with an overview of parallellisation through distributed computing. In a final section of the chapter, we discussed some example applications such as traffic state estimation, sustainability effects of traffic management systems, and assessing the impacts of traffic control measures.

Part V

Conclusions and Perspectives

Chapter 8

General conclusions and future research

In the final chapter of this dissertation, we give a concise overview of the results obtained in our research, centred around the state-of-the-art in the literature, numerical data analysis, and integrated dynamic traffic assignment. This is followed by an extensive account of some of the issues encountered during our explorations, requiring and suggesting further research. Note that although most of the discussed methods are also applicable to city traffic, the work in this dissertation is primarily aimed towards motorways.

8.1 Discussion and summary

The research elaborated upon in this dissertation, spanned a broad range going from a discussion of the models encountered in transportation planning and traffic flow modelling, over the concept of traffic cellular automata models, towards a numerical analysis of traffic data, ending with a framework for performing integrated dynamic traffic assignment. In the following sections, we take a look at each of these aspects in detail.

8.1.1 The physics of road traffic and transportation

Considering the plethora of notations encountered in different scientific fields related to traffic flow modelling, Chapter 2 provided an extensive account of what we suggest what would be the current state-of-the-practice, detailing several aspects related to the description of traffic flows. Most importantly, we have introduced a nomenclature convention, built upon a consistent set of notations. Besides the classical traffic flow variables and performance indicators, we discussed some of the different points of view with respect to the causes of congestion, as adopted by the traffic engineering community. In this latter aspect, we compared two different mainstream philosophies, based on congestion being deterministic (i.e., bottleneck-induced) and stochastic (i.e., spontaneous breakdown) in nature, respectively. We noted how both theories are quite different, but nevertheless compatible with each other. As choosing which school to follow is largely a matter of personal taste, we conclude that research into the behaviour of traffic at bottlenecks is one of the most important aspects in the context of traffic flow theory.

With respect to the existing literature, Chapter 3 elucidated on transportation planning models, operating on a high level, and traffic flow models that explicitly describe the physical propagation of traffic flows, typically on a lower level. The incentive for such an elaborate description was fuelled by the fact that we encounter a frequent confusion among traffic engineers and policy makers when it comes to transportation planning models and the role that traffic flow models play therein. One of the main advances to the currently existing body of literature, is our comprehensive overview, which is unique on a global scale. As of yet, when diving into the field of traffic-related research, people had to read tons of course texts, papers, … most of the time spread over different scientific areas. Our contribution to the state-of-the-art in the literature, is an integrated overview that is able to help any researcher wishing to partake in the field (note that our work excludes fields such as (agent-based) traffic control theory and practice, as this is not the focus of our research). Within our survey, we elaborated upon land-use models, trip-based and activity-based transportation models, as well as transportation economics, discussing pros and contras, and the links between them. On the traffic flow modelling side, the debate on whether or not to use macro-/meso- or microscopic models still continues to spawn many intriguing discussions (in the upcoming Section 8.2.1, we shed some light on the sense and nonsense of the development of yet another traffic flow model).

8.1.2 Cellular automata models of road traffic

Continuing our survey, Chapter 4 dived into the field of traffic cellular automata models, as being one of the most promising microscopic simulation models; they allow for computationally efficient, yet still detailed enough, calculations of the propagation of traffic flows. They found their roots in the physics discipline of statistical mechanics, having the goal of reproducing the correct macroscopic behaviour based on a minimal description of microscopic interactions. Already, several reviews of traffic cellular automata models exist, but none of them considers all the models exclusively from the behavioural point of view, as we do. As this kind of survey did not hitherto exist in the current scientific field, our overview fills this void, caused by the need for researchers to have such a comprehensive insight. We also introduced a classification based on single-cell versus multi-cell TCA models, whereby the latter class can lead to some surprising behaviour with respect to unexpected hysteresis phenomena. In a sense, we also believe that developing yet another TCA model is no longer a necessity; rather, putting the existing models to good use has become more of an issue in our opinion (e.g., the use of the brake-light TCA model in the *On-Line SIMulator* (OLSIM) [Chr04; Pot04]).

Bridging the gap between microscopic and macroscopic models, we presented in Chapter 5 an alternate methodology that implicitly incorporates the STCA's stochasticity into the macroscopic first-order LWR model. The innovative aspect of our approach, is that we derive the LWR's fundamental diagram directly from the STCA's rule set, by assuming a stationarity condition that converts the STCA's rules into a set of linear inequalities. In turn, these constraints define the shape of the fundamental diagram that is then specified to the LWR model. The main insight gained from our approach, is that there can be a significant difference between an average fundamental diagram (STCA) and a stationary fundamental diagram (LWR). As a result, the STCA model is able to temporarily operate under larger flows and densities than those possible for the LWR's stationary fundamental diagram. As such, it becomes very important to correctly capture the capacities in both the STCA and LWR models in the presence of noise.

8.1.3 Numerical analysis of traffic data

Chapter 6 revolved around an exploratory analysis of traffic data, mostly stemming from single inductive loop detectors embedded in Flanders' motorways. Our techniques were aimed at the raw, noisy data, containing outliers, missing values, ... To this end, we implemented a methodology that tracks outliers from a statistical point of view. We also developed a visual technique, based on maps, that allows a quick assessment of structural and incidental detector malfunctioning.

In contrast to the many existing approaches for travel time estimation, we considered the use of cumulative curves, based on flow measurements, for the off-line estimation of travel times. A central aspect in this methodology was dealing with synchronisation issues and systematic errors; afterwards, it was possible to estimate the distribution of the travel time. We applied our methodology to case studies on the E19 motorway and the R0 ring road, thereby uncovering the differences in travel time distributions on, e.g., Mondays and Fridays.

The final part of our research on numerical analysis of traffic data, dealt with reliability and robustness properties related to traffic flow dynamics, giving us an extra instrument for the analysis of recurrent congestion. These tempo-spatial maps of the traffic evolution provide us with a powerful method for assessing structural congestion on a typical weekday. This can assist policy makers in deciding where to spend attention when tackling congestion, e.g., indicating the hot spots that are sensitive to disturbances.

8.1.4 Integrated dynamic traffic assignment

Finally, our research described in Chapter 7 of this dissertation, consists of the development of a framework that allows us to perform simulation-based dynamic traffic assignment. Nowadays, it has become more or less mandatory to include departure time choice and dynamic route choice, coupled with a dynamic network loading model. Our proposed framework describes a straightforward method for doing all three, whereby we opt for a sequential inclusion of both departure time choice and dynamic route choice models. The underlying dynamic network loading model is represented as a computationally efficient cellular automaton. In order to furthermore increase the efficiency, we explain a technique that adopts the concept of parallellisation through distributed computing, i.e., dividing the total work load over several distinct central processing nodes.

8.2 Future research

With respect to future research, we now provide an extensive account of some of the issues encountered during the course of our research; we have ordered them into four distinct groups, i.e., (i) traffic flow models, (ii) numerical data analysis, (iii) integrated dynamic traffic assignment, and (iv) general road traffic-related remarks.

8.2.1 Traffic flow models

- The lack of a unified notational standard for traffic flow variables has bothered many scientific fields that are drawn to the conglomerate that traffic flow modelling has become (i.e., the integration of engineering, mathematics, physics, economics, psychology, ...). The adoption of a logical and consistent terminology is a necessity when it comes to creating order in the 'zoo of notations' that currently exists. We believe this largely exceeds the idea of an intuitive notation that is different in each scientific field separately.

- There is a need for a consequent analysis of all kinds of developed traffic flow models, their mathematical properties and physical soundness (e.g., are there phantom phenomena occurring that are only encountered in the model structure itself ?). As researchers in the scientific field seem to spawn many traffic flow models, the question arises as to how do these relate to each other and what are their respective strengths and weaknesses ?

- Closely related to the previous remark, is the sense and nonsense of developing yet another 'new' traffic flow model. Although it has benefits in that we may get renewed insights into already existing model formulations, their (hidden) assumptions and properties, ..., it nevertheless is time to put the models to good use. In many cases, the modelling done can be viewed as a mere mathematical exercise. Irrespective of the previous comments, the quest remains to construct the most simplistic model that has the greatest explanatory strength, whilst still being based on tangible real-world principles.

- It is easier to model and predict the behaviour of groups of people (corresponding to socio-economics and statistical mechanics), than the behaviour of a single individual (corresponding to psychology). As such, it remains a task for the humanities and social sciences to devote research in this area with respect to the modelling of transportation demand. The inclusion of psychological aspects is especially useful when assessing how human beings comply and react to advanced traveller information systems (ATIS), e.g., route guidance, …

- Related to the previous remark and the conclusions 3.3 in Chapter 3, there are two more aspects that deserve attention:

  - On the one hand there are the *self-organising aspects* of a transportation system. In nearly all cases of traffic flow models, this factor is neglected or even not an issue. We could argue that traffic cellular automata models and stochastic models in general have self-organising aspects, as they can lead to the spontaneous formation of jams. However, the real power of self-organisation lies in how we can model traffic not just as a medium but as a complete interactive environment, in which human behaviour plays an important role. How can the presence of information influence this behaviour ? Can this be incorporated in ‘intelligent’ traffic flow models ? How is the interaction between a driver and his vehicle, between a driver and other drivers in his neighbourhood, and between a driver and his environment (i.e., the road infrastructure, sign posts, traffic lights, …) ?
  - On the other hand there is a large concern with respect to traffic safety. The traditional approach is centred around statistics based on accidents, sometimes including more elaborated methods such as classes of road types, intersection layout, … With the arrival of more powerful computers, it has become possible to execute detailed simulations that incorporate all kinds of elements, such as physical vehicle and engine characteristics, human behaviour and tactical decisions when accelerating, decelerating, changing lanes, and crossing intersections, the interaction between a vehicle and a human’s physiology, … An example of such a promising model is the PELOPS submicroscopic traffic flow model, described in Section 3.2.4.

8.2.2 Data quality, travel time estimation, and reliability

- Performing robust statistics based on large-scale data sets (i.e., a high number of observations n and a high number of dimensions p) still remains a challenge. The MCD estimator used in Section 6.2.2.2 is not suited for data sets containing more than say 50,000 data points. If this is the case, it is more advisable to use an alternative, such as the median and median absolute deviation (MAD) as robust estimates for the mean and variance.

- With the advent of upcoming technologies such as GSM and GPS probe vehicles, we can envision a trend towards a true integration of data from all kinds of sources on a country-wide scale. This includes, e.g., traffic counts from single and double inductive loop detectors as well as cameras, information from traffic lights, travel times from probe vehicles, weather information and forecasts, ... The issue then comes down to the most effective way of mining all the information in this 'national data warehouse'. With respect to the control of traffic flows and the dissemination of information towards the travellers, this becomes a significant challenge for future stake holders.

- There is the need for reliable travel time estimation in an on-line setting; this is useful in a control-oriented context for real-time advanced traffic management systems (ATMS) and advanced traveller information systems (ATIS). When a new measurement from the system becomes available, how can we translate it into another form of information, i.e., an updated travel time, such that we can tune our controller ?

- The travel time estimation procedure explained in Section 6.3.2 should be compared to the existing START/SITTER[1] system in Flanders, which bases its travel times on the inverse of the speed as recorded by single inductive loop detectors. Preliminary results in this direction are already obtained by Logghe and Van Hove [Log05a]; in congestion, the latter exhibits much more variability than the former technique, due to the fluctuations at low speeds caused by stop-and-go traffic.

- When estimating the cumulative N curves in Section 6.3.2, we did not make a distinction between vehicle classes, leading to an average travel time. In this respect, a further refinement would be to estimate them for cars and trucks separately. Because on a longer road section, vehicles will probably tend to travel at different speeds, implying overtaking manoeuvres, we can assume that the FIFO condition no longer holds. How can we now incorporate this in the procedure for estimating travel times based on cumulative N curves ?

- Another remark related to the previous comment, is how to do travel time estimation for complete roads, i.e., sequences of road segments. How do we treat the time spent in the complexes of on-/off-ramps in between ? Can we use traffic counts from the detectors located at the on-/off-ramps themselves ? If this is not possible (e.g., no detectors exist), can we then use the mean speed as calculated by the inverse of the travel time upstream/downstream of the complex, or the mean speed as reported by the upstream/downstream detectors of each section

[1] START/SITTER is an acronym for *"Systeem Trafiek op Autosnelwegen Reële Tijd – Système Intelligent Trafic en Temps Réel"*; it is a system that processes the traffic measurements on the Belgian motorway network in real time.

nearest to the complex ? Do we have to take into account the current regime, i.e., congestion upstream or downstream when selecting the correct speed ?

- One of the main challenges in capturing valuable real-time traffic data (e.g., actual travel times), is the correct map-matching from GSM probes onto the underlying road network. An even better approach is to directly use the position of GPS-probe vehicles; influential players on this market include transportation companies that equip their truck fleet with GPS devices, allowing the data to be used in order to get an accurate picture of traffic conditions on motorways, in cities, … One of the most complete scenarios in this respect is when the technology is implanted in each vehicle, converting them all into probes that gather floating car data.

- Related to the illustrative detector maps of Section 6.2.3.2, it is also possible to coalesce the three used statistics S_1, S_2, and S_3, by assigning a distinct color to each one of them (e.g., red, yellow, and blue). Summing them will result in a coloured map, which is reminiscent of the expression levels on micro-arrays in the discipline of bio-informatics.

- Finally, when constructing the reliability maps in Section 6.3.3, we based our statistics on the median as a robust location estimator. It is however also possible to use, e.g., the 95% or even the 99% percentile, which will give us a possible indication of very rare events such as incidents (e.g., football games) and accidents. This can give the road operator a clue as to where traffic safety might be a concern (e.g., the so-called black spots), which areas are more sensitive to disturbances, …

8.2.3 Integrated dynamic traffic assignment

- A first item to tackle, is constructing a practical computer implementation of the framework proposed in Chapter 7, thereby creating a model that can be used for scenario evaluation and the applications mentioned therein. Crucial in the development will be the behaviour of the model with respect to convergence.

- One of the major questions in the field is whether or not an equilibrium always exists. In this respect, it is advisable to first define what is meant by an equilibrium (e.g., reaching some sort of a stationary state). Does an equilibrium then always exists in simulation-based traffic assignment, what about gridlocks that cause problems for these kind of DTA models (e.g., TRANSIMS does not perform DTC as of yet, but its DRC component averages route travel times in 15 minute intervals, feeding them back to the assignment module; from time to time, oscillations caused by rerouting tend to occur) ? Assuming perfect information and deterministic flows, the existence of an equilibrium can be expected, but the world is not made up of perfectly informed commuters, et cetera. How about convergence and stability issues, existence proofs, … as explained in the work of Szeto and Lo [Sze06] ? And what if there exist road networks that might not be able to achieve an equilibrium due to their layout: jam lengths and times can vary from day to day, resembling an unstable system.

- Related to the previous remark, is the notion that, notwithstanding the possibility on convergence, there is another issue that needs to be addressed. When the DNL model is executed, we obtain the result of one such a Monte Carlo simulation. In this respect, it is necessary to perform multiple simulation runs, as they each are the outcome of a stochastic process. How do we assess the combination of these runs ?

- How can we define the set of available paths that are considered by a commuter ? Is this a large diverse set, a small one, which paths are dropped, ... ?

- How are extensions such as multi-modal traffic, multiple trip-legs, et cetera included in the modelling approach ? We believe an activity-based setup is best suited for this goal.

- How to calibrate a framework for integrated dynamic traffic assignment such as ours, by means of data stemming from traffic detectors (e.g., cameras, single inductive loop detectors, ...) ?

8.2.4 General road traffic-related remarks

- How can we incorporate travel time reliability in the a priori design of a road network ? Can we adapt existing networks to this end, for example through the introduction of some form of ‘controlled’ flexibility (e.g., the allowed use of hard shoulder lanes, which implies extra risks in case an incident occurs and they are needed for ambulances and the like) ?

- What is the role of public transportation ? How can this sector benefit in the sense of traffic flows, i.e., using dedicated lanes, giving priority to buses and trams by means of the explicit control of traffic lights, using buffer zones at intersections, ... ?

- There is an increasing need for the adaptation of the function of a road to its design. E.g., the implementation of policy measures such as zones with maximum speeds of 30 km/h near schools, may require a change in the road’s infrastructure; this will discourage fast driving, making it more logical and sensical for humans to stick to the imposed speed limit.

- Besides caused by incidents, congestion always occurs due to a traffic demand that exceeds the supply. An increased demand also leads to an increased probability of congestion, as driver fluctuations will play an important role near the critical density. As a result, advanced traffic management systems (ATMS) and advanced driver assistance systems (ADAS) may result in a slight amelioration of the traffic situation. We therefore believe that the main focus towards the alleviation of congestion should be on a higher level, i.e., where route choice and assignment occurs, possibly through the implementation of road pricing policies and the like.

- Is it useful or just pointless to equip every commuter with route guidance and actual information ? Some may win, some may lose as the excess of demand needs to go somewhere. Everybody can be more or less satisfied if all measures are fully integrated with each other (i.e., the ATMS on the road, as the integration of public transportation, as the shifting of departure times, . . .). In the end, the game might boil down to the question: *"How intelligent is my route planner and is it able to beat every on else on the block, or in other words: how can I beat the system ?"*. A key question is how this influences network equilibria, and if we can use this knowledge to anticipate on traffic conditions and steer them towards some set point.

Part VI

Appendices

Appendix A

Glossary of terms

A.1 Acronyms and abbreviations

4SM	four step model
AADT	annual average daily traffic
ABM	activity-based modelling
ABS	anti-locking brake system
ACC	adaptive cruise control
ACF	average cost function
ADAS	advanced driver assistance systems
AIMSUN2	Advanced Interactive Microscopic Simulator for Urban and Non-Urban Networks
ALBATROSS	A Learning BAsed TRansportation Oriented Simulation System
AMI	average mutual information
AMICI	Advanced Multi-agent Information and Control for Integrated multi-class traffic networks
AON	all-or-nothing
ARIMA	autoregressive integrated moving average
ASDA	Automatische StauDynamikAnalyse
ASEP	asymmetric simple exclusion process
ATIS	advanced traveller information systems
ATMS	advanced traffic management systems
BCA	Burgers cellular automaton
BJH	Benjamin, Johnso, and Hui
BJH-TCA	Benjamin-Johnson-Hui traffic cellular automaton
BL-TCA	brake-light traffic cellular automaton
BML	Biham, Middleton, and Levine
BML-TCA	Biham-Middleton-Levine traffic cellular automaton

BMW	Beckmann, McGuire, and Winsten
BPR	Bureau of Public Roads
BTS	base transceiver station
CA	cellular automaton
CA-184	Wolfram’s cellular automaton rule 184
CAD	computer aided design
CATSIM	Cellular Automata Traffic SIMulation
CBD	central business district
CCL	Creative Commons Licences
CFD	computational fluid dynamics
CFL	Courant-Friedrichs-Lewy
CFVD	Cellular Floating Vehicle Data
ChSch-TCA	Chowdhury-Schadschneider traffic cellular automaton
CLO	camera Linkeroever
CML	coupled map lattice
CONTRAM	CONtinuous TRaffic Assignment Model
COMF	car-oriented mean-field theory
CPM	computational process models
CSA	client-server architecture
CTM	cell transmission model
DARPA	Defense Advanced Research Projects Agency
DCE	delay coordinate embedding
DDE	delayed differential equation
DFI-TCA	deterministic Fukui-Ishibashi traffic cellular automaton
DGP	dissolving general pattern
DLC	discretionary lane change
DLD	double inductive loop detector
DNL	dynamic network loading
DoD	Department of Defense
DRC	dynamic route choice
DRIP	dynamic route information panel
DSA	Daily Statistics Algorithm
DTA	dynamic traffic assignment
DTC	dynamic traffic control departure time choice
DTM	dynamic traffic management
DUE	deterministic user equilibrium
DynaMIT	Dynamic network assignment for the Management of Information to Travellers
DYNASMART	DYnamic Network Assignment-Simulation Model for Advanced Roadway Telematics
ECA	elementary cellular automaton
EDA	exploratory data analysis
ELA	emergency lane assist
EM	expectation-maximisation

EP	expanded congested pattern
EPS	electronic power steering
ER-TCA	Emmerich-Rank traffic cellular automaton
ESP	electronic stability programme
FCD	floating car data
FDE	finite difference equation
FIFO	first-in, first-out
FNN	false nearest neighbours
FOTO	Forecasting of Traffic Objects
FVD	floating vehicle data
GETRAM	Generic Environment for TRaffic Analysis and Modeling
GHR	Gazis-Herman-Rothery
GIS	geographical information systems
GNSS	Global Navigation Satellite System (e.g., Europe's Galileo)
GoE	Garden of Eden state
GP	general pattern
GPL	GNU General Public Licence
GPRS	General Packet Radio Service
GPS	Global Positioning System (e.g., USA's NAVSTAR)
GRP	generalised Riemann problem
GSM	Groupe Spéciale Mobile
GSMC	Global System for Mobile Communications
HAPP	household activity pattern problem
HCM	Highway Capacity Manual
HCT	homogeneously congested traffic
HDM	human driver model
HKM	human-kinetic model
HPC	high-performance computing
HRB	Highway Research Board
HS-TCA	Helbing-Schreckenberg traffic cellular automaton
HTC	high-throughput computing
ICC	intelligent cruise control
IDM	intelligent driver model
INDY	INteractive DYnamic traffic assignment
ITS	intelligent transportation systems
IVP	initial value problem
JDK	Java™ Development Kit
JRE	Java™ Runtime Environment
JVM	Java™ Virtual Machine
KKT	Karush-Kuhn-Tucker
KKW-TCA	Kerner-Klenov-Wolf traffic cellular automaton
KWM	kinematic wave model
LIBRA	Library for Robust Analysis
LGA	lattice gas automaton
LOD	level of detail

LOS	level of service
LPF	low-pass filter
LSP	localised synchronised-flow pattern
LTM	link transmission model
LWR	Lighthill, Whitham, and Richards
MAD	median absolute deviation
MADT	monthly average daily traffic
MCD	minimum covariance determinant
MCMC	Markov chain Monte Carlo
MC-STCA	multi-cell stochastic traffic cellular automaton
MD	Mahalanobis distance
MesoTS	Mesoscopic Traffic Simulator
MFT	mean-field theory
MI	multiple imputation
MINDAT	Minute Data
MITRASIM	MIcroscopic TRAffic flow SIMulator
MITSIM	MIcroscopic Traffic flow SIMulator
MIXIC	Microscopic model for Simulation of Intelligent Cruise Control
MLC	mandatory lane change
	moving localised cluster
MOE	measure of effectiveness
MPA	matrix-product ansatz
MPC	model predictive control
MPCF	marginal private cost function
MPI	Message Passing Interface
MSA	method of successive averages
MSCF	marginal social cost function
MSP	moving synchronised-flow pattern
MT	movement time
MTS	Mobile Traffic Services
MUC-PSD	multi-class phase-space density
NaSch	Nagel and Schreckenberg
NAVSTAR	Navigation Satellite Timing and Ranging
NCCA	number conserving cellular automaton
NRW	North-Rhine-Westphalia
NSE	Navier-Stokes equations
OCT	oscillatory congested traffic
	optimal control theory
OD	origin-destination
ODE	ordinary differential equation
OLSIM	On-Line SIMulator
OSI	Open-Source Initiative
OSS	Open-Source Software
OVF	optimal velocity function
OVM	optimal velocity model

Paramics	Parallel microscopic traffic simulator
PAT	preferred arrival time
PATH	California Partners for Advanced Transit and Highways Program on Advanced Technology for the Highway
PCE	passenger car equivalent
PCU	passenger car unit
PDE	partial differential equation
PELOPS	Program for the dEvelopment of Longitudinal micrOscopic traffic Processes in a Systemrelevant environment
PeMS	California Freeway Performance Measurement System
PHF	peak hour factor
PLC	pinned localised cluster
pMFT	paradisiacal mean-field theory
PMT	total person miles travelled
PQM	point-queue model
PRT	perception-reaction time
PSD	phase-space density
PVM	Parallel Virtual Machines
PW	Payne-Whitham
QoS	quality of service
RD	robust distance
RDG	relative duality gap
RMI	Remote Method Invocation
RP	recurrence plots
RQA	recurrence quantification analysis
SCF	sustainable cost function
SFI-TCA	stochastic Fukui-Ishibashi traffic cellular automaton
Simone	Simulation model of Motorways with Next generation vehicles
SLD	single inductive loop detector
SMARTEST	Simulation Modelling Applied to Road Transport European Scheme Tests
SMS	space-mean speed
SOC	self-organised criticality
SOMF	site-oriented mean-field theory
SP	synchronised-flow pattern
SQM	spatial-queue model
SSEP	symmetric simple exclusion process
SSNN	state-space neural networks
STA	static traffic assignment
STARCHILD	Simulation of Travel/Activity Responses to Complex Household Interactive Logistic Decisions
START/SITTER	Systeem Trafiek op Autosnelwegen Reële Tijd/ Système Intelligent Trafic en Temps Réel
STCA	stochastic traffic cellular automaton
STCA-CC	stochastic traffic cellular automaton with cruise control

SUE	stochastic user equilibrium
SUMO	Simulation of Urban MObility
SUV	sport utility vehicle
T^2-TCA	Takayasu-Takayasu traffic cellular automaton
TASEP	totally asymmetric simple exclusion process
TCA	traffic cellular automaton
TCS	traction control system
TDF	travel demand function
TISEAN	TIme SEries ANalysis
TMC	Traffic Message Channel
TMS	time-mean speed
TOCA	time-oriented traffic cellular automaton
TRANS	TRAffic Network Simulator
TRANSIMS	TRansportation ANalysis and SIMulation System
TRB	Transportation Research Board
TRP	thresholded recurrence plot
TSA	time series analysis
TSG	triggered stop-and-go traffic
UDM	ultra-discretisation method
UMTS	Universal Mobile Telecommunications System
VDR-TCA	velocity-dependent randomisation traffic cellular automaton
VDT	total vehicle distance travelled
VHT	total vehicle hours travelled
VI	variational inequality
VIP	video image processor
VMS	variable message sign
VMT	total vehicle miles travelled
VOT	value of time
WLT	wireless location technology-based
WORE	write once, run anywhere
WSP	widening synchronised-flow pattern
WYA	whole year analysis

A.2 List of symbols

Traffic flow theory

a_i	the acceleration of vehicle i
C	the number of substreams in a traffic flow
η	the efficiency of a road section (according to Chen et al, [Che01b])
ΔT_i	the time taken by vehicle i to travel the distance ΔX
ΔX_i	the distance travelled by vehicle i during the time interval ΔT
E	the efficiency of a road section (according to Brilon, [Bri00])
F	the free-flow curve in three-phase traffic theory
g_{s_i}	the space gap of vehicle i
$g_{\mathrm{s}_i}^{\mathrm{l,b}}$	the space gap at the left-back of vehicle i
$g_{\mathrm{s}_i}^{\mathrm{l,f}}$	the space gap at the left-front of vehicle i
$g_{\mathrm{s}_i}^{\mathrm{r,b}}$	the space gap at the right-back of vehicle i
$g_{\mathrm{s}_i}^{\mathrm{r,f}}$	the space gap at the right-front of vehicle i
g_{t_i}	the time gap of vehicle i
$g_{\mathrm{t}_i}^{\mathrm{l,b}}$	the time gap at the left-back of vehicle i
$g_{\mathrm{t}_i}^{\mathrm{l,f}}$	the time gap at the left-front of vehicle i
$g_{\mathrm{t}_i}^{\mathrm{r,b}}$	the time gap at the right-back of vehicle i
$g_{\mathrm{t}_i}^{\mathrm{r,f}}$	the time gap at the right-front of vehicle i
$\overline{h}_{\mathrm{s}}$	the average space headway
h_{s_i}	the space headway of vehicle i
$h_{\mathrm{s}_i}^{\mathrm{l,b}}$	the space headway at the left-back of vehicle i
$h_{\mathrm{s}_i}^{\mathrm{l,f}}$	the space headway at the left-front of vehicle i
$h_{\mathrm{s}_i}^{\mathrm{r,b}}$	the space headway at the right-back of vehicle i
$h_{\mathrm{s}_i}^{\mathrm{r,f}}$	the space headway at the right-front of vehicle i
$\overline{h}_{\mathrm{t}}$	the average time headway
h_{t_i}	the time headway of vehicle i
$h_{\mathrm{t}_i}^{\mathrm{l,b}}$	the time headway at the left-back of vehicle i
$h_{\mathrm{t}_i}^{\mathrm{l,f}}$	the time headway at the left-front of vehicle i
$h_{\mathrm{t}_i}^{\mathrm{r,b}}$	the time headway at the right-back of vehicle i
$h_{\mathrm{t}_i}^{\mathrm{r,f}}$	the time headway at the right-front of vehicle i
J	the wide-moving jam line J in three-phase traffic theory
k	the density
k_c	the density of the c-th substream in a traffic flow
k_{c}	the critical density
k_{crit}	the critical density

k_{j}	the jam density
k_{jam}	the jam density
k_{max}	the jam density
k_l	the density in lane l
k_{out}	the density associated with the queue discharge capacity
$k(t)$	the density at time t
K	the length of a measurement region (i.e., a certain road section)
K_{ld}	the length of a detection zone
$\overline{l}$	the average length of a vehicle
l_i	the length of vehicle i
L	the number of lanes on a road
N	the number of vehicles in a measurement region
N_l	the number of vehicles in the measurement region in lane l
$N_l(t)$	the number of vehicles in the measurement region in lane l at time t
$N(t)$	a cumulative count function
$\widetilde{N}(t)$	a smooth approximation of $N(t)$
$\overline{o}_{\text{t}}$	the average on-time of a set of vehicles
o_{t_i}	the on-time of vehicle i
$o_{\text{t}_{i,l}}$	the on-time of vehicle i in lane l
q	the flow
$\overline{q}_{\|15}$	the peak flow rate during one quarter hour within an hour
$\overline{q}_{\|60}$	the average flow during the hour with the maximum flow in one day
q_{b}	a background flow
q_c	the flow of the c-th substream in a traffic flow
q_{c}	the capacity flow
q_{cap}	the capacity flow
$q_{\text{e}}(k)$	an equilibrium relation between the flow and the density
q_l	the flow in lane l
q_{max}	the capacity flow
q_{out}	the outflow from a (wide-moving) jam, the queue discharge capacity
$q(t)$	the flow at time t
ρ	the occupancy
ρ_i	the occupancy time of vehicle i
ρ_l	the occupancy in lane l
R_{s}	a spatial measurement region at a fixed time instant
R_{t}	a temporal measurement region at a fixed location
$R_{\text{t,s}}$	a general measurement region
σ_{s}^2	the statistical sample variance of the space-mean speed

σ_{t}^2	the statistical sample variance of the time-mean speed
S	the synchronised-flow region in three-phase traffic theory
τ_i	the reaction time of vehicle i's driver
t_*	a time instant
T_{mp}	the duration of a measurement period
$T(t_0)$	the experienced dynamic travel time, starting at time instant t_0
$\widetilde{T}(t_0)$	the experienced instantaneous travel time, starting at time instant t_0
$\overline{v}_{\text{c}}$	the capacity-flow speed
$\overline{v}_{\text{cap}}$	the capacity-flow speed
$\overline{v}_{\text{ff}}$	the free-flow speed
v_i	the speed of vehicle i
$v_{i,l}$	the speed of vehicle i in lane l
$v_{i,l}(t)$	the speed of vehicle i in lane l at time t
v_{max}	the maximum allowed speed (e.g., by an imposed speed limit)
$\overline{v}_{\text{s}}$	the space-mean speed
$\overline{v}_{\text{s}_c}$	the space-mean speed of the c-th substream
$\overline{v}_{\text{s}_\text{e}}(\overline{h}_{\text{s}})$	an equilibrium relation between the SMS and the average space headway
$\overline{v}_{\text{s}_\text{e}}(k)$	an equilibrium relation between the SMS and the density
$\overline{v}_{\text{s}_\text{e}}(q)$	an equilibrium relation between the SMS and the flow
$\overline{v}_{\text{sust}}$	the sustained speed during a period of high flow
$\overline{v}_{\text{t}}$	the space-mean speed
$\overline{v}_{\text{t}_c}$	the time-mean speed of the c-th substream
$v(t,x)$	the local instantaneous vehicle speed at time instant t and location x
w	the characteristic/kinematic wave speed (of a wide-moving jam)
x_*	a location
x_i	the longitudinal position of vehicle i

Traffic flow models

a_{max}	the maximum acceleration in the IDM
$c(k)$	the sound speed of traffic
$C(q)$	the economical cost associated with the travel demand q
$\Delta f(x)$	the forward difference operator applied to $f(x)$
Δk	the difference in density up- and downstream of a shock wave
Δq	the difference in flow up- and downstream of a shock wave
ΔT	the size of a time step in a numerical discretisation scheme
ΔX	the width of a cell in a numerical discretisation scheme
D_j	a destination zone j
ϵ	a small diffusion constant for the viscosity ν
$g_{\text{s}}^*(v_i, \Delta v_i)$	the effective desired space gap in the IDM
κ	a kinetic coefficient related to τ, k, and Θ
k_{t}	the partial derivative of $k(t,x)$ with respect to time
k_{x}	the partial derivative of $k(t,x)$ with respect to space
$\widetilde{k}(t,x,\overline{v}_{\text{s}})$	the phase-space density at (t,x) associated with SMS $\overline{v}_{\text{s}}$
$\widetilde{k}_{\text{t}}$	the partial derivative of $\widetilde{k}(t,x)$ with respect to time
$\widetilde{k}_{\text{x}}$	the partial derivative of $\widetilde{k}(t,x)$ with respect to space
λ	the sensitivity to the stimulus in a car-following model
	the arrival rate at a server in queueing theory
μ	the service rate of a server in queueing theory
$\nabla f(x)$	the backward difference operator applied to $f(x)$
	the gradient vector of $f(x)$
ν	the kinematic traffic viscosity coefficient
ν_{W}	a parameter in Whitham's sound speed of traffic
O_i	an origin zone i
π	the probability of overtaking (as opposed to slowing down)
P	the traffic pressure
P_{x}	the partial derivative of the traffic pressure with respect to space
$P(t,x,\overline{v}_{\text{s}})$	the distribution of the vehicles with SMS $\overline{v}_{\text{s}}$ at (t,x)
q_{pc}	the practical capacity
q_{so}	travel demand associated with a system optimum
q_{ue}	travel demand associated with a user equilibrium
S	a traffic state in the human-kinetic model
τ	a driver's reaction time
Θ	the variance of the speed

$\Theta_{\mathrm{e}}(k, \overline{v}_{\mathrm{s}})$	an equilibrium relation between the speed variance, the density, and the SMS
T	a travel time
T_{ff}	the travel time under free-flow conditions
T_{r}	a relaxation parameter (in Pipes' car-following model)
u	the velocity (in the context of a Navier-Stokes fluid)
v_{des}	the desired speed of drivers
$\overline{v}_{\mathrm{s_t}}$	the partial derivative of the space-mean speed with respect to time
$\overline{v}_{\mathrm{s_x}}$	the partial derivative of the space-mean speed with respect to space
$\overline{v}_{\mathrm{s_e}}(k, \Theta)$	an equilibrium relation between the SMS, the density, and the speed variance
$V()$	the optimal velocity function
w_{shock}	the speed of a shock wave

Cellular automata

$\mathcal{C}(0)$	a CA's initial configuration
$\mathcal{C}(t)$	a CA's global configuration at time step t
δ	a CA's local transition rule
G	a CA's global map
G^{-1}	a reversible CA's inverse global map
$K_{\mathcal{L}}$	the number of cells in one lane of a TCA's lattice
$\mathcal{L}$	a CA's lattice (e.g., $\mathbb{Z}^2$)
$\mathcal{N}_i$	the (partially) ordered set of cells in the neighbourhood of the i^{th} cell
$\lvert\mathcal{N}\rvert$	the number of cells in the neighbourhood of each cell
$\mathcal{O}^{-}_{\mathcal{C}(t)\lvert G^{-1}}$	the backward orbit of the configuration $\mathcal{C}(t)$ under G^{-1}
$\mathcal{O}^{+}_{\mathcal{C}(0)\lvert G}$	the forward orbit of the initial configuration $\mathcal{C}(0)$ under G
$\sigma_i(t)$	the state of the i^{th} cell at time step t
Σ	the set of all possible states a CA's cells can be in (e.g., $\mathbb{Z}_2$)
$\Sigma^{\mathcal{L}}$	the set of all possible global configurations of a CA
$\Sigma^{\mathcal{N}}$	the set of all possible configurations of a cell's neighbourhood
$\lvert\Sigma^{\Sigma^{\mathcal{N}}}\rvert$	the number of all possible rules of a CA
$\mathcal{T}_{\mathcal{C}(0)\lvert G}$	the trajectory/orbit of the initial configuration $\mathcal{C}(0)$ under G

Traffic cellular automata

$\wedge$	the logical binary conjunction operator ‘AND’
α	the entry rate of particles in the TASEP model
α_i	the anticipatory driving parameter of vehicle i
a	the acceleration capability of a vehicle in the KKW-TCA model
β	the exit rate of particles in the TASEP model
b	the deceleration capability of a vehicle in the KKW-TCA model
$b_i(t)$	the state of the brake light of vehicle i at time t in the BL-TCA model
δ	the probability for a particle to move to the right in the TASEP model
Δ_{acc_i}	the deterministic acceleration of vehicle i in the KKW-TCA model
ΔT	a TCA’s temporal discretisation
ΔV	a TCA’s speed discretisation
ΔX	a TCA’s spatial discretisation
D_0	a parameter for the synchronisation distance in the KKW-TCA model
D_1	a parameter for the synchronisation distance in the KKW-TCA model
D_i	the synchronisation distance of vehicle i in the KKW-TCA model
η_i	the stochastic acceleration of vehicle i in the KKW-TCA model
γ	the probability for a particle to move to the left in the TASEP model
$\overline{g}_{\text{s}}$	the average space gap
$g^*_{\text{s}_i}(t)$	the effective space gap of vehicle i at time t in the BL-TCA model
$g_{\text{s}_{\text{security}}}$	a security constraint for the space gap in the BL-TCA model
$\overline{g}_{\text{t}}$	the median time gap
$\overline{g}_{\text{t}_\text{s}}$	the safe time gap in the TOCA model
h	the interaction range of the brake light in the BL-TCA model
$\overline{h}_{\text{s}_\text{c}}$	the average space headway corresponding to the critical density
$\overline{h}_{\text{s}_\text{j}}$	the average space headway corresponding to the jam density
k_{g}	the global density of a TCA’s lattice
k_{l}	the local density of a TCA’s lattice
$K_{\mathcal{L}}$	the number of cells in one lane of a TCA’s lattice
$\mathcal{L}$	a TCA’s lattice
l_i	the length of vehicle i
$\overline{l}$	the average length of all vehicles on a TCA’s lattice
$L_{\mathcal{L}}$	the number of lanes in a TCA’s lattice
$M_{g_{\text{s}_i},v_i}$	the gap-speed matrix of the ER-TCA model
p	the slowdown probability in $[0, 1]$
p_0	the slow-to-start probability in $[0, 1]$
p_{a}	the acceleration probability in $[0, 1]$ in the KKW-TCA model

p_{a_1}	a parameter for the acceleration probability in the KKW-TCA model
p_{a_2}	a parameter for the acceleration probability in the KKW-TCA model
p_{acc}	the acceleration probability in $[0, 1]$ in the TOCA model
p_{b}	the braking probability in $[0, 1]$ in the BL-TCA model
	the deceleration probability in $[0, 1]$ in the KKW-TCA model
p_{d}	the slowdown probability in $[0, 1]$ in the BL-TCA model
p_{dec}	the deceleration probability in $[0, 1]$ in the TOCA model
p_{s}	the slow-to-start probability in $[0, 1]$ in the BJH-TCA model
p_{t}	the slow-to-start probability in $[0, 1]$ in the T^2-TCA model
$P_n(v)$	the probabilities of finding a space gap of n cells
	for a vehicle driving with speed v
q_{g}	the global flow of a TCA's lattice
q_{l}	the local flow of a TCA's lattice
t_{s_i}	the interaction horizon in the BL-TCA model
v_{des_i}	the desired speed of vehicle i in the KKW-TCA model
v_{p}	a parameter for the acceleration probability in the KKW-TCA model
$\overline{v}_{\text{s}_{\text{ff}}}$	the space-mean speed in the free-flow regime
$\overline{v}_{\text{s}_{\text{g}}}$	the global space-mean speed of a TCA's lattice
$\overline{v}_{\text{s}_{\text{l}}}$	the local space-mean speed of a TCA's lattice
$\xi(t)$	a random number in $[0, 1[$ drawn at time t from a uniform distribution
$x_i^{\text{l,b}}$	the longitudinal position of vehicle i's left-back neighbour
$x_i^{\text{l,f}}$	the longitudinal position of vehicle i's left-front neighbour
$x_i^{\text{r,b}}$	the longitudinal position of vehicle i's right-back neighbour
$x_i^{\text{r,f}}$	the longitudinal position of vehicle i's right-front neighbour

Data quality, travel time estimation, and reliability

α_{ld}	the slope of an SLD's threshold's linear function
ΔN	the number of vehicles between two measurement posts
Δt	the time needed to travel between two consecutive measurement posts
$\overline{\Delta t}$	the average time difference between a pair of $t(N)$ curves
d	the order of integration in an ARIMA model
Φ	the cumulative distribution function of the normal distribution
g	the g-factor for an SLD
k	the density
K	the length of a road section
	the distance between two measurement posts
K_{ld}	the length of an SLD

λ_{ld_i}	a threshold calculated for SLD i
l_j	the length of the j^{th} vehicle
$\bar{l}$	the average vehicle length
$\bar{l}_{\mathrm{c}}$	the average length of a car
$\bar{l}_{\mathrm{t}}$	the average length of a truck
m	the number of multiple imputations
m_{DCE}	the embedding dimension for delay coordinate embedding
μ	the mean of a distribution
$\hat{\mu}_{\mathrm{MCD}}$	a robust estimation of the mean of a distribution
n	the number of data points (observations)
N_{down}	the cumulative number of vehicles (downstream)
N_{up}	the cumulative number of vehicles (upstream)
o_{t_j}	the on-time of the j^{th} vehicle over an SLD
$\overline{o}_{\mathrm{t}_{\mathrm{c}_i}}$	the average on-time for a car at SLD i
$\overline{o}_{\mathrm{t}_{\mathrm{c}_{\mathrm{min}}}}$	the minimum average on-time for all cars at an SLD
$\overline{o}_{\mathrm{t}_{\mathrm{c}_{\mathrm{max}}}}$	the maximum average on-time for all cars at an SLD
p	the number of dimensions (variables)
	the order of autoregression in an ARIMA model
$p(\rho_i)$	an estimated probability density function of the occupancies ρ_i
$\chi^2_{p,0.975}$	a specified threshold for identifying outliers
q	the flow
	the order of the moving average in an ARIMA model
q_{down}	the flow (downstream)
q_{up}	the flow (upstream)
q_{c_i}	the number of cars driving by SLD i
q_{t_i}	the number of trucks driving by SLD i
ρ^*	a threshold occupancy
ρ_{c}	the critical occupancy
ρ_i	the occupancy of SLD i
$\mathbf{r}$	a point on the trajectory of a delay coordinate embedded time series
σ	the standard deviation of a distribution
Σ	the covariance matrix of a distribution
$\hat{\Sigma}_{\mathrm{MCD}}$	a robust estimation of the covariance matrix of a distribution
S_1	the number of samples with zero occupancy for the DSA algorithm
S_2	the number of samples with a high occupancy for the DSA algorithm
S_3	the entropy of the occupancy samples for the DSA algorithm
τ_{DCE}	the delay for delay coordinate embedding
τ_{ld_i}	a threshold calculated for SLD i

$\tau_{\mathrm{ld}_{\mathrm{min}}}$	the minimum threshold for all SLDs
$\tau_{\mathrm{ld}_{\mathrm{max}}}$	the maximum threshold for all SLDs
t_{down}	the inverse of the cumulative number of vehicles (downstream)
t_{up}	the inverse of the cumulative number of vehicles (upstream)
$\overline{T}$	the average experienced dynamic travel time of all the vehicles
$T(t_0)$	the experienced dynamic travel time starting at t_0
$\widetilde{T}(t_0)$	the experienced instantaneous travel time starting at t_0
$\widetilde{T}_{\mathrm{ff}}$	the experienced instantaneous travel time under free-flow conditions
T_{DSA}	the length of the aggregation period for the DSA algorithm
T_{mp}	the duration of a measurement period
v_j	the speed of the j^{th} vehicle
$\overline{v}_{\mathrm{ff}}$	the free-flow speed
$\overline{v}_{\mathrm{t}_i}$	the time-mean speed of all vehicles driving by SLD i
x_i	a univariate sample taken from a distribution
$\mathbf{x}_i$	a multivariate sample taken from a distribution
$\overline{x}_{\mathrm{med}}$	the median of a distribution
x_{ref}	a reference measurement used for interpolation
z_i	the z-score for identifying outliers

Dynamic traffic assignment based on cellular automata

C_{β_i}	the schedule delay cost for agent i when arriving too early
C_{γ_i}	the schedule delay cost for agent i when arriving too late
C_{μ_i}	the cost associated with agent i's queueing time
C_{sd_i}	the schedule delay costs for agent i
C_{T_i}	the cost associated with agent i's travel time
C_{total_i}	the generalised travel cost associated with agent i
ϵ_i	a stochastic error term in the route logit model
μ	a dispersion factor in the route logit model
μ_i	the waiting time of agent i in a queue at an origin
N	the number of agents in a disaggregated population
p_i	the logit probability for selecting a certain route
$\mathcal{R}$	the set of all available routes between an OD-pair
$\mathcal{R}_{\mathrm{pre}}$	the set of predefined feasible routes
$t_{\mathrm{demand,end}}$	the end of the demand generation period
$t_{\mathrm{demand,start}}$	the start of the demand generation period
$t_{\mathrm{departure}_i}$	the departure time of agent i
t_{PAT_i}	agent i's preferred arrival time
T_i	the travel time of agent i
$U_{\mathrm{departure}}$	a utility function for agent i's time of departure
U_i	the utility of route i

Appendix B

TCA+ Java™ software

As already briefly mentioned in Chapter 4, all simulations therein were performed by means of our Traffic Cellular Automata + *software [Mae04d]. It was developed for the Java™ Virtual Machine (JVM), and can be downloaded*[1] *from:*

> `http://smtca.dyns.cx`

The software is also referenced on the Traffic Forum[2] *(see Section 'Links', Subsection 'Online Traffic Simulation or Visualization (Java Applets)', Item 'Java (Swing) application for several cellular automata models').*

In this appendix, we summarise our rudimentary TCA+ software. We start with an overview of its features, explain how to run the software, and conclude with some technical details with respect to the implementation of its code base.

B.1 Overview and features

The TCA+ software package's goal is two-fold: on the one hand, it provides an *intuitive didactical tool* for getting acquainted with the concept of single-lane traffic cellular automata models. On the other hand, it provides a rich enough code base to perform hand-tailored *simulation experiments*, as well as giving insight into the details of programming TCA models.

In a nutshell, our software considers one-dimensional traffic cellular automata with periodic boundary conditions, i.e., vehicles driving on a unidirectional circular road. Different sets of rules can be chosen, and for each set its parameters (e.g., stochastic

[1] From May 2002 until June 2005, the software has been downloaded some 800 times, of which one third appears to be traffic coming from search engines' indexing robots.

[2] `http://www.trafficforum.org`

noise) can be changed at run time. Both local and global measurements can be performed on the lattice by means of artificial loop detectors. A traffic light with cyclical red and green phases was also added, allowing to study elementary queueing behaviour. In the software, we have implemented the TCA models listed in Table B.1.

TCA model	*Refer to Section*	*TCA model*	*Refer to Section*
CA-184	4.3.1.1	stochastic T^2-TCA	4.3.3.1
DFI-TCA	4.3.1.2	VDR-TCA	4.3.3.3
STCA	4.3.2.1	VDR-CC-TCA	4.3.3.3
STCA-CC	4.3.2.2	TOCA	4.3.3.4
SFI-TCA	4.3.2.3	MC-STCA	4.4.1
TASEP	4.3.2.4	HS-TCA	4.4.2.1
ER-TCA	4.3.2.5	BL-TCA	4.4.2.2
deterministic T^2-TCA	4.3.3.1	KKW-TCA	4.4.2.3

Table B.1: All TCA models implemented in our TCA+ software, accompanied by references to the respective sections in Chapter 4 where they are extensively discussed.

In Figure B.1, we show a screenshot of the main graphical user interface (GUI). As can be judged from the image, the TCA+'s GUI is rather huge, spanning approximately 1400x1200 pixels (scrollbars are automatically placed if it does not fit on the screen). It consists of several panels:

- a scrolling time-space diagram containing vehicle trajectories and an animation of the road situation,
- a panel containing some simulation statistics,
- several simulator controls,
- and scrolling loop detector plots and plots of the (k,q), (k,$\overline{v}_s$), and (q,$\overline{v}_s$) diagrams.

In the following paragraphs, we describe each of these features in more detail. Note that there currently are two versions of the GUI: a standard version for all the single-cell TCA models, and a modified multi-cell TCA version with limited functionality (mainly for creating coloured tempo-spatial diagrams).

Vehicle animation

Looking at the time-space diagram in the upper-left panel, we can discern the individual vehicle trajectories, as well as the typical backwards-travelling shock waves of congestion. In this scrolling diagram, the time axis goes from the left to the right, while the space axis goes from the bottom to the top (and is a one-to-one mapping of

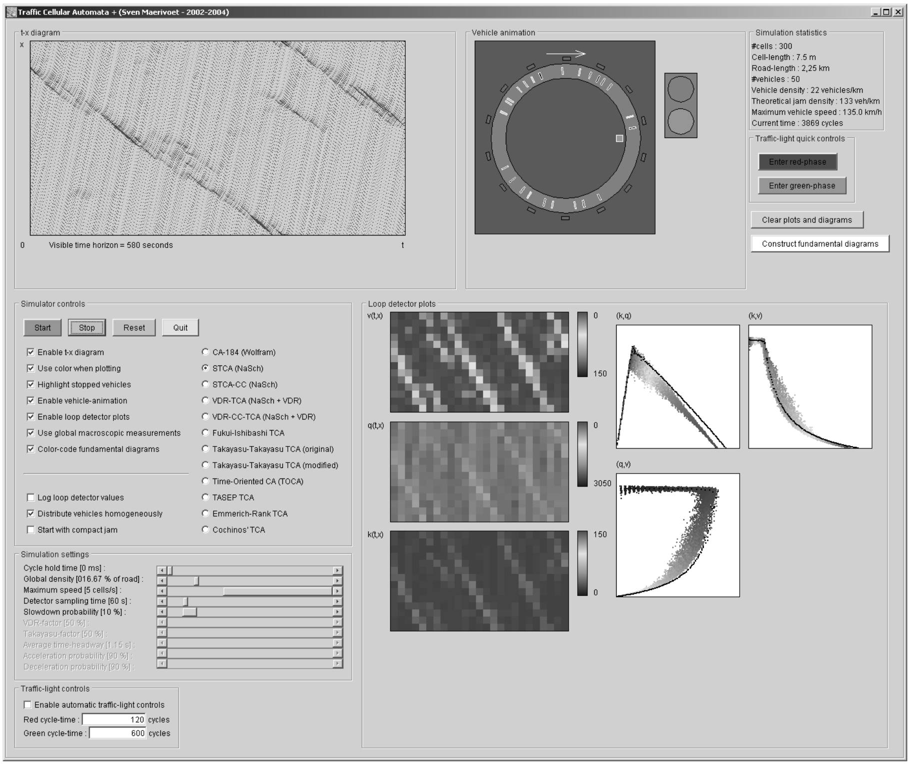


Figure B.1: A screenshot of the TCA+'s main graphical user interface (GUI) for single-cell TCA models. The GUI is rather huge, spanning approximately 1400x1200 pixels, consisting of several panels: a scrolling time-space diagram containing vehicle trajectories, an animation of the road situation, a panel containing some simulation statistics, several simulator controls, scrolling loop detector plots and plots of the (k,q), $(k,\overline{v}_s)$, and $(q,\overline{v}_s)$ diagrams.

the consecutive cells on the ring road). Each pixel here corresponds to a unique cell of the simulator and each vehicle is coloured with a certain shade of yellow (in order to easily distinguish between different neighbouring vehicles). There is also a setting available that allows stopped vehicles to be coloured red. In the upper-middle panel, the actual geometrical configuration of the ring road is depicted. This allows us to view the current physical situation on the road, i.e., the positions of all the vehicles. Each vehicle can be coloured with a certain shade of yellow (the same as in the time-space diagram). The current phase of the traffic light is also shown, as well as the positions of all the loop detectors: their positions are indicated by the small purple boxes alongside the road. The small green box indicates the position of the traffic light, with vehicles travelling in clockwise fashion.

Simulation statistics

In the upper-right panel, we can find the length of the ring road (expressed in the number of cells), the number of vehicles currently in the simulator, the global vehicle density, and the current time step. There is also a small panel that allows to quickly set the status of the traffic light to either red or green.

Simulator controls and settings

The middle-left panel contains buttons for starting, stopping (i.e., pausing), resetting, and quitting the simulator. Several preferences can also be specified, i.e., whether or not to activate several panels containing the simulator's output. There is also the possibility to log the measurements from the loop detectors to a default file (called *detector-values.data*). And finally, the type of traffic cellular automaton (i.e., its rule set) can also be selected from a list, specified by radio control buttons.

Note that there are several initial conditions possible for each density level: it is possible to start with a homogeneous state (all vehicles are spaced evenly), with a compact superjam of vehicles that are all stopped, or with a random initialisation (see also the introduction of Section 4.3).

If the simulation goes (visually) too fast, the cycle hold time can be increased, thereby freezing the simulation for a while between two consecutive time steps. Besides this, the ring road's global density and the vehicles' maximum speed can be specified. The sampling time for the artificial loop detectors can be adjusted (to increase or smooth out fluctuations). And finally, all probabilities can be adjusted between 0% and 100% in incremental steps of 1%.

The red and green cycle times for the traffic light can be specified, such that the light can operate automatically, thereby inducing artificial queues at regular intervals. One can also control the traffic light manually (enabling the red or green phase) using the small upper-right panel; but if applied, the traffic-light controls override these manual settings.

Plots of macroscopic measurements

The software has the ability to extract both local and global macroscopic flow measurements from several uniformly road-side placed loop detectors which record flows, densities, and space-mean speeds.

The three large coloured regions in the middle panel represent the measured (and averaged) values of the local flows, local densities, and local space-mean speeds of the loop detectors. Pair-wise correlating these values, results in the plots of the (k,q), $(k,\overline{v}_{\text{s}})$, and $(q,\overline{v}_{\text{s}})$ diagrams in the lower-right panel. The coloured dots indicate locally obtained measurements, whereas the black dots represent globally obtained ones.

Note the small button that allows to construct these diagrams: when it is pressed, the global density is incrementally increased from 0% to 100%, each time adding a single vehicle to the ring road. The simulation is then ran for a certain amount of time and the measurements from all the loop detectors are recorded. When all densities are processed (an indicator of the total time left is shown), the diagrams should be clearly visible in the loop detector plots in the lower-right panel.

B.2 Running the software

When visiting the website mentioned in the introduction of this appendix, there are two options for downloading the software. One is by downloading the *compiled classes*, whereas the other is to download the programme's *source code*. Once the compiled software has been downloaded, it is relatively easy to start the graphical user interface. Considering the single-cell setup GUI, the software is ran by executing the following command:

```
java -jar tca.jar
```

> Note that a Java™ Development Kit (JDK) (preferably Sun's at `http://java.sun.com`) should be installed. Furthermore, due to a change in the multi-threading implementation of the Java™ Swing™ API, it appears that only JDK/JRE 1.3.1 is suitable.

B.3 Technical implementation details

It should be noted that the software is not implemented as an applet, but instead as a full Java™ application because it uses Swing™ components that are not standard supported by most browsers (at least not without installing a necessary plugin). The source itself logically consists of three different parts:

- the TCA engine with different rule sets,
- the graphical user interface,
- and a whole range of predefined experiments.

The geometrical configuration used in the single-cell TCA engine is a unidirectional ring road with a single lane. Vehicles are located in cells of $\Delta X = 7.5$ m and can have speeds of 0 to 5 cells/time step (corresponding to a maximum speed of 135 km/h). One iteration in the simulation corresponds to a time step of $\Delta T = 1$ s.

A number of artificial loop detectors are uniformly placed alongside the road, aggregating various macroscopic traffic measurements (i.e., flows, densities and space-mean speeds). In the GUI, global measurements on the entire lattice are performed according to the methodology explained in Section 4.2.3.2, whereas local measurements are performed according to Section 4.2.3.1. Note that for the TCA software itself, it is also possible to perform local measurements using a detector of unit length, according to the methodology explained in Section 4.2.3.3.

Besides the standard single-cell GUI and the limited multi-cell GUI, there also exist some predefined experiments. These allow to create the (k,q), $(k,\overline{v}_{\text{s}})$, and $(q,\overline{v}_{\text{s}})$ diagrams, histograms of the vehicles' speeds, space gaps, and time gaps, as well as several order parameters (density correlations, nearest neighbours, and an inhomogeneity measure that compares the locally recorded densities to the current global density).

Inside the TCA+ software, several packages are available:

- `tca.base` containing the definitions of cells, global states, loop detectors, and the traffic cellular automaton's lattice,
- `tca.automata` containing implementations of all the TCA models mentioned in Section B.1,
- `tca.simulator` containing the classes related to the single-cell and multi-cell GUIs,
- `tca.experiments.fundamentaldiagrams`,
  `tca.experiments.histograms`,
  and `tca.experiments.orderparameters` containing setups for the previously mentioned experiments.

Appendix C

Some thoughts on obtaining a PhD

In this appendix, we share our thoughts with respect to the process of obtaining a PhD degree in the Faculty of Engineering at the Katholieke Universiteit Leuven. We focus on what we believe to be the requirements of a PhD candidate, after which we give some reflections on the hassles in the doctoral training programme every graduate student is expected to participate in.

C.1 Preliminaries

- This note was written in the pluralis majestatis for aesthetic purposes, furthermore, we use the words *'must'* and *'should'* liberally and interchangeably, but a negative connotation is never implied.
- The **vision** set forth in this note, is mainly based on the process of obtaining a doctoral degree in the Faculty of Engineering (Katholieke Universiteit Leuven). Other faculties focus on other aspects, putting more or less emphasis on some of the points touched upon in our train of thought.

"*Be passionate about your scientific field, feel driven by your research.*"

C.2 Requirements of a PhD candidate

PhD candidates are those among the few people in the world who can do research at any time. They can pursue own interests, noble causes, . . . all in the name of research.

But, at the same time, this great **freedom** also implies **a sense of responsibility** which we think is a **necessary** ingredient !

Consider the following implicit minimal requirements (*'implicit' meaning that a mature PhD candidate will take the following points for granted*); a PhD candidate:

- must **be aware of the structure** in which he/she is expected to work/operate; here we are talking on the level of the research group, the department, and to a lesser extent the university's structure (more specifically about financial possibilities/opportunities for funding, et cetera), be **concerned** with his/her working environment.
- must, to a high degree, **be able to work independently**.
  $\Rightarrow$ Additionally, a PhD candidate should not be afraid of talking to other people, in fact, he/she should consult others if necessary (*note here that one of the roles for the promotor is to point out possible persons*).
- closely related to the previous point, a PhD candidate must **construct a research network** in which he/she actively creates and manages contacts at several levels (e.g., locally within the research group, globally with fellow researchers in other departments, and even at conferences et cetera).
  $\Rightarrow$ Conferences are mainly intended as a means to develop and sustain your network.
- must be able to coordinate a project, i.e., **have organisational skills** (with initial guidance if needed), and **have a (broad) sense of responsibility**, taking initiatives, et cetera. Behaving in a **professional** way is mandatory when interacting with other people.
- must **be enthusiastic about his/her research**, and **have ambition** (the optimal situation is when the candidate 'lives' by his research, thinks about it at the most odd occasions, is absorbed by the scientific field... but is still able to draw a line !).
- must **be interdisciplinary minded** (i.e., being interested in all kinds of knowledge, not only those within the own field of research), in contrast to this, a PhD candidate should be able to fluently handle the large doses of incoming information by selecting the relevant parts.
  $\Rightarrow$ This implies that a PhD candidate is supposed to be extremely curious: **he/she is like a sponge**, absorbing as much knowledge as possible. In this respect, a PhD candidate should strive for an almost encyclopaedial knowledge of literature.

Furthermore, we believe a PhD candidate:

- should **have a critical attitude** towards science, research, and triviality,
- should **have a global world view**, and his/her position in it,
- should **be creative** about his/her research,
- should **adopt his/her own research style**, create a personal profile, have a unique character (as opposed to the default grey mass in which most PhD candidates seem to dissolve).

Note that independence comes in at least two degrees: taking initiatives, coming (1) from the promotor and (2) from oneself. Furthermore, we acknowledge the fact that there are different kinds of doctoral students, with respect to being able to work independently. In the case where the PhD candidate needs guidance, this should be initially provided by (1); the promotor is not obliged to guide the candidate in persona, but should at least be obliged to **provide** the means for guidance.

With respect to this last item, we partially agree with the K.U.Leuven's *'Profile of a good promotor'*[1]: a promotor can only guide a limited number of doctoral students; if this number increases, this requires other means of guidance (e.g., post-doctoral researchers). The exception we make, is when the PhD candidates are able to function **truly independently**. But do note that in any other case, the primary role of the promotor is to 'take care' of the PhD candidate, such that in the end, the same results are achieved as if the candidate was working independently.

Universities are not large scale PhD factories; instead, PhD's craft themselves to a certain degree, with a careful eye for detail.

To end, we would like to draw some attention to the following question:

"What is the initial motivation for starting as a PhD candidate ?"

We believe each individual should think about this question at one time or another, and be able to give a definite answer for him/herself. Doing a PhD is certainly not a 'nine to five job', as it entails a whole philosophy in a certain sense. Obtaining the PhD degree is a daunting task, in which the candidate learns to plan over a course of three to six years, getting more mature in the process.

[1]See `http://www.kuleuven.ac.be/doctoreren/profiel.htm` for more information.

C.3 About the doctoral training programme (DOCOP)

(note: DOCOP means 'DOCtoraatsOPleiding' in Dutch)

Consider the original intent of the regulations:

1. *"The first goal of the DOCOP regulation is to broaden the knowledge of the PhD candidate and to immerse him/her in the field of research."*
2. *"As a secondary goal of the regulations, they allow the process of obtaining the doctoral degree more efficiently by providing better guidance and tracking abilities."*
3. *"The regulations also aspire to play a supportive role, in that they want to prepare the PhD candidate for his/her later professional functioning."*
4. *"They furthermore stimulate the research dynamics and contribute to a doctoral culture."*

Putting these intentions into practice, the doctoral training consists of the following requirements that reflect the expectations towards a *'good PhD candidate'*:

- publications at an international level,
- giving and following of doctoral seminars,
- actively participating to international congresses,
- and reporting on the doctoral research on a regular base.

In contrast to this, we claim that:

- this regulation should, in principle, be **redundant**, because a 'good' PhD candidate:
  - will spontaneously follow courses, go to conferences, publish in journals, et cetera, when it's interesting to him/her, it should not necessarily be directly related to the field of research (although it can sometimes be preferred),
  - should not be obliged to take doctoral exams (except of course the thesis defence).
    $\Rightarrow$ This means that the famous requirement of 'following a doctoral course with evaluation' is dismissed on the grounds that when a course (or part of it) is interesting to the researcher, he/she will already try to master it, *without* the need for an evaluation. There are many more courses, and the fact that the DOCOP rules stipulate that only one course is necessary, also reflects the artificial sounding to this rule.

- the DOCOP rules **don't** guarantee the fulfillment of the second intent: guidance is not provided at all, tracking the research progress is done in a way that is too artificial (grading),

Note that our claims are based on the preposition that many things that are *stipulated* in the original DOCOP regulation, are in fact expected to be *automatically satisfied* by the PhD candidate. This means that we ask the following central question:

"Should a PhD candidate be enforced to obtain these goals ?"

We say **no**, because in our opinion, the other PhD candidates are 'unworthy' to obtain their doctoral degree. **In order to receive the PhD title, one has to earn it**. This last remarks clearly goes beyond the requirement of a thesis with an accompanying dissertation. From our point of view, we believe this is the original motivation from which the DOCOP regulation took root. However, the current regulation has a pertinent fixation on the grading system, and not as much appreciation of the qualitative content.

Addendum:

Other universities base their doctoral training requirements on more or less the same philosophies:

- "*To deepen the PhD candidate's knowledge of the discipline and scientific field and to broaden his/her knowledge outside this discipline.*"
  — Universiteit Antwerpen
- "*To deepen and broaden the PhD candidate's knowledge and skills.*"
  — Universiteit Gent
- "*To stimulate a high research quality, and to provide a increased level of support for PhD candidates.*"
  — Vrije Universiteit Brussel
- "*To provide profound, systematic, and functional guidance for PhD candidates, to provide a thorough education in all aspects of the research methodology, to learn to work independently.*"
  — Universiteit Hasselt

Note that at the Universiteit Gent, following the doctoral training is *advised*, but *not made obligatory*. And at the Vrije Universiteit Brussel, they internally challenged the use of the doctoral training programme, changing it from a mere administrative task of obtaining points, to a more dedicated guidance of PhD candidates (by means of peer support, knowledge management, ...).

Appendix D

Nederlandse samenvatting

Modelleren van Verkeer op Autosnelwegen:
State-of-the-Art, Numerieke Data Analyse,
en Dynamische Verkeerstoedeling

Hoofdstuk 1: Inleiding

Gezien de huidige problematiek omtrent de drukte in het wegverkeer in steden en landen, wordt het met de dag duidelijker dat we filevorming niet volledig kunnen oplossen. Alles is echter niet verloren, daar we kunnen proberen om de toestand te verzachten door de ritten zo comfortabel mogelijk te maken. Het blijft niettemin een zware taak om filevorming op globale schaal aan te pakken, iets wat een geïntegreerde aanpak vereist, en waarbij verschillende regeltechnieken gecombineerd worden. Voorbeelden van dit laatste zijn de geavanceerde verkeersbeheerssystemen (ATMS) zoals dynamische route geleiding, toeritdosering, snelheidsharmonisatie, getijdengolven, …, en beleidsmaatregelen die beslist worden door (lokale) overheden. Deze maatregelen worden uitgevoerd onder de vorm van bijvoorbeeld rekeningrijden, beter en goedkoper openbaar vervoer, … tot zelfs sommige bizarre voorstellen zoals dubbeldek-autosnelwegen, voorgesteld door de liberale senator Jean-Marie De Decker. Met dit laatste wil men de capaciteit van het wegennet uitbreiden en zo de files verminderen. In tegenstelling tot sommige van deze extreme maatregelen, zou vlotter verkeer bewerkstelligd moeten worden door gebruik te maken van het bestaande wegennet, zonder de noodzaak om nieuwe weginfrastructuur aan te leggen (let wel dat lokale aanpassingen nog steeds toegestaan zijn).

Een eerste luik van ons onderzoek is het aanreiken van een goede achtergrond met betrekking tot het modelleren van wegverkeer. Tot op heden is er nog vaak een regelmatige verwarring tussen verkeersdeskundigen enerzijds en beleidsmakers anderzijds, wanneer het aankomt op transportplanningsmodellen en de rol die verkeersstroommodellen daarin spelen. Het literatuuroverzicht in dit werk is uniek omdat het een redelijk volledig geheel vormt. Hierdoor wordt de noodzaak weggenomen om te gaan kijken in de zoo van artikels en notaties die er op dit moment heerst.

Een tweede luik van ons onderzoek is gericht op de numerieke data analyse van ruwe verkeersmetingen (tellingen). We geven onderzoekers middelen om statistische uitschieters op te sporen, om op een snelle manier structurele en incidentele storingen van detectors te beoordelen, om reistijden te schatten op een off-line manier, gebaseerd op ruwe cumulatieve tellingen, en om een visuele voorstelling van de dynamica van verkeersstromen in tijd en ruimte te verkrijgen.

Het derde en laatste luik van ons onderzoek komt voort uit de dynamische verkeerstoedeling. De huidige evolutie in het wetenschappelijk domein is om op een endogene wijze zowel keuze van vertrektijdstip als route te combineren. We stellen een duidelijke methode voor die toelaat om beide problemen op sequentiële wijze aan te pakken, op basis van een verkeersstroommodel dat uitgewerkt wordt als een computationeel efficiënte cellulaire automaat. Deze efficiëntie wordt verder nog verbeterd door het concept parallelliseren door middel van gedistribueerd rekeken.

Hoofdstuk 2: Verkeersstroomtheorie

Omwille van de grote diversiteit van het vakgebied (ingenieurs, fysici, wiskundigen, . . .), is een van de belangrijkste doelstellingen van dit hoofdstuk om zowel een logische als consistente terminologie te definiëren. We geloven er sterk in dat een dergelijke standaard een noodzaak is, zeker wanneer het er op aankomt om een graad van orde te creëren in de 'zoo van notaties' die op dit moment ons insziens bestaat (voor een volledig overzicht van alle afkortingen en notaties die we voorstellen, verwijzen we de lezer naar appendix A).

In dit hoofdstuk geven we een overzicht van wat op dit moment de state-of-the-art is met betrekking tot verkeersstroomtheorie. We beginnen met een korte geschiedenis, waarna we de tussentijden en -ruimtes als microscopische karakteristieken van wegverkeer geven, met dichtheden en intensiteiten als hun macroscopische tegenhangers. Verdergaand, bespreken we enkele performantie-indicatoren die ons toelaten om de kwaliteit van de verkeersoperaties te beoordelen. Hierbij besteden we ook kort aandacht aan het afleiden van reistijden op basis van cumulatieve curves. Het laatste deel van dit hoofdstuk bespreekt enkele relaties tussen verkeerskundige karakteristieken, door middel van het fundamenteel diagram, en werpt een licht op verschillende standpunten inzake de oorzaken van filevorming.

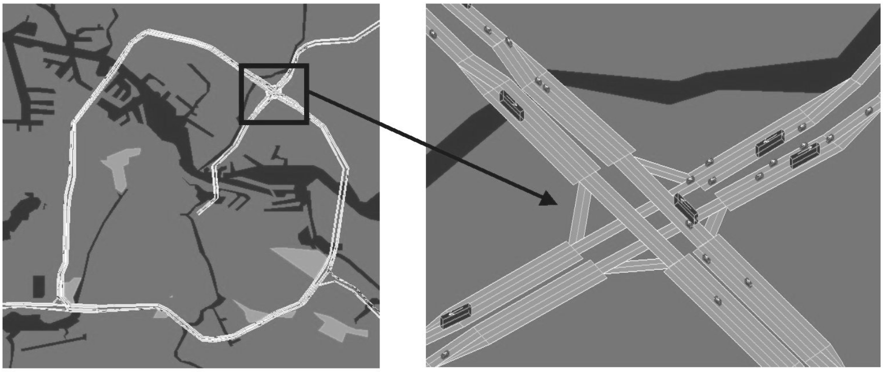
Hoofdstuk 3: Transportplannings- en verkeersstroommodellen

Daar waar het vorige hoofdstuk ging over de notaties en terminologie die geassocieerd worden met karakteristieken van verkeersstromen, legt dit hoofdstuk de nadruk op de verschillende verkeersmodellen die er bestaan.

Gedurende ons onderzoek merkten we vaak een verwarring op tussen verkeerskundigen aan de ene kant en beleidsmakers aan de andere kant. Om hieraan tegemoet te komen, gaan we dieper in op transportplanningsmodellen enerzijds en verkeersstroommodellen anderzijds. De eerste klasse is typisch werkzaam op een hoger niveau, waarbij gezinnen bepaalde beslissingen nemen, wat aanleiding geeft tot transport en het gebruik van de weginfrastructuur. De tweede klasse is typisch werkzaam op een lager niveau, waarbij expliciet het stromen van het verkeer op een wegennetwerk wordt beschreven; deze klasse wordt typisch als onderdeel in de eerste klasse gebruikt. Voor de transportplanningsmodellen bespreken we modellen voor het gebruik van land, zowel in een klassieke als moderne context, waarna we het traditionele vierstapsmodel (4SM) uit de doeken doen, gevolgd door een uitwerking van activiteiten-gebaseerd modelleren (ABM). Alvorens in te gaan op de stroommodellen, geven we nog een beknopt overzicht van enkele basisprincipes in het vakgebied rond de transporteconomie, waarbij we afsluiten met een discussie omtrent rekeningrijden. Vanaf dan gaat het hoofdstuk verder met gedetailleerde informatie over macroscopische, mesoscopische en microscopische stroommodellen (zie Figuur D.1 voor een voorbeeld van deze laatste klasse); van deze drie vereisen de microscopische modellen vaak ook meer rekenkracht omwille van de complexere interacties tussen individuele voertuigen, daar waar macro- en mesoscopische modellen eerder uitgaan van geaggregeerd gedrag (bijvoorbeeld zoals in samendrukbare vloeistoffen en gassen).

Onze doelstelling is niet om een volledig overzicht te geven, maar is het eerder onze intentie om de lezer een grondig gevoel te geven voor de verschillen tussen transportplannings- en verkeersstroommodellen. Omwille van de snelle vooruitgang gedurende het laatste decennium (en zelfs de laatste vijf jaren), probeert dit hoofdstuk zowel oudere, meer klassieke modellen, als de laatste ontwikkelingen in het vakgebied te bespreken. Merk op dat ons overzicht beperkt is, en bijvoorbeeld geen behandeling geeft van disciplines zoals verkeersregeling, milieumodellering, . . .

Tot op heden gebruiken veel transportplanningsbureaus statische modellen voor het evalueren van beleidsmaatregelen; de noodzaak tot dynamische modellen wordt meer en meer uitgesproken [Mae04b]. Zelfs na meer dan zestig jaar onderzoek op verkeerskundig vlak, blijft het debat omtrent wat nu de juiste modelleringsaanpak is nog steeds even intens. Aan de kant van de transportplanningsmodellen, opteren veel studiebureaus nog voor het klassieke vierstapsmodel, omdat het de meest intuïtieve en begrepen aanpak vormt. In tegenstelling tot dit model, wint activiteiten-gebaseerd modelleren aan aandacht, al blijft het voor veel mensen een min of meer obscure, niet transparante discipline. Aan de basis van deze twijfels tegenover de ABM aanpak ligt de afwezigheid van een algemeen aanvaarde omkadering zoals die wel aanwe-



Figuur D.1: Een voorbeeld van een microscopische verkeerssimulator, Mitrasim *2000* [Mae01b]. Links is een deel van de A10 ringweg rond Amsterdam te zien, rechts een gedetailleerder beeld van een op-/afrittencomplex. Elk tiende van een seconde worden de individuele voertuigen (personen- en vrachtwagens) voortbewogen doorheen het netwerk, waarbij ze telkens een bepaalde route volgen en proberen om hun bestemming botsingsvrij te bereiken.

zig is voor het 4SM. Het lijkt dan ook verleidelijk om de ABM aanpak rechtstreeks te vertalen naar die van de 4SM (waarbij bijvoorbeeld de creatie van de synthetische populatie overeenkomt met de productie- en attractiecycli, de distributie-stap en de opsplitsing naar mode), waarbij tijdsafhankelijke herkomst-bestemmingstabellen worden afgeleid. In gelijkaardige trend, kan men de simulatie van individuen in ABM zien als een implementatie van de verkeerstoedeling in het 4SM. Niettemin blijft het echter moeilijk om inzicht te krijgen in een dergelijke rechtstreekse vertaling en het resulterende reisgedrag van een bevolking. We merken bij dit laatste op dat het vakgebied rond ABM zich in een toestand van beweging bevindt, dankzij de steeds meer toenemende rekenkracht van computers.

Op het vlak van verkeersmodellering, blijft het debat omtrent welke nu de juiste aanpak is (micro-/meso/- of macroscopische modellen), steeds stof tot intrigerende discussies op te werpen. Ondanks de respectievelijke kritieken, is het algemeen aanvaard dat het modelleren van bestuurdersgedrag complexe mens-mens, mens-voertuig en voertuig-voertuig interacties vereist. Dit vraagt dan ook om interdisciplinair onderzoek, met invloeden van vakgebieden zoals wiskunde, fysica, en ingenieurstechnieken, alsmede sociologie en psychologie (zie bijvoorbeeld ook het overzicht van Helbing en Nagel [Hel04]).

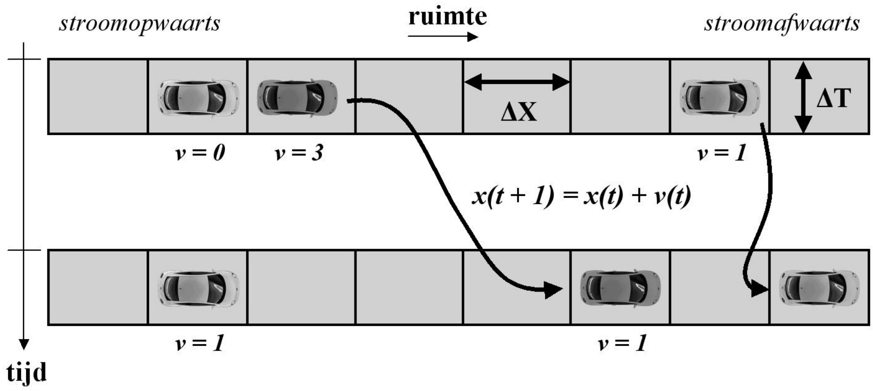
Hoofdstuk 4: Cellulaire automaten voor wegverkeer

Na de bespreking in het vorige hoofdstuk, leggen we in dit hoofdstuk de nadruk op computationeel efficiënte microscopische stroommodellen. Cellulaire automaten (*traffic cellular automata*, afgekort als TCA) passen netjes in deze beschrijving. Overeenkomstig met de principes van de statistische mechanica, hebben deze TCA modellen niet de bedoeling om een realistische microscopische beschrijving van de verkeersstroom als grondslag te zijn. Ze zijn er eerder op gericht om het macroscopisch gedrag correct weer te geven, en als dusdanig zeer geschikt om eerste- en tweede-orde macroscopische effecten van verkeersstromen te vatten.

Er bestaan reeds enkele overzichten van TCA modellen (bijvoorbeeld de meer theoretisch-georiënteerde werken van Chowdhury et al. [Cho00], Santen [San99], Knospe et al. [Kno04], en Mahnke et al. [Mah05]). Echter, geen enkele van deze overzichten beschouwt de TCA modellen exclusief vanuit hun gedragingsdynamiek. Ons onderzoek vult deze leegte in het huidige vakgebied in en geeft een antwoord op de vraag van onderzoekers naar een dergelijk bevattend inzicht. Merk op dat dit hoofdstuk in zijn geheel ook als opzichzelf-staand overzicht werd gepubliceerd in *Physics Reports* [Mae05].

In het hoofdstuk bespreken we eerst de historische achtergrond van cellulaire automaten (CA). We gaan dieper in op de ingrediënten van een CA en de wiskundige achtergrond, waarbij we een weg discretiseren in een aantal kleine cellen die elk een breedte hebben van bijvoorbeeld $\Delta X = 7.5$ m. Ook de tijd wordt gediscretiseerd in eenheden van ongeveer $\Delta T = 1$ s. Gedurende een tijdsstap van t naar $t + 1$, wordt op alle voertuigen tegelijk een verzameling van gedragsregels toegepast. Deze regels beschrijven hoe een voertuig zijn snelheid aanpast, hierbij rekening houdend met enerzijds dat het zo snel mogelijk wil rijden, en anderzijds dat het niet wil botsen met zijn voorligger. Hierna wordt de nieuwe positie berekend uitgaande van deze snelheid (zie Figuur D.2 voor een voorbeeld). We geven ook methodes aan om op een rooster van cellen macroscopische karakteristieken te meten. Verder beschrijven we ook hoe deze omgezet kunnen worden naar eenheden in de werkelijke wereld en vice versa.

Vervolgens behandelen we uitgebreid de gedragsmatige aspecten van enkele TCA modellen die in de literatuur voorkomen. We maken hierbij gebruik van tijd-ruimte diagrammen, empirische verbanden tussen de dichtheid en de intensiteit, en histogrammen die de verdeling van de snelheden, tussenruimtes en tussentijden van de voertuigen weergeven. We maken in het hoofdstuk een onderscheid tussen enkelvoudige en meervoudige celmodellen (telkens slechts 1 rijstrook). In de eerste soort neemt een voertuig exact 1 cel in, terwijl in de tweede soort langere voertuigen ondersteund worden doordat deze meerdere opeenvolgende cellen kunnen innemen. We besluiten met een kort overzicht van TCA modellen voor meerdere rijstroken, en modellen die werken op twee-dimensionale roosters van cellen. Het laatste deel van dit hoofdstuk illustreert enkele van de meer gebruikte analytische benaderingen van enkelvoudige TCA modellen.



Figuur D.2: Een schematisch overzicht van de werking van een cellulaire automaat voor een verkeersstroom met slechts 1 rijstrook. We zien twee tijdsstappen (voertuigen rijden van links naar rechts), waarbij tussen de tijdstippen t en $t + 1$ voor alle voertuigen wordt gekeken naar hun huidige snelheid. Deze snelheid wordt dan op basis van een verzameling gedragsregels berekend (bijvoorbeeld zo snel mogelijk proberen te rijden zonder op de voorligger te botsen), waarna de voertuigen hun nieuwe posities innemen.

Gegeven de huidige stand van zaken op het vlak van TCA modellen, toont onze analyse aan dat het vakgebied gedurende het laatste decennium sterk geëvolueerd is. Dit is voornamelijk te danken aan de toename in de rekenkracht van de computers. Er worden complexere modellen ontwikkeld, waarvan het *brake-light TCA* model (zie ook paragraaf 4.4.2.2 voor meer details) het meestbelovend lijkt. Aansluitend bij het overzicht van het vorige hoofdstuk, merken we op dat er een evoluerende trend is om deze TCA modellen ook in te schakelen als fysieke modellen die het wegverkeer beschrijven in multi-agent systemen (bijvoorbeeld in de activiteiten-gebaseerde aanpak van transportplanningsmodellen). Men beschrijft hier dan het gedrag van een ganse bevolking voor grootschalige wegennetwerken (i.e., landelijk niveau).

Hoofdstuk 5: Overeenkomende dynamiek van de STCA en het LWR model

Dit hoofdstuk is gewijd aan het dichten van de kloof die er heerst tussen microscopische en macroscopische modellen, in het bijzonder tussen het stochastische TCA model van Nagel en Schreckenberg (STCA) [Nag92b] en het eerste-orde macroscopische vloeistofmodel van Lighthill, Whitham en Richards (LWR) [Lig55; Ric56]. We gebruiken hiervoor een alternatieve methode die impliciet de stochasticiteit van het STCA model in het LWR model betrekt. We veronderstellen dat een stationaire voorwaarde geldt voor de gedragsregels van de STCA, waarna we deze laatste omzetten in een verzameling lineaire ongelijkheden. Deze definiëren dan de vorm van het fundamentele diagram dat vervolgens als parameter van het LWR model kan gezien worden. We passen onze methode toe op een kleine theoretische gevalsstudie. Hieruit besluiten we dat het zeer belangrijk is om de capaciteiten in beide systemen goed te vatten, zeker naarmate de stochastische ruis toeneemt.

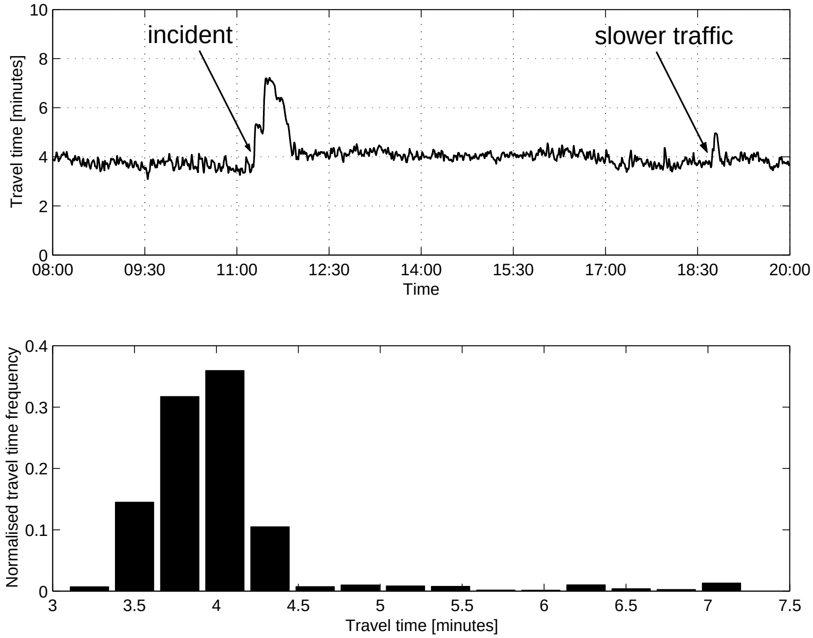
Hoofdstuk 6: Data kwaliteit, reistijdschattingen en betrouwbaarheid

Dit hoofdstuk legt zich toe op exploratieve data analyse (EDA); we beschouwen hierbij alle verkeersmetingen (tellingen) die vergaard worden op het Vlaamse autosnelwegennet. Eerst beschrijven we verschillende manieren om deze metingen te bekomen: enkelvoudige lusdetectoren in het wegdek, of camera's geplaatst langs de weg. Verder bekijken we hoe al deze gegevens in een databank worden bewaard en hoe we deze kunnen bevragen om bijvoorbeeld dagelijkse patronen te visualiseren. We bestuderen ook de gemiddelde snelheden die door de vermelde lusdetectoren berekend worden, en wat hier de effecten van zijn. Vervolgens werken we een methode uit om statistische uitschieters in de metingen op te sporen, en reiken we verschillende oplossingen aan om ontbrekende waarden op te vullen.

We ontwikkelen ook een visuele techniek die gebaseerd is op intuïtieve kaarten die een duidelijk overzicht geven van structurele en incidentele storingen van detectors. Daarnaast werken we een methode uit voor het off-line schatten van reistijden, gebaseerd op ruwe voertuigtellingen. Hiertoe construeren we eerst twee opeenvolgende cumulatieve curves van een gesloten wegsectie, waarna we een geautomatiseerde synchronisatie van deze curves uitvoeren, rekening houdend met systematische fouten. Daarna is het mogelijk om de verdeling van de reistijd te schatten met behulp van een histogram (zie bijvoorbeeld Figuur D.3). Het laatste deel van dit hoofdstuk bespreekt enkele tijd-ruimte kaarten die de verkeersdynamica weergeven (voor de E19 autosnelweg tussen Antwerpen en Brussel en de R0 ringweg rond Brussel); dit laat toe om punten met structurele filevorming te identificeren.

Hoofdstuk 7: Dynamische verkeerstoedeling gebaseerd op cellulaire automaten

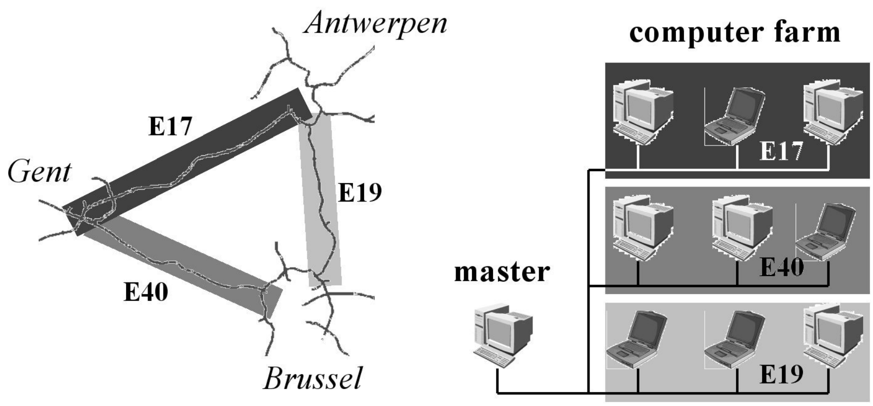
Binnen het kader van de modellering van de transportvraag, bestaan er drie grote methodologieën, namelijk trip-gebaseerd, activiteiten-gebaseerd en evenwichtsgebaseerd (zie ook paragrafen 3.1.2 en 3.1.3) [Boy98]. Deze diversiteit in het vakgebied is een duidelijk teken dat verschillende technieken worden beschouwd, elk gebaseerd op een eigen reeks aan ideeën. Ondanks deze verscheidenheid, gebruiken de verschillende technieken bepaalde aspecten van elkaar. Dit impliceert dat er enige algemene overeenkomsten tussen de modellen te vinden zijn. Bijgevolg, zal een voorspellingsmodel



Figuur D.3: *Bovenaan:* De evolutie van de reistijd gedurende een dag, berekend op basis van cumulatieve curves (zie ook paragraaf 2.3.2.2); de data was afkomstig van enkelvoudige lusdetectoren over drie rijstroken van de E40 autosnelweg tussen Erpe-Mere and Wetteren, dit voor maandag 4 april 2003. Zoals te zien, trad er hoogstwaarschijnlijk een incident op rond 11u00, waardoor de reistijd van vier naar zeven minuten toenam. Verder is te zien dat rond 18u45 ’s avonds al het verkeer gelijktijdig vertraagde gedurende een periode van ongeveer tien minuten. *Onderaan:* Gebaseerd op de berekende reistijden gedurende de dag, kunnen we een histogram opstellen dat een benadering geeft van de onderliggende verdeling van de reistijden. Op de grafiek is te zien dat het gemiddelde rond de vier minuten ligt.

voor transport een geven-en-nemen zijn tussen de verschillende vereisten en wensen en de huidige stand van zaken.

Kijkende naar de structuur die achter deze methodologieën zit, weten we dat een kerncomponent in elk van hen gevormd wordt door verkeerstoedeling [Boy04a]. In dit hoofdstuk beschrijven we eerst enkele benaderingen om dynamische verkeerstoedeling te doen, dit vanuit een zowel analytisch als simulatie-gebaseerd standpunt. Daarna stellen we een methode voor om dynamische verkeerstoedeling uit te voeren, waarbij we de keuze van het vertrektijdstip sequentieel integreren (dit leidt tot het fenomeen piekspreiding) met dynamische routekeuze. De methode is gebouwd rond een verkeersstroommodel dat uitgewerkt wordt als een computationeel efficiënte cellulaire automaat; we geven een functionele beschrijving en enkele implementatie details, waarna we een techniek bespreken die de efficiëntie nog groter maakt op basis van parallellisatie door middel van gedistribueerd rekenen. Bij dit laatste wordt de werklast verdeeld over een aantal gescheiden rekeneenheden (zie ook Figuur D.4 voor een illustratief overzicht). In het laatste deel van het hoofdstuk, geven we een kort overzicht van enkele mogelijke toepassingen, zoals schatting van de verkeerstoestand, duurzaamheidseffecten van verkeersbeheerssystemen en het schatten van de impact van verkeersmaatregelen.



Figuur D.4: Het idee achter gedistribueerd rekenen in een verkeersstroommodel: een computer (*master*) stuurt enkele rekeneenheden (*computer farm*) aan in een heterogene rekenomgeving (e.g., verschillende soorten computers). Al deze rekeneenheden werken samen, waarbij de rekenlast van het totale wegennet over hen verdeeld wordt. In het voorbeeld zien we drie grote autosnelwegen gemodelleerd, waarbij verschillende gegroepeerde rekeneenheden de verantwoordelijkheid over een autosnelweg krijgen toegewezen.

Hoofdstuk 8: Besluiten

In dit proefschrift werd gekeken naar de stand van zaken met betrekking tot de modellering van het verkeer op autosnelwegen, de numerieke data analyse van ruwe verkeersmetingen (tellingen) en de integratie van de keuze van vertrektijdstip en route in dynamische verkeerstoedeling. Met betrekking tot de literatuur onderscheiden onze bijdragen zich doordat ze een synthese vormen van de benaderingen voor het beschrijven van wegverkeer, terwijl dergelijke samenvattingen tot op heden enkel verspreid bestonden. Om een globaal beeld te krijgen met betrekking tot de kwaliteit van verkeersmetingen, bieden wij daarnaast methodes aan die kunnen omgaan met grootschalige data, dit in tegenstelling tot het meeste onderzoek naar de numerieke analyse van verkeersmetingen wat vaak slechts op beperkte data wordt uitgevoerd. Tenslotte met betrekking tot de vele benaderingen van het paradigma van simulatie-gebaseerde dynamische verkeerstoedeling, stellen wij een methodologie voor die de keuze van het vertrektijdstip sequentieel met de routekeuze integreert.

References

[Abr98] T. Abrahamsson. *Estimation of Origin-Destination Matrices Using Traffic Counts: A Literature Survey*. IR-98-021, International Institute for Applied Systems Analysis, May 1998.

[AD94] J. A. Acha-Daza and F. L. Hall. The application of catastrophe theory to traffic flow variables. *Transportation Research B,* volume 28:pages 235–250, 1994.

[Aer96] M. V. Aerde, B. Hellinga, M. Baker, and H. Rakha. INTEGRATION: An overview of traffic simulation features. In *Proceedings of the 75th Annual Meeting of the Transportation Research Board*. Washington, D.C., 1996.

[Ahm99] K. I. Ahmed. *Modeling Driver's Acceleration and Lane Changing Behavior*. Ph.D. thesis, Massachusetts Institute of Technology, Feb 1999.

[Ahn04] S. Ahn, M. J. Cassidy, and J. Laval. Verification of a simplified car-following theory. *Transportation Research B,* volume 38(5):pages 431–440, Jun 2004.

[AI04] L. Alvarez-Icaza, L. Muñoz, X. Sun, and R. Horowitz. Adaptive observer for traffic density estimation. In *Proceedings of the 2004 American Control Conference*, pages 2705–2710. Boston, Massachusetts, Jun 2004.

[Alg98] S. Algers, E. Bernauer, M. Boero, L. Breheret, and C. D. Taranto. *Simulation Modelling Applied to Road Transport European Scheme Tests (SMARTEST) – Review of Micro-Simulation Models*. SMARTEST Project Deliverable D3, Institute for Transport Studies, University of Leeds, Mar 1998. Transport RTD Programme of the 4th European Framework Programme.

[Alt03] E. Altman and L. Wynter. Equilibrium, games, and pricing in transportation and telecommunication networks. *Networks and Spatial Economics,* volume 2, 2003.

[Ang05] O. Angel, A. E. Holroyd, and J. B. Martin. *The Jammed Phase of the Biham-Middleton-Levine Traffic Model,* Mar 2005.

[Ans90] R. Ansorge. What does the entropy solution mean in traffic flow theory ? *Transportation Research B,* volume 24(2):pages 133–143, 1990.

[Are97] B. van Arem, A. de Vos, and M. Vanderschuren. *The microscopic traffic simulation model MIXIC 1.3*. INRO-VVG 1997-02b, TNO Inro, Department of Traffic and Transport, Delft, The Netherlands, Jan 1997.

[Are00] T. Arentze and H. Timmermans. *ALBATROSS: A Learning BAsed TRansportation Oriented Simulation System*. CIP-DATA Koninklijke Bibliotheek, Den Haag, The Netherlands, 2000. ISBN 90-6814-100-7.

[Arn94a] R. Arnott. Alleviating traffic congestion: Alternatives to road pricing. In *Proceedings of the Taxation, Resources and Economic Development (TRED) Conference on "Alternative Strategies for Managing Externalities"*. Lincoln Institute of Land Policy, Cambridge, Massachusetts, Sep 1994.

[Arn94b] R. Arnott and K. Small. The economics of traffic congestion. *American Scientist*, volume 82:pages 446–455, Sep 1994.

[Arn98] R. Arnott, A. de Palma, and R. Lindsey. Recent developments in the bottleneck model. In K. J. Button and E. T. Verhoef, editors, *Road pricing, traffic congestion and the environment*, pages 79–110. Edward Elgar Pub, Cheltenham, U.K., 1998.

[Arr05] S. Arroyo and A. L. Kornhauser. Modeling travel time distributions on a road network. In *Proceedings of the 2005 TRB Annual Conference*. Washington, D.C., Jan 2005.

[Ast02] V. Astarita. Node and link models for network traffic flow simulation. *Mathematical and Computer Modelling*, (35):pages 643–656, 2002.

[Aw00] A. Aw and M. Rascle. Resurrection of "second order" models of traffic flow ? *SIAM Journal on Applied Mathematics*, volume 60(3):pages 916–938, 2000.

[AWV04] *Het wegenpatrimonium in cijfers*. Administratie Wegen en Verkeer, Departement Leefmilieu en Infrastructuur (LIN) van het Ministerie van de Vlaamse Gemeenschap, Aug 2004.
(URL: http://wegen.vlaanderen.be/wegen/patrimonium).

[AWV06] *Knelpunten en 'missing links'*. Administratie Wegen en Verkeer, Departement Leefmilieu en Infrastructuur (LIN) van het Ministerie van de Vlaamse Gemeenschap, Feb 2006.
(URL: http://wegen.vlaanderen.be/wegen/missinglinks).

[Axh00] K. Axhausen. Activity-based modelling: Research directions and possibilities. In D. Simmonds and J. Bates, editors, *New Look at Multi-Modal Modelling*. Department of Environment, Transport and the Regions, London, Cambridge and Oxford, Oct 2000.

[BA85] M. Ben-Akiva and S. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge Massachusetts, 1985.

[BA96] M. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, and R. G. Mishalani. DynaMIT: Dynamic network assignment for the management of information to travelers. In *Proceedings of the 4th Meeting of the EURO Working Group on Transportation*. Newcastle, United Kingdom, 1996.

[BA98] M. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, and R. G. Mishalani. DynaMIT: a simulation-based system for traffic prediction. In *Proceedings of the DACCORD Short Term Forecasting Workshop*. Delft, The Netherlands, Feb 1998.

[BA99] M. Ben-Akiva and M. Bierlaire. Discrete choice methods and their applications to short term travel decisions. In R. Hall, editor, *Handbook of Transportation Science*. Kluwer Academic Publishers, Massachusetts, 1999.

[Bag05] P. Bagnerini, R. M. Colombo, and A. Corli. *On the role of source terms in continuum traffic flow models*, Mar 2005.

[Bak88] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. *Physical Review A*, volume 38:page 368, 1988.

[Bak02] M. Baker, R. Buyya, and D. Laforenza. Grids and grid technologies for wide-area distributed computing. In *Software – Practice and Experience*. 2002.

[Bal02] R. Balakrishna. *Calibration of the Demand Simulator in a Dynamic Traffic Assignment System*. Master's thesis, Massachusetts Institute of Technology, Jun 2002.

[Bal04a] M. Balmer, N. Cetin, K. Nagel, and B. Raney. Towards truly agent-based traffic and mobility simulations. In *Third International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS'04)*, volume 1, pages 60–67. New York City, New York, USA, Jul 2004.

[Bal04b] M. Balmer, K. Nagel, and B. Raney. Large-scale multi-agent simulations for transportation applications. *Journal of Intelligent Transportation Systems*, volume 8(4):pages 205–222, May 2004.

[Bal04c] M. Balmer, B. Raney, and K. Nagel. Coupling activity-based demand generation to a truly agent-based traffic simulation – activity time allocation. In *Proceedings of the EIRASS workshop on Progress in activity-based analysis*. Maastricht, The Netherlands, 2004.

[Bal05] M. Balmer, M. Rieser, A. Vogel, K. W. Axhausen, and K. Nagel. Generating day plans based on origin-destination matrices. In *Proceedings of the 5th Swiss Transport Research Conference (STRC)*. Monte Verità, Ascona, Mar 2005.

[Bal06] M. Balmer, K. W. Axhausen, and K. Nagel. A demand generation framework for large scale micro simulations. In *Proceedings of the 85th Annual Meeting of the Transportation Research Board*. Washington, D.C., Jan 2006.

[Ban95] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama. Dynamic model of traffic congestion and numerical simulation. *Physical Review E*, volume 51(2):pages 1035–1042, 1995.

[Bar95] C. L. Barrett, S. Eubank, K. Nagel, S. Rasmussen, J. Riordan, and M. Wolinsky. *Issues in the representation of traffic using multi-resolution cellular automata*. LA-UR-95:2658, Los Alamos National Laboratory, TRANSIMS Report Series, 1995.

[Bar98] R. Barlović, L. Santen, A. Schadschneider, and M. Schreckenberg. Metastable states in cellular automata for traffic flow. *European Physics Journal B*, volume B5(793), 1998.

[Bar99] R. Barlović, J. Esser, K. Froese, W. Knospe, L. Neubat, M. Schreckenberg, and J. Wahle. Online traffic simulation with cellular automata. *Traffic and Mobility: Simulation-Economics-Environment*, pages 117–134, Jul 1999. Institut für Kraftfahrwesen, RWTH Aachen, Duisburg.

[Bar02a] J. Barceló. Microscopic traffic simulation: A tool for the analysis and assessment of its systems. In *Proceedings of the Half Year Meeting of the Highway Capacity Committee*. Lake Tahoe, Jul 2002.

[Bar02b] J. Barceló and J. Casas. *Heuristic Dynamic Assignment based on Microscopic Traffic Simulation*. 9th Meeting of the Euro Working Group on Transportation, Jun 2002. Bari, Italy.

[Bar02c] R. Barlović, T. Huisinga, A. Schadschneider, and M. Schreckenberg. Open boundaries in a cellular automaton model for traffic flow with metastable states. *Physical Review E*, volume 66(4):pages 6113–6123, Oct 2002.

[Bar03] R. Barlović. *Traffic Jams – Cluster formation in low-dimensional cellular automata models for highway and city traffic*. Ph.D. thesis, Universität Duisburg-Essen, Standort Duisburg, Oct 2003.

[Bas04] F. Basile, C. Carbone, P. Chiacchio, R. Boel, and C. Avram. A hybrid model for urban traffic control. In *Proceedings of the 2004 IEEE SMC Conference*. Den Haag, The Netherlands, Oct 2004.

[Bat] R. Batley, T. Fowkes, D. Watling, G. Whelan, A. Daly, and E. Hato. *Models for analysing route choice*. Institute for Transport Studies, University of Leeds, UK.

[Bat01] R. P. Batley and R. G. Clegg. Driver route and departure time choices: The evidence and the models. In *Proceedings of the 33rd Annual UTSG conference*. St. Anne's College, Cambridge, 2001.

[Bec55] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation*. RM-1488. Yale University Press, May 1955.

[Bel91] M. G. Bell. The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research B*, volume 25(1):pages 13–22, 1991.

[Bel99] T. Bellemans, B. D. Schutter, and B. D. Moor. On data acquisition, modelling and simulation of highway traffic. In *Proceedings of the 9th IFAC Symposium on Control in Transportation Systems*, pages 22–27. Braunschweig, Germany, Jun 1999.

[Bel00] T. Bellemans, B. D. Schutter, and B. D. Moor. Data acquisition, interfacing and pre-processing of highway traffic data. In *Proceedings of the Telematics Automotive 2000*, volume 1, pages 1–7. Birmingham, U.K., Apr 2000.

[Bel03] T. Bellemans. *Traffic Control on Motorways*. Ph.D. thesis, Katholieke Universiteit Leuven, Department of Electrical Engineering ESAT-SCD (SISTA), May 2003.

[Ben96] S. C. Benjamin, N. F. Johnson, and P. Hui. Cellular automata models of traffic flow along a highway containing a junction. *Journal of Physics A: Mathematical and General*, volume 29:pages 3119–3127, Jun 1996.

[Ben04] I. Benenson and P. M. Torrens. *Geosimulation: Automata-based Modeling of Urban Phenomena*. John Wiley and Sons, Ltd., Jul 2004. ISBN 0470843497.

[Ber03] R. Bertini. Toward the systematic diagnosis of freeway bottleneck activation. In *Proceedings of the IEEE 6th Annual Conference on Intelligent Transportation Systems*. Shanghai, China, Oct 2003.

[Ber05] R. L. Bertini, R. V. Lindgren, D. Helbing, and M. Schonhof. Empirical observations of dynamic traffic flow phenomena on a German Autobahn. In *Proceedings of the 16th International Symposium on Transportation and Traffic Theory (ISTTT16)*. College Park, Maryland, Jul 2005.

[Bic04] P. Bickel, C. Chen, J. Kwon, J. Rice, E. van Zwet, and P. Varaiya. *Measuring Traffic*, May 2004.

[Bie95] M. Bierlaire. *Mathematical models for transportation demand analysis*. Ph.D. thesis, Facultés Universitaires Notre-Dame de la Paix de Namur, Dec 1995.

[Bih92] O. Biham, A. A. Middleton, and D. Levine. Self-organization and a dynamical transition in traffic-flow models. *Physical Review A*, volume 46(10):pages R6124–R6217, Nov 1992.

[Bli] M. C. Bliemer, R. J. Castenmiller, and P. H. Bovy. *Analytical multiclass dynamic traffic assignment using a dynamic network loading procedure*.

[Bli01] M. Bliemer. *Analytical dynamic traffic assignment with interacting user-classes: theoretical advances and applications using a variational inequality approach*. Ph.D. thesis, Faculty of Civil Engineering and Geosciences, TU Delft, Delft, The Netherlands, 2001.

[Bli04] M. Bliemer, E. Versteegt, and R. Castenmiller. Indy: A new analytical multiclass dynamic traffic assignment model. In *TRISTAN V: The Fifth Triennial Symposium on Transportation Analysis*. Le Gosier, Guadeloupe, Jun 2004.

[Boc00] N. Boccara and H. Fukś. Critical behavior of a cellular automaton highway traffic model. *Journal of Physics A: Mathematical and General*, volume 33:pages 3407–3415, 2000.

[Bol98] J. Bolker. *Writing Your Dissertation in Fifteen Minutes a Day: A Guide to Starting, Revising, and Finishing Your Doctoral Thesis*. Owl Books, 1st edition, Aug 1998. ISBN 080504891X.

[Bor01] B. D. Borger and S. Proost, editors. *Reforming Transport Pricing in the European Union, A modelling approach*. Edward Elgar Publishing, 2001. ISBN 1840641290.

[Bot00] J. A. Bottom. *Consistent Anticipatory Route Guidance*. Ph.D. thesis, Massachusetts Institute of Technology, Sep 2000.

[Bou02] E. Bourrel and V. Henn. Mixing micro and macro representations of traffic flow: A first theoretical step. In *Proceedings of the 9th meeting of the Euro Working Group on Transportation*, pages 610–616. Bari, Italia, Polytechnic of Bari, Jun 2002.

[Bov00] P. L. Bovy and R. Thijs, editors. *Estimators of Travel Time for Road Networks – New developments, evaluation results, and applications*. Delft University Press, 2000. ISBN 90-407-2048-7.

[Bow95] J. Bowman and M. Ben-Akiva. Activity-based model system of urban passenger travel demand. In *Proceedings of the 76th Annual Meeting of the Transportation Research Board*. Washington, D.C., 1995.

[Box76] G. E. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Cambridge University Press, Cambridge, 1976.

[Box94] G. E. Box, G. M. Jenkins, and G. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, New Jersey, 3rd edition, 1994.

[Boy98] D. Boyce. A practitioners guide to urban travel forecasting models. In *Proceedings of the Metropolitan Conference on Public Transportation Research*, pages 309–319. Chigaco, Illinois, Jun 1998.

[Boy01] D. Boyce, D.-H. Lee, and B. Ran. Analytical models of the dynamic traffic assignment problem. *Networks and Spatial Economics*, volume 1(3–4), 2001.

[Boy04a] D. E. Boyce. Forecasting travel on congested urban transportation networks: Review and prospects for network equilibrium models. In *Proceedings of TRISTAN V: The Fifth Triennial Symposium on Transportation Analysis*. Le Gosier, Guadeloupe, Jun 2004.

[Boy04b] D. E. Boyce, H. S. Mahmassani, and A. Nagurney. *A Retrospective on Beckmann, McGuire and Winsten's Studies in the Economics of Transportation*. To appear in revised form in Papers in Regional Science, Apr 2004.

[Boy06] D. Boyce and A. Nagurney. In memoriam: C. Bartlett McGuire (1925-2006) and Christopher B. Winsten (1923-2005). *Transportation Science*, volume 40(1):pages 1–2, Feb 2006.

[BPR64] Bureau of Public Roads, U.S. Department of Commerce, Urban Planning Division, Washington D.C. *Traffic Assignment Manual*, 1964.

[Bra69] D. Braeß. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, (12):pages 256–268, 1969.

[Bra76] D. Branston. Models of single lane time headway distributions. *Transportation Science*, volume 10:pages 125–148, 1976.

[Bra95] G. Braga, G. Cattaneo, P. Flocchini, and C. Q. Vogliotti. Pattern growth in elementary cellular automata. *Theoretical Computer Science*, volume 45:pages 1–26, 1995.

[Bra00] M. Brackstone and M. McDonald. Car following: A historical review. *Transportation Research F*, volume 2(4):pages 181–196, 2000.

[Bri99] W. Brilon and N. Wu. Evaluation of cellular automata for traffic flow simulation on freeway and urban streets. *Traffic and Mobility: Simulation-Economics-Environment*, pages 163–180, Jul 1999. Institut für Kraftfahrwesen, RWTH Aachen, Duisburg.

[Bri00] W. Brilon. Traffic flow analysis beyond traditional methods. In *Transportation Research Circular E-C018: 4th International Symposium on Highway Capacity*, pages 26–41. Transportation Research Board and National Research Council, Jun 2000.

[Bri03] P. A. Brinkman. *The Ethical Challenges and Professional Responses of Travel Demand Forecasters*. Ph.D. thesis, University of California, Berkeley, 2003.

[Bro03] E. Brockfeld, R. D. Kühne, A. Skabardonis, and P. Wagner. Towards a benchmarking of microscopic traffic flow models. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington, D.C., Jan 2003.

[Buc68] D. Buckley. A semi-Poisson model of traffic flow. *Transportation Science*, volume 2(2):pages 107–132, 1968.

[Buc74] D. Buckley and S. Yagar. Capacity funnels near on-ramps. In D. Buckley, editor, *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, pages 105–123. London, United Kingdom, 1974.

[Bud00] S. Budiansky. The physics of gridlock. *The Atlantic Online*, volume 286(6):pages 20–24, Dec 2000.

[Bui92] D. Bui, P. Nelson, and S. Narasimhan. *Computational realizations of the entropy condition in modelling congested traffic flow*. FHWA/TX-92/1232-7, Texas Transportation Institute, 1992. Federal Highway Administration report.

[Bur25] E. W. Burgess. The growth of the city. In R. E. Park, E. W. Burgess, and R. D. McKenzie, editors, *The City*, pages 47–62. University of Chicago Press, 1925.

[Bur48] J. M. Burgers. A mathematical model illustrating the theory of turbulence. *Advances in Applied Mechanics*, 1948. Academic Press, New York.

[Bur04a] W. Burghout. *A note on the number of replication runs in stochastic traffic simulation models*. Centre for Traffic Research (CTR), 2004.

[Bur04b] W. Burghout. *Hybrid microscopic-mesoscopic traffic simulation*. Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden, 2004.

[Bur05] W. Burghout, H. N. Koutsopoulos, and I. Andreasson. Hybrid mesoscopic-microscopic traffic simulation. In *Proceedings of the 84th Annual Meeting of The Transportation Research Board*. Jan 2005.

[Cal01] Caliper Corporation. *TransCAD*, 2001.

[Cam94] G. Cameron, B. J. N. Wylie, and D. McArthur. *PARAMICS – Moving Vehicles on the Connection Machine*. Technical report, Edinburgh Parallel Computing Centre (EPCC), The University of Edinburgh, 1994. IEEE, ISSN 1063-9535.

[Car05] K. Carlier, R. V. Nes, J. Schrijver, and S. Catalano. *TRANSFER: A new equilibrium model for analysing multimodal passenger trips*. 2005-01, Transport & Mobility Leuven, Aug 2005.

[Cas88] E. Cascetta and S. Nguyen. A unified framework for estimation or updating origin-destination matrices from traffic counts. *Transportation Research B*, volume 22(6):pages 437–455, 1988.

[Cas91] E. Cascetta and C. Cantarella. A day-to-day and within-day dynamic stochastic assignment model. *Transportation Research A*, volume 25(5):pages 277–291, 1991.

[Cas95] M. Cassidy and J. Windover. Methodology for assessing dynamics of freeway traffic flow. *Transportation Research Record*, volume 1484:pages 73–79, 1995.

[Cas97] M. J. Cassidy and B. Coifman. Relation among average speed, flow, and density and analogous relation between density and occupancy. *Transportation Research Record*, volume 1591:pages 1–6, 1997.

[Cas98] M. J. Cassidy. Bivariate relations in nearly stationary highway traffic. *Transportation Research B*, volume 32(1):pages 49–59, 1998.

[Cas99] M. Cassidy and R. Bertini. Some traffic features at freeway bottlenecks. *Transportation Research B*, volume 33:pages 25–42, 1999.

[Cas01] M. Cassidy and M. Mauch. An observed traffic pattern in long freeway queues. *Transportation Research A*, volume 35:pages 149–162, 2001.

[Cay97] R. Cayford, W.-H. Lin, and C. F. Daganzo. *The NETCELL simulation package: technical description*. UCB-ITS-PRR-97-23, California PATH Research Project, University of California, Berkeley, May 1997. ISSN 1055-1425.

[Cet03] N. Cetin, A. Burri, and K. Nagel. A large-scale agent-based traffic microsimulation based on queue model. In *Proceedings of the 3rd Swiss Transport Research Conference (STRC)*. Monte Verità, Ascona, Switserland, Mar 2003.

[Ceu01] M. D. Ceuster and L. Immers. *Verkeersindices voor het gebruik van het autosnelwegennet in België*. Technical report, Transport & Mobility Leuven, Oct 2001.

[Cha58] R. Chandler, R. Herman, and E. Montroll. Traffic dynamics: studies in car following. *Operations Research*, volume 6:pages 165–184, 1958.

[Cha03] S. Chanut and C. Buisson. Godunov discretization of a two-flow macroscopic model for mixed traffic with distinguished speeds and lengths. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington, D.C., Jan 2003.

[Che98] O. J. Chen. *Integration of Dynamic Traffic Control and Assignment*. Ph.D. thesis, Massachusetts Institute of Technology, Jun 1998. Promotor: Moshe E. Ben-Akiva.

[Che01a] A. Chen and W. W. Recker. Considering risk taking behavior in travel time reliability. In *Proceedings of the 80th Annual Meeting of The Transportation Research Board*. 2001. UCI-ITS-WP-00-24.

[Che01b] C. Chen, Z. Jia, and P. Varaiya. Causes and cures of highway congestion. *IEEE Control Systems Magazine*, pages 26–32, Dec 2001.

[Che02] C. Chen, E. van Zwet, P. Varaiya, and A. Skabardonis. *Travel Time Reliability as a Measure of Service*, 2002.

[Che03] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya. Detecting errors and imputing missing data for single loop surveillance systems. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington, D.C., 2003.

[Cho95] B. Chopard, P.-A. Queloz, and P. O. Luthi. Traffic models of a 2D road network. In *Proceedings of the 3rd CM Users' Meeting*. Parma, Oct 1995.

[Cho97] B. Chopard, A. Dupuis, and P. Luthi. A cellular automata model for urban traffic and its application to the city of Geneva. In M. Schreckenberg and D. Wolf, editors, *Proceedings of the Workshop on Traffic and Granular Flow '97*, pages 153–168. Springer, Duisburg, 1997.

[Cho98] D. Chowdhury, A. Pasupathy, and S. Sinha. Distributions of time- and distance-headways in the Nagel-Schreckenberg model of vehicular traffic: Effects of hindrances. *European Physics Journal B - Condensed Matter*, volume 5(3):pages 781–786, Oct 1998.

[Cho99a] D. Chowdhury, L. Santen, and A. Schadschneider. Vehicular traffic: A system of interacting particles driven far from equilibrium. *Current Science*, volume 77(411), 1999.

[Cho99b] D. Chowdhury and A. Schadschneider. Self-organization of traffic jams in cities: Effects of stochastic dynamics and signal periods. *Physical Review E*, volume 59(2):pages 1311–1314, Feb 1999.

[Cho00] D. Chowdhury, L. Santen, and A. Schadschneider. Statistical physics of vehicular traffic and some related systems. *Physics Reports*, volume 329:pages 199–329, 2000.

[Cho04] D. Chowdhury, K.Nishinari, and A. Schadschneider. Self-organized patterns and traffic flow in colonies of organisms. *Phase Transitions*, volume 77:pages 601–624, 2004.

[Cho05] D. Chowdhury, A. Schadschneider, and K. Nishinari. *Physics of Transport and Traffic Phenomena in Biology: from molecular motors and cells to organisms*, Sep 2005.

[Chr00] R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg. Three categories of traffic data: Historical, current and predictive. In E. Schnieder and U. Becker, editors, *9th IFAC Symposium Control in Transportation Systems*, pages 250–255. IFAC, Braunschweig, 2000, 2000.

[Chr04] R. Chrobok, S. F. Hafstein, and A. Pottmeier. Olsim: A new generation of traffic information systems. In V. Macho and K. Kremer, editors, *Forschung und wissenschaftliches Rechnen*, 63, pages 11–25. 2004.

[Chu02] K. Chung and M. J. Cassidy. *Testing Daganzo's Behavioral Theory for Multi-lane Freeway Traffic*. UCB-ITS-PWP-2002-9, University of California, Berkeley, Nov 2002.

[Chu04a] L. Chu, H. X. Liu, J.-S. Oh, and W. Recker. A calibration procedure for microscopic traffic simulation. In *Proceedings of the 83rd Transportation Research Board Annual Meeting*. Washington, D.C., USA, 2004.

[Chu04b] L. Chua, V. Sbitnev, and S. Yoon. A nonlinear dynamics perspective of Wolfram's New Kind of Science. part III: Predicting the unpredictable. *International Journal of Bifurcation and Chaos*, volume 14(11):pages 3689–3820, 2004.

[Cir02] C. Cirillo and K. W. Axhausen. Comparing urban activity travel behaviour. In *Proceedings of the 81st Annual Meeting of the Transportation Research Board*. Washington (DC), Jan 2002.

[Cle64] D. Cleveland and D. Capelle. Queueing theory approaches. In D. Gerlough and D. Capelle, editors, *An Introduction to Traffic Flow Theory*, pages 49–96. Highway Research Board, Washington, D.C., 1964. Special Report 79.

[Coi00] B. Coifman and M. Cassidy. *Automated Travel Time Measurement Using Vehicle Lengths From Loop Detectors*. UCB-ITS-PRR-2000-12, Institute of Transportation Studies, University of California, Berkeley, Aug 2000.

[Coi01] B. Coifman. Improved velocity estimation using single loop detectors. *Transportation Research A*, volume 35(10):pages 863–880, 2001.

[Coi02a] B. Coifman. Estimating travel times and vehicle trajectories on freeways using dual loop detectors. *Transportation Research Part A*, volume 36(4):pages 351–364, 2002.

[Coi02b] B. Coifman and M. Cassidy. Vehicle reidentification and travel time measurement on congested freeways. *Transportation Research Part A*, volume 36(10):pages 899–917, 2002.

[Coi03a] B. Coifman. Estimating density and lane inflow on a freeway segment. *Transportation Research A*, volume 37(8):pages 689–701, 2003.

[Coi03b] B. Coifman. Estimating median velocity instead of mean velocity at single loop detectors. *Transportation Research C*, volume 11(3–4):pages 211–222, Jun 2003.

[Cor05] J. Corbett. *Torsten Hägerstrand: Time Geography*. Regents of University of California, Santa Barbara, 2005.

[Cra04] M. S. Cramer. *Foundations of Fluid Mechanics*. Cambridge University Press, 2004. (URL: `http://www.navier-stokes.net`).

[Cre81] M. Cremer and M. Papageorgiou. Parameter identification for a traffic flow model. *Automatica*, volume 17:pages 837–843, 1981.

[Cre86] M. Cremer and J. Ludwig. A fast simulation model for traffic flow on the basis of boolean operations. *Mathematics and Computers in Simulation*, volume 28(4):pages 297–303, Aug 1986.

[Cro01] C. Cronström and M. Noga. Third-order phase transition and superconductivity in thin films. *Czechoslovak Journal of Physics*, volume 51(2):pages 175–184, Feb 2001.

[Cru87] J. P. Crutchfield and K. Kaneko. Phenomenology of spatiotemporal chaos. In B. Hao, editor, *Directions in Chaos*, pages 272–353. World Scientific, 1987.

[Cue93] J. A. Cuesta, F. C. Martínez, J. M. Molera, and A. Sánchez. Phase transitions in two-dimensional traffic flow models. *Physical Review E*, volume 48(6):pages 4175–4178, 1993.

[Čul88] K. Čulik and S. Yu. Undecidability of CA classification schemes. *Complex Systems*, volume 2(2):pages 177–190, 1988.

[Dag77] C. Daganzo and Y. Sheffi. On stochastic models of traffic assignment. *Transportation Science*, volume 11(3):pages 253–274, 1977.

[Dag94] C. F. Daganzo. The cell transmission model. part I: A simple dynamic representation of highway traffic. *Transportation Research B*, volume 28:pages 269–287, 1994.

[Dag95a] C. F. Daganzo. The cell transmission model. part II: Network traffic. *Transportation Research B*, volume 29(2):pages 79–93, 1995.

[Dag95b] C. F. Daganzo. A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Research B*, volume 29(4):pages 261–276, 1995.

[Dag95c] C. F. Daganzo. Properties of link travel time functions under dynamic loads. *Transportation Research B*, volume 29(2):pages 95–98, 1995.

[Dag95d] C. F. Daganzo. Requiem for second-order fluid approximations of traffic flow. *Transportation Research B*, volume 29(4):pages 277–286, 1995.

[Dag96] C. F. Daganzo. The nature of freeway gridlock and how to prevent it. In J.-B. Lesort, editor, *Proceedings of the 13th Symposium on Transportation and Traffic Theory*, pages 629–646. Lyon, France, Jul 1996.

[Dag97a] C. F. Daganzo. A continuum theory of traffic dynamics for freeways with special lanes. *Transportation Research B*, volume 31(2):pages 83–102, Apr 1997.

[Dag97b] C. F. Daganzo. *Fundamentals of Transportation and Traffic Operations*. Elsevier Science Ltd., 2 edition, 1997. ISBN 0-08-042785-5.

[Dag97c] C. F. Daganzo. A simple physical principle for the simulation of freeways with special lanes and priority vehicles. *Transportation Research B*, volume 31(2):pages 103–125, Apr 1997.

[Dag99a] C. F. Daganzo. The lagged cell-transmission model. In *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT)*. Jerusalem, Israel, 1999.

[Dag99b] C. F. Daganzo. Remarks on traffic flow modelling and its applications. *Traffic and Mobility: Simulation-Economics-Environment*, pages 105–115, Jul 1999. Institut für Kraftfahrwesen, RWTH Aachen, Duisburg.

[Dag99c] C. F. Daganzo, M. J. Cassidy, and R. L. Bertini. Possible explanations of phase transitions in highway traffic. *Transportation Research A*, volume 33:pages 365–379, 1999.

[Dag02a] C. F. Daganzo. A behavioral theory of multi-lane traffic flow. part I: Long homogeneous freeway sections. *Transportation Research B*, volume 36(2):pages 131–158, 2002.

[Dag02b] C. F. Daganzo. A behavioral theory of multi-lane traffic flow. part II: Merges and the onset of congestion. *Transportation Research B*, volume 36(2):pages 159–169, 2002.

[Dag02c] C. F. Daganzo. Reproducible features of congested highway traffic. *Mathematical and Computational Modeling*, (35):pages 509–515, 2002.

[Dag02d] C. F. Daganzo. Reversibility of the time-dependent shortest path problem. *Transportation Research B*, volume 36(7):pages 665–668, 2002.

[Dag03a] C. F. Daganzo. *A variational formulation of kinematic waves: Solution methods*. UCB-ITS-RR-2003-7, Institute of Transportation Studies and Department of Civil and Environmental Engineering, University of California, Berkeley, California, Nov 2003.

[Dag03b] C. F. Daganzo and J. A. Laval. *On the Numerical Treatment of Moving Bottlenecks*. UCB-ITS-PWP-2003-10, University of California, Berkeley, Jul 2003.

[Dag05] C. F. Daganzo. A variational formulation of kinematic waves: basic theory and complex boundary solutions. *Transportation Research B*, volume 39(2):pages 187–196, 2005.

[Dag06] C. F. Daganzo. In traffic flow, cellular automata = kinematic waves. *Transportation Research B*, volume 40(5):pages 351–436, Jun 2006.

[Dai97] D. Dailey. Travel time estimates using a series of single loop volume and occupancy measurements. In *Proceedings of the 76th Annual Meeting of the Transportation Research Board*. Washington, D.C., Jan 1997.

[Dai99] D. Dailey. A statistical algorithm for estimating speed from single loop volume and occupancy measurements. *Transportation Research B*, volume 33(5):pages 313–322, 1999.

[Dam02] A. van Dam. TraFlowPACK software documentation. In *A Moving Mesh Finite Volume Solver for Macroscopic Traffic Flow Models*, chapter 5, pages 57–71. TNO-INRO, May 2002.

[Dav66] K. Davidson. A flow travel time relationship for use in transportation planning. *Proceedings of the Australian Road Research Board*, volume Part I(3):pages 183–194, 1966.

[Del95a] J. Del Castillo and F. Benítez. On the functional form of the speed-density relationship – I: General theory. *Transportation Research B*, volume 29(5):pages 373–389, 1995.

[Del95b] J. Del Castillo and F. Benítez. On the functional form of the speed-density relationship – II: Empirical investigation. *Transportation Research B*, volume 29(5):pages 391–406, 1995.

[Del98] M. Delorme. An introduction to cellular automata. In M. Delorme and J. Mazoyer, editors, *Cellular Automata – a parallel model*. Kluwer Academic Publishers Group, Dec 1998. ISBN 0792354931.

[del05] J. del Río and M. Lárraga. *Transient situations in traffic flow: Modelling the Mexico City Cuernavaca Highway*, Jan 2005.

[Dem77] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, volume 39:pages 1–8, 1977.

[Dem03] C. Demir, B. Kerner, R. Herrtwich, S. Klenov, H. Rehborn, M. Aleksic, T. Reigber, M. Schwab, and A. Haug. FCD for urban areas: method and analysis of practical realisations. In *Proceedings of the 10th World Congress and Exhibition on Intelligent Transport Systems and Services (ITSS03)*. Madrid, Spain, Nov 2003.

[Der92] B. Derrida, E. Domany, and D. Mukamel. An exact solution of a one-dimensional asymmetric exclusion model with open boundaries. *Journal of Statistical Physics*, volume 69:pages 667–687, 1992.

[Des01] F. Desmyter. *Ontwerp Mobiliteitsplan Vlaanderen – Naar een duurzame mobiliteit in Vlaanderen*. Ministerie van de Vlaamse Gemeenschap, Departement Leefmilieu en Infrastructuur, Mobiliteitscel, Jun 2001.

[Det01] D. Dettloff. *Mapping a Prism*. Course “Cartography 2”, University of Colorado, Boulder, Jul 2001.

[Die00] G. Diedrich, L. Santen, A. Schadschneider, and J. Zittartz. Effects of on- and off-ramps in cellular automata models for traffic flow. *International Journal of Modern Physics C*, volume 11(2):pages 335–345, 2000.

[Dij59] E. W. Dijkstra. A note on two problems in connection with graphs. *Numerische Mathematik*, volume 1:pages 269–271, 1959.

[Dow02] R. Dowling, J. Holland, and A. Huang. *Guidelines for Applying Traffic Microsimulation Modeling Software*. California Department of Transportation, Sep 2002.

[Dra67] J. Drake, J. Schofer, and A. May. A statistical analysis of speed density hypotheses. In L. Edie, R. Herman, and R. Rothery, editors, *Vehicular Traffic Science*, pages 112–117. American Elsevier Publishing Company, Inc., New York, 1967. Proceedings of the Third International Symposium on the Theory of Traffic Flow.

[D’S05] R. D’Souza. Geometric structure of coexisting phases found in the Biham-Middleton-Levine traffic model. *Physical Review E*, 2005.

[Dub01] J.-C. Dubacq, B. Durand, and E. Formenti. Kolmogorov complexity and cellular automata classification. *Theoretical Computer Science*, volume 259(1–2):pages 271–285, 2001.

[Dup98] A. Dupuis and B. Chopard. Parallel simulation of traffic in Geneva using cellular automata. *Parallel and Distributed Computing Practices Journal*, volume 1(3):pages 79–92, Sep 1998.

[DYN03] Maryland Transportation Initiative, University of Maryland. *DYNASMART-X 0.930 User’s Guide*, Dec 2003.

[Eck85] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, volume 57(3):pages 617–656, 1985.

[Eck87] J.-P. Eckmann, S. Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, volume 4:pages 973–977, 1987.

[Edi61] L. Edie. Car following and steady-state theory for non-congested traffic. *Operations Research*, volume 9:pages 66–76, 1961.

[Edi65] L. Edie. Discussion on traffic stream measurements and definitions. In J. Almond, editor, *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, pages 139–154. OECD, Paris, France, 1965.

[Ehm00] D. Ehmanns and A. Hochstädter. Driver-model of lane change maneuvers. In *7th World Congress on Intelligent Transport Systems*. Turijn, Nov 2000.

[Eis98] B. Eisenblätter, L. Santen, A. Schadschneider, and M. Schreckenberg. Jamming transition in a cellular automaton model for traffic flow. *Physical Review E*, volume 57:pages 1309–1314, 1998.

[Eis03] N. Eissfeldt and P. Wagner. Effects of anticipatory driving in a traffic flow model. *European Physical Journal B*, volume 23:pages 121–129, 2003.

[Els81] P. Else. A reformulation of the theory of optimal congestion taxes. *Journal of Transport Economics and Policy*, volume 15:pages 217–232, 1981.

[Els82] P. Else. A reformulation of the theory of optimal congestion taxes: a rejoinder. *Journal of Transport Economics and Policy*, volume 16:pages 299–304, 1982.

[Emm97] H. Emmerich and E. Rank. An improved cellular automaton model for traffic flow simulation. *Physica A*, volume 234:pages 676–686, 1997.

[Emm98] H. Emmerich, T. Nagatani, and K. Nakanishi. From modified Korteweg-de Vries equation to a second-order cellular automaton for traffic flow. *Physica A*, volume 254:pages 548–556, 1998.

[Ess97] J. Esser and M. Schreckenberg. Microscopic simulation of urban traffic based on cellular automata. *International Journal of Modern Physics C*, volume 8(5):pages 1025–1036, Jun 1997.

[Ett03] D. Ettema, G. Tamminga, and H. Timmermans. A micro-simulation model system of departure time and route choice under travel time uncertainty. In *Proceedings of the 10th International Conference on Travel Behaviour Research*. Lucerne, Aug 2003.

[Eva92] A. Evans. Road congestion pricing: when is it a good policy ? *Journal of Transport Economics and Policy*, volume 26:pages 213–234, 1992.

[Eva93] A. Evans. Road congestion pricing: when is it a good policy ? A rejoinder. *Journal of Transport Economics and Policy*, volume 27:pages 99–105, 1993.

[Fas] U. Fastenrath. *Floating Car Data on a Larger Scale*. DDG Gesellschaft für Verkehrsdaten mbH.

[Fat03] N. Fatès. Experimental study of elementary cellular automata dynamics using the density parameter. In M. Morvan and É. Rémila, editors, *Discrete Models for Complex Systems – DMCS03*, volume AB, pages 155–166. 2003. Discrete Mathematics & Theoretical Computer Science.

[Fon05] M. D. Fontaine. *Improving the effectiveness of wireless location technology-based traffic monitoring: an exploration of system design issues*. Ph.D. thesis, The Faculty of the School of Engineering and Applied Science, University of Virginia, Jan 2005.

[For58] T. Forbes, H. Zagorski, E. Holshouser, and W. Deterline. Measurement of driver reactions to tunnel conditions. In *Proceedings of the Highway Research Board*, volume 37, pages 345–357. 1958.

[Fra56] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, volume 3:pages 95–110, 1956.

[Fra86] A. Fraser and H. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, volume 33(2):pages 1134–1140, 1986.

[Fra99] A. Frangioni and G. Gallo. A bundle type dual-ascent approach to linear multi commodity min-cost flow problems. *Informs Journal on Computing*, volume 11:pages 370–393, 1999.

[Fre90] E. Fredkin. Digital mechanics: An informational process based on reversible universal CA. *Physica D*, volume 45:pages 254–270, 1990.

[Fre95] J. Freund and T. Pöschel. A statistical approach to vehicular traffic. *Physica A*, volume 219:pages 95–113, 1995.

[Fre99] E. Freeman, S. Hupfer, and K. Arnold. *JavaSpaces: Principles, Patterns and Practice*. Addison-Wesley, 3rd edition, Nov 1999. ISBN 0-201-30955-6.

[Fuk93] M. Fukui and Y. Ishibashi. Evolution of traffic jam in traffic flow model. *Journal of the Physical Society of Japan*, volume 62(11):pages 3841–3844, Nov 1993.

[Fuk96] M. Fukui and Y. Ishibashi. Traffic flow in 1D cellular automaton model including cars moving with high speed. *Journal of the Physical Society of Japan*, volume 65(6):pages 1868–1870, 1996.

[Fuk97] M. Fukui and Y. Ishibashi. Effect of delay in restarting of stopped cars in a one-dimensional traffic model. *Journal of the Physical Society of Japan*, volume 66(2):pages 385–387, Feb 1997.

[Fuk98] H. Fukś and N. Boccara. Generalized deterministic traffic rules. *International Journal of Modern Physics C*, volume 9:pages 1–12, 1998.

[Fuk99] H. Fukś. Exact results for deterministic cellular automata traffic models. *Physical Review E*, volume 60:pages 197–202, 1999.

[Fuk01] H. Fukś and N. Boccara. Convergence to equilibrium in a class of interacting particle systems evolving in discrete time. *Physical Review E*, volume 64(1):page 016117, Jul 2001.

[Fuk04] H. Fukś. Critical behaviour of number-conserving cellular automata with nonlinear fundamental diagrams. *Journal of Statistical Mechanics – Theory and Experiment*, volume 5:page P07005, Jul 2004.

[Gar70] M. Gardner. Mathematical games – The fantastic combinations of John Conway's new solitaire game "Life". *Scientific American*, pages 120–123, Oct 1970.

[Gar97] N. Gartner, H. Mahmassani, C. J. Messer, H. Lieu, R. Cunard, and A. K. Rathi. *Traffic Flow Theory: A State-of-the-Art Report*. Technical report, Transportation Research Board, Dec 1997.

[Gaw97] C. Gawron, S. Krauß, and P. Wagner. Dynamic user equilibria in traffic simulation models. In *Proceedings of the Workshop on Traffic and Granular Flow '97*, 307. Duisburg, Germany, 1997.

[Gaw98a] C. Gawron. An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. *International Journal of Modern Physics C*, volume 9(3):pages 393–407, Feb 1998.

[Gaw98b] C. Gawron. *Simulation-Based Traffic Assignment*. Ph.D. thesis, Universität zu Köln, Feb 1998.

[Gaz59] D. Gazis, R. Herman, and R. Potts. Car-following theory of steady-state traffic flow. *Operations Research*, volume 7:pages 499–506, 1959.

[Gaz61] D. Gazis, R. Herman, and R. Rothery. Nonlinear follow-the-leader models of traffic flow. *Operations Research*, volume 9:pages 545–567, 1961.

[Gaz92] D. Gazis and R. Herman. The moving and 'phantom' bottleneck. *Transportation Science*, volume 26:pages 223–229, 1992.

[Gaz02] D. C. Gazis. The origins of traffic theory. *Operations Research*, volume 50(1):pages 69–77, Jan 2002.

[Gei94] A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manchek, and V. Sunderam. *PVM: Parallel Virtual Machine – A Users' Guide and Tutorial for Networked Parallel Computing*. Massachusetts Institute of Technology Press, 1994.

[Gel96] D. Gelernter. *Linda Spaces*. Yale Linda Group, Yale University, 1996.

[Geo02] Y. Georget. *A Game of Life in C for X11*, Mar 2002.

[Ger64] D. Gerlough. Simulation of traffic flow. In D. Gerlough and D. Capelle, editors, *An Introduction to Traffic Flow Theory*, pages 97–118. Highway Research Board, Washington, D.C., 1964. Special Report 79.

[Gho98] K. Ghosh, A. Majumdar, and D. Chowdhury. Distribution of time-headways in a particle-hopping model of vehicular traffic. *Physical Review E*, volume 58(3):pages 4012–4015, Sep 1998.

[Gip81] P. Gipps. A behavioural car following model for computer simulation. *Transportation Research B*, volume 15(2):pages 105–111, 1981.

[Gir03] R. Girvan. Going with the flow. *Scientific Computing World*, Mar 2003.

[Gob01] S. Gobron and N. Chiba. Crack pattern simulation based on 3D surface cellular automaton. *The Visual Computer – Special Issue: Computer Graphics International 2000*, volume 17(5):pages 287–309, 2001.

[God59] S. K. Godunov. A difference scheme for numerical computation of discontinuous solutions of equations of fluid dynamics. *Matematicheskii Sbornik*, volume 47(89):pages 271–306, 1959.

[Gof04] K. Goffin. Elke reis zijn prijs. *Dossier mobiliteit – snelbericht*, volume 19:pages 11–13, Dec 2004.

[Goo06] Google. *Google Maps*, Feb 2006.
(URL: `http://maps.google.com`).

[Gos74] R. Gosper. Life is universal ! In E. Berlekamp, J. Conway, and R. Guy, editors, *Winning Ways for your Mathematical Plays*, volume 2 (1982). Proceedings of the Winter Simulation Conference, Washington D.C., 1974, Academic Press, New York, 1974. Chapter 25.

[Gra01] S. Grabolus. *Numerische Untersuchungen zum Nagel-Schreckenberg-Verkehrsmodell und dessen Varianten*. Master's thesis, Institut für Theoretische Physik, Universität zu Köln, 2001.

[Gra03] L. Gray. A mathematician looks at Wolfram's New Kind of Science. *Notices of the AMS*, volume 50(2):pages 200–211, Feb 2003.

[Gre35] B. D. Greenshields. A study of traffic capacity. In *Highway Research Board Proceedings*, volume 14, pages 448–477. 1935.

[Gre59] H. Greenberg. An analysis of traffic flow. *Operations Research*, volume 7:pages 78–85, 1959.

[Gro02] W. Grosso. *Java RMI*. O'Reilly, 2002. ISBN 1-56592-452-5.

[Gun04] P. Gundaliya, V. Tom, and S. Dhingra. Heterogeneous traffic flow modelling using cellular automata for an arterial. In *Proceedings of the 6th International Workshop on Transportation Planning and Implementation Methodologies for Developing Countries (TPMDC04)*. Transportation Systems Engineering, IIT Bombay, Dec 2004.

[Gut96] H. Gutowitz. Cellular automata and the sciences of complexity (part I). *Complexity*, volume 1(5), 1996.

[Häg70] T. Hägerstrand. What about people in regional science ? *Papers of the Regional Science Association*, volume 24(1):pages 7–21, 1970.

[Hag01] J. N. Hagstrom and R. A. Abrams. Characterizing Braeß's paradox for traffic networks. In *Proceedings of IEEE 2001 Conference on Intelligent Transportation Systems*, pages 837–842. 2001.

[Hai63] F. Haight. *Mathematical theories of traffic flow*. New York Academic Press, 1963.

[Hal86] F. Hall, B. Allen, and M. Gunter. Empirical analysis of freeway flow-density relationships. *Transportation Research A*, volume 20:pages 197–210, 1986.

[Hal89] F. Hall and B. Persaud. Evaluation of speed estimates made with single-detector data from freeway traffic management systems. *Transportation Research Record*, volume 1232:pages 9–16, 1989.

[Han06] S. Han and B. Heydecker. Consistent objectives and solution of dynamic user equilibrium models. *Transportation Research Part B*, volume 40(1):pages 16–40, Jan 2006.

[Har45] C. D. Harris and E. L. Ullman. The nature of cities. *Annals of the American Academy of Political and Social Science*, volume 242:pages 7–17, 1945.

[Har01a] W. Harrington, A. J. Krupnick, and A. Alberini. Overcoming public aversion to congestion pricing. *Transportation Research A*, volume 35(2):pages 87–105, Feb 2001.

[Hår01b] B. Hårsman. Urban road pricing acceptance. In *Proceedings of the IMPRINT-EUROPE seminar*. Brussels, Belgium, Nov 2001.

[Hau85] A. Haurie and P. Marcotte. On the relationship between Nash-Cournot and Wardrop equilibria. *Networks*, volume 15:pages 295–309, 1985.

[Hau05a] T. D. Hau. Economic fundamentals of road pricing: A diagrammatic analysis, part I – Fundamentals. *Transportmetrica*, volume 1(2):pages 81–117, 2005.

[Hau05b] T. D. Hau. Economic fundamentals of road pricing: A diagrammatic analysis, part II – Relaxation of assumptions. *Transportmetrica*, volume 1(2):pages 119–149, 2005.

[HCM00] *2000 Highway Capacity Manual*. Transportation Research Board, National Academy of Science, Washington, D.C., 2000. ISBN 9991332944.

[Heg99] R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos*, (9):pages 413–435, 1999.

[Heg01] A. Hegyi, B. D. Schutter, J. Hellendoorn, S. Hoogendoorn, and C. Tampère. Gelijke behandeling voor verkeersstroommodellen. *Verkeerskunde*, volume 52(4):pages 32–36, 2001.

[Heg04] A. Hegyi. *Model Predictive Control for Integrating Traffic Control Measures*. Ph.D. thesis, The Netherlands TRAIL Research School, Delft University of Technology, Feb 2004.

[Hei99] D. Heidemann. Some critical remarks on a class of traffic flow models. *Transportation Research B*, volume 33(2):pages 153–155, Apr 1999.

[Hei01] D. Heidemann. A queueing theory model of nonstationary traffic flow. *Transportation Science*, volume 35(4):pages 405–412, Nov 2001.

[Hel61] W. Helly. Simulation of bottlenecks in single lane traffic flow. In *Theory of Traffic Flow*, pages 207–238. Elsevier Publishing Co., 1961.

[Hel96] D. Helbing. Gas-kinetic derivation of Navier-Stokes-like traffic equations. *Physical Review E*, volume 53(3):pages 2266–2381, 1996.

[Hel97] D. Helbing. *Verkehrsdynamik*. Springer-Verlag, Berlin, Heidelberg, 1997. ISBN 3-540-61927-5.

[Hel98a] D. Helbing. From microscopic to macroscopic traffic models. In J. Parisi, S. Muller, and W. Zimmermann, editors, *A Perspective Look at Nonlinear Media. From Physics to Biology and Social Sciences*, pages 122–139. Springer, Berlin, 1998.

[Hel98b] D. Helbing and B. Huberman. Coherent moving states in highway traffic. *Nature*, volume 396(738):pages 738–740, Dec 1998.

[Hel98c] B. R. Hellinga. *Requirements for the Calibration of Traffic Simulation Models*. Canadian Society of Civil Engineering (CSCE), University of Waterloo, Ontario, Canada, 1998.

[Hel99a] D. Helbing, A. Hennecke, and M. Treiber. Phase diagram of traffic states in the presence of inhomogeneities. *Physical Review Letters*, volume 82:pages 4360–4363, 1999.

[Hel99b] D. Helbing and M. Schreckenberg. Cellular automata simulating experimental properties of traffic flow. *Physical Review E*, volume 59:pages 2505–2508, 1999.

[Hel99c] D. Helbing and Tamás Vicsek. Optimal self-organization. *New Journal of Physics*, volume 1(13):pages 1–17, 1999.

[Hel99d] D. Helbing and M. Treiber. Numerical simulation of macroscopic traffic equations. *Computing in Science and Engineering*, volume 1(5):pages 89–99, Sep 1999.

[Hel01a] D. Helbing. MASTER: macroscopic traffic simulation based on a gas-kinetic, non-local traffic model. *Transportation Research B*, volume 35(2):pages 183–211, Feb 2001.

[Hel01b] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, volume 73:pages 1067–1141, Oct 2001.

[Hel02a] D. Helbing, A. Hennecke, V. Shvetsov, and M. Treiber. Micro- and macrosimulation of freeway traffic. *Mathematical and Computer Modelling*, volume 35:pages 517–547, 2002.

[Hel02b] D. Helbing and M. Treiber. Critical discussion of synchronized flow. *Cooper@tive Tr@nsport@tion Dyn@mics*, volume 1:pages 2.1–2.24, 2002.

[Hel04] D. Helbing and K. Nagel. The physics of traffic and regional development. *Contemporary Physics*, volume 45(5):pages 405–426, Sep 2004.

[Hen81] C. Hendrickson and G. Kocur. Schedule delay and departure time decisions in a deterministic model. *Transportation Science*, volume 15(1):pages 62–77, 1981.

[Hen03] V. Henn. A wave-based resolution scheme for the hydrodynamic LWR traffic flow model. In P. Bovy, S. Hoogendoorn, M. Schreckenberg, and D. Wolf, editors, *Proceedings of the Workshop on Traffic and Granular Flow '03*, pages 105–124. Delft University of Technology, Springer, 2004, Delft, The Netherlands, Oct 2003.

[Her45] E. Herrey and H. Herrey. Principles of physics applied to traffic movements and road conditions. *American Journal of Physics*, volume 13:pages 1–14, 1945.

[Her59] R. Herman, E. Montroll, R. Potts, and R. Rothery. Traffic dynamics: analysis of stability in car-following. *Operations Research*, volume 7:pages 86–106, 1959.

[Her79] R. Herman and I. Prigogine. A two-fluid approach to town traffic. *Science*, volume 204:pages 148–151, 1979.

[Hil93] P. Hills. Road congestion pricing: when is it a good policy ? A comment. *Journal of Transport Economics and Policy*, volume 27:pages 91–99, 1993.

[Hol98] E. Holland. A generalised stability criterion for motorway traffic. *Transportation Research B*, volume 32(2):pages 141–154, Feb 1998.

[Hoo99] S. P. Hoogendoorn. *Multiclass continuum modelling of multilane traffic flow*. Ph.D. thesis, Technical University Delft, 1999. Delft University Press.

[Hoo00] S. P. Hoogendoorn and P. H. Bovy. Continuum modeling of multiclass traffic flow. *Transportation Research B*, volume 34(2):pages 123–146, 2000.

[Hoo01] S. P. Hoogendoorn and P. H. Bovy. State-of-the-art of vehicular traffic flow modelling. *Proceedings of the Institution of Mechanical Engineers*, volume 215(1):pages 283–303, 2001.

[Hoo02a] S. Hoogendoorn, P. Bovy, and H. van Lint. Short-term prediction of traffic flow conditions in a multilane multi-class network. In M. A. Taylor, editor, *Transportation and Traffic Theory in the 21st Century*, pages 625–651. Oxford: Pergamon, Elsevier Science, 2002.

[Hoo02b] S. Hoogendoorn, H. van Lint, and H. van Zuylen. Macroscopic network model for multiclass multilane traffic flow prediction. In M. A. Taylor, editor, *Transportation and traffic flow theory in the 21st century*, pages 625–651. Oxford: Pergamon, Elsevier Science, 2002.

[Hoo03] S. Hoogendoorn, H. V. Zuylen, M. Schreuder, B. Gorte, and G. Vosselman. Microscopic traffic data collection by remote sensing. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington, D.C., Jan 2003.

[Hor99] A. J. Horowitz and M. H. Patel. Through trip tables for small urban areas: A method for quick response travel forecasting. *Transportation Research Record*, volume 1685:pages 57–64, 1999.

[Hou03] J. Hourdakis, P. Michalopoulos, and J. Kottommannil. A practical procedure for calibrating microscopic traffic simulation models. *Transportation Research Record*, volume 1852:pages 130–139, 2003.

[Hoy39] H. Hoyt. *The Structure and Growth of Residential Neighborhoods in American Cities*. Landmark Publication Housing, 1939.

[Igl93] B. Iglewicz and D. Hoaglin. *How to Detect and Handle Outliers*. American Society for Quality Control, Milwaukee, Wisconsin, 1993.

[Imm98a] B. Immers, M. Westerman, and H. de Ruiter. Sturen zonder structuren. In *Proceedings Colloquium Vervoersplanologisch Speurwerk*. Amsterdam, The Netherlands, Nov 1998.

[Imm98b] L. Immers and J. Stada. *Verkeersmodellen*. Departement Burgerlijke Bouwkunde, Faculteit Toegepaste Wetenschappen, Katholieke Universiteit Leuven, Oct 1998. Cursus H111.

[Imm04] L. Immers, A. Bleukx, and M. Snelder. Betrouwbaarheid van reistijden – het belang van netwerkstructuur en optimale wegcapaciteit. *Tijdschrift Vervoerswetenschappen*, volume 40(4):pages 18–28, 2004.

[Iwa98] J. S. Iwanski and E. Bradley. Recurrence plots of experimental data: To embed or not to embed ? *Chaos*, volume 8:pages 861–871, 1998.

[Jac90] L. Jacobson, N. Nihan, and J. Bender. Detecting erroneous loop detector data in a freeway traffic management system. *Transportation Research Record*, volume 1287:pages 151–166, 1990.

[Jac98] R. Jacob, M. Marathe, and K. Nagel. *A Computational Study of Routing Algorithms for Realistic Transportation Networks*. LA-UR-98-2249, Los Alamos National Laboratory, 1998.

[Jay00] R. Jayakrishan, J.-S. Oh, and A.-E. K. Sahraoui. *Calibration and Path Dynamics Issues In Microscopic Simulation for ATMS and ATIS*. UCI-ITS-WP-00-22, Institute of Transportation Studies, University of California, Irvine, CA 92697-3600, U.S.A., Dec 2000.

[Jia01] Z. Jia, C. Chen, B. Coifman, and P. Varaiya. The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors. In *Proceedings of IEEE Intelligent Transportation Systems 2001*, pages 536–541. 2001.

[Jin05] W.-L. Jin. *A kinematic wave theory of lane-changing vehicular traffic*, Mar 2005.

[Jok02] D. Joksimovic and M. Bliemer. *Dynamic Assignment Models for Assessing Road Pricing Policies*. MD-PIT Meeting, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Amsterdam, The Netherlands, Sep 2002.

[Jon83] P. Jones, M. Dix, M. Clarke, and I. Heggie. *Understanding Travel Behavior*. Gower, Aldershot, UK, 1983.

[Jos02] D. Jost. *Breakdown and recovery in traffic flow models*. Master's thesis, Department of Computer Science, ETH Zürich, Aug 2002.

[Jos03a] D. Jost and K. Nagel. Probabilistic traffic flow breakdown in stochastic car following models. In *Transportation Research Board Annual Meeting*. Washington, D.C., 2003. Paper 03-4266.

[Jos03b] D. Jost and K. Nagel. Probabilistic traffic flow breakdown in stochastic car following models. In P. Bovy, S. Hoogendoorn, M. Schreckenberg, and D. Wolf, editors, *Proceedings of the Workshop on Traffic and Granular Flow '03*, pages 87–103. Delft University of Technology, Springer, 2004, Delft, The Netherlands, Oct 2003.

[Júl05] J. Júlvez and R. Boel. Modelling and controlling traffic behaviour with continuous Petri nets. In *Proceedings of the 16th IFAC World Congress*. Prague, Jul 2005.

[Jün02] A. Jüngel. *Modeling and Numerical Approximation of Traffic Flow Problems*. Universität Mainz, Dec 2002.

[Kan90] K. Kaneko. Simulating physics with coupled map lattices. In K. Kawasaki, A. Onuki, and M. Suzuki, editors, *Formation, Dynamics, and Statistics of Patterns*, pages 1–52. World Scientific, 1990.

[Kar97] I. Karafyllidis and A. Thanailakis. A model for predicting forest fire spreading using cellular automata. *Ecological Modelling*, volume 99:pages 87–97, 1997.

[Kat63] J. H. Katz. Simulation of a traffic network. *Source Communications of the ACM archive*, volume 6(8):pages 480–486, Aug 1963.

[Kat04] R. van Katwijk, P. van Koningsbruggen, B. D. Schutter, and J. Hellendoorn. A test bed for multi-agent control systems in road traffic management. In P. H. Bovy, editor, *Proceedings of the 8th TRAIL Congress 2004 – A World of Transport, Infrastructure and Logistics*, pages 309–329. The Netherlands Research School for Transport, Infrastructure and Logistics (TRAIL), Delft University Press, Delft, The Netherlands, Nov 2004.

[Kay01] C. Kayatz. *Stability Analysis of Traffic Flow Models*. Master's thesis, Eidgenössische Technische Hochschule in Zürich, Aug 2001.

[Kem] C. Kemper. Dynamic traffic assignment with static OD matrices. In *Proceedings of the 2nd International Symposium*.

[Kem05] C. Kemper. Dynamic traffic flow model – a new approach with static data. In *Proceedings of the 5th European Congress and Exhibition on Intelligent Transport Systems (ITS)*. Hannover, Germany, Jun 2005.

[Ken92] M. Kennel, R. Brown, and H. Abarbanel. Determining minimum embedding dimension using a geometrical construction. *Physical Review A*, volume 45:pages 3403–3411, 1992.

[Ker94] B. Kerner and P. Konhäuser. Structure and parameters of clusters in traffic flow. *Physical Review E*, volume 50(1):pages 54–83, Jul 1994.

[Ker96a] B. Kerner and H. Rehborn. Experimental features and characteristics of traffic jams. *Physical Review E*, volume 53(2):pages 1297–1300, Feb 1996.

[Ker96b] B. Kerner and H. Rehborn. Experimental properties of complexity in traffic flow. *Physical Review E*, volume 53(5):pages 4275–4278, May 1996.

[Ker97] B. Kerner and H. Rehborn. Experimental properties of phase transitions in traffic flow. *Physical Review Letters*, volume 79(20):pages 4030–4033, Nov 1997.

[Ker98] B. Kerner. Experimental features of self-organization in traffic flow. *Physical Review Letters*, volume 81(17):pages 3797–3800, Oct 1998.

[Ker01] B. Kerner, H. Rehborn, M. Aleksić, and A. Haug. Methods for tracing and forecasting of congested traffic patterns on highways. *Traffic Engineering and Control*, volume 42:pages 282–287, 2001.

[Ker02] B. S. Kerner, S. K. Klenov, and D. E. Wolf. Cellular automata approach to three-phase traffic theory. *Journal of Physics A: Mathematical and General*, volume 35:pages 9971–10013, Nov 2002.

[Ker03] B. Kerner and S. Klenov. Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks. *Physical Review E*, volume 68(3), Sep 2003.

[Ker04] B. S. Kerner. *The Physics of Traffic – Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Understanding Complex Systems. Springer, 2004. ISBN 3-540-20716-3.

[Kim02] Y. Kim. *Online Traffic Flow Model Applying Dynamic Flow-Density Relations*. Ph.D. thesis, Technische Universität München, Mar 2002.

[Kla96] A. Klar, R. D. Kühne, and R. Wegener. Mathematical models for vehicular traffic. *Surveys on Mathematics for Industry*, (6):pages 215–239, 1996.

[Kle72] L. Kleinrock. Nets, stochastic message flow and delay. *Communications*, 1972.

[Klu03] G. Klunder, H. Versteegt, and C. Tampère. *PredicTime: Model specifications and initial test results*. 037N-142-73196, TNO INRO, The Netherlands, Jul 2003.

[Kni24] F. H. Knight. Some fallacies in the interpretation of social cost. *Quarterly Journal of Economics*, volume 38:pages 582–606, 1924.

[Kno00] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg. Towards a realistic microscopic description of highway traffic. *Journal of Physics A: Mathematical and General*, volume 33:pages 477–485, 2000.

[Kno02a] W. Knospe. *Synchronized traffic – Microscopic modeling and empirical observations*. Ph.D. thesis, Universität Duisburg, Jun 2002.

[Kno02b] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg. Human behavior as origin of traffic phases. *Physical Review E*, volume 65, 2002.

[Kno02c] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg. Single-vehicle data of highway traffic: microscopic description of traffic phases. *Physical Review E*, volume 65(05):pages 6133–6147, May 2002.

[Kno04] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg. An empirical test for cellular automaton models of traffic flow. *Physical Review E*, volume 70(016115), Jul 2004.

[Koh41] J. G. Kohl. *Der Verkehr und die Ansiedelungen der Menschen in ihrer Abhängigkeit des Gestaltung der Erdoberfläche*. Dresden, Leipzig, Germany, 1841.

[Kol98] A. B. Kolomeisky, G. M. Schütz, E. B. Kolomeisky, and J. P. Straley. Phase diagram of one-dimensional driven lattice gases with open boundaries. *Journal of Physics A: Mathematical and General*, volume 31:pages 6911–6919, 1998.

[Kom58] E. Kometani and T. Sasaki. On the stability of traffic flow. *Journal of Operations Research, Japan*, volume 2:pages 11–26, 1958.

[Kom61] E. Kometani and T. Sasaki. Dynamic behavior of traffic with a nonlinear spacing. In R. Herman, editor, *Theory of Traffic Flow*, pages 105–119. Elsevier, Amsterdam, 1961.

[Kos83] M. Koshi, M. Iwasaki, and I. Okhura. Some findings and an overview on vehicular flow characteristics. In V. Hurdle, E. Hauer, and G. Steuart, editors, *Proceedings of the 8th International Symposium on Transportation and Traffic Flow Theory*, pages 403–426. University of Toronto Press, Toronto, Canada, 1983.

[Kra97a] S. Krauß. Towards a unified view of microscopic traffic flow theories. In M. Papageorgiou and A. Pouliezos, editors, *IFAC/IFIP/IFORS Symposium*, volume 2. IFAC - International Federation of Automatic Control, Chania, Greece, Jun 1997.

[Kra97b] S. Krauß, P. Wagner, and C. Gawron. Metastable states in a microscopic model of traffic flow. *Physical Review E*, volume 55(304):pages 5597–5602, 1997.

[Kra98] S. Krauß. *Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics*. Ph.D. thesis, Deutsches Zentrum für Luft- und Raumfahrt, Apr 1998.

[Kra99] S. Krauß, K. Nagel, and P. Wagner. The mechanism of flow breakdown in traffic flow models. In *Proceedings of the International Symposium on Traffic and Transportation Theory (ISTTT99)*. Jerusalem, 1999.

[Kra04] D. Krajzewicz. *Simulation of Urban MObility – SUMO*. Deutsches Zentrum für Luft- und Raumfahrt (DLR), Dec 2004.

[Küh84] R. Kühne. Macroscopic freeway model for dense traffic-stop-start waves and incident detection. In N. Vollumler and R. Hamerslag, editors, *Proceedings of the 9th International Symposium on Transportation and Traffic Theory (ISTTT9)*, pages 21–42. 1984.

[Kul02] A. A. Kulkarni. *Modeling Activity Pattern Generation and Execution*. Ph.D. thesis, University of California, Irvine, 2002.

[Kur03] D. Kurzyniec, T. Wrzosek, D. Drzewiecki, and V. Sunderam. Towards self-organizing distributed computing frameworks: The H2O approach. *Parallel Processing Letters*, volume 13(2):pages 273–290, 2003.

[Kwo95] E. Kwon and P. Machalopoulos. Macroscopic simulation of traffic flows in complex freeway segments on a personal computer. In *Proceedings of the 6th Vehicle Navigation and Information Systems Conference*, pages 342–345. Seattle, Washington, D.C., 1995.

[Kwo03] J. Kwon, P. Varaiya, and A. Skabardonis. Estimation of truck traffic volume from single loop detector using lane-to-lane speed correlation. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington, D.C., Jan 2003.

[Lag03a] A. Lago. *Spatial Models of Morning Commute Consistent with Realistic Traffic Behavior*. UCB-ITS-DS-2003-1, Institute of Transportation Studies, University of California at Berkeley, 2003. ISSN 0192 4109.

[Lag03b] A. Lago and C. F. Daganzo. *A Network Model of Departure Time Choice with Spillovers and Merging Effects. Part I: A Building Block*. UCB-ITS-RR-2003-5, Institute of Transportation Studies, University of California at Berkeley, Jul 2003.

[Lár04] M. Lárraga, J. del Río, and A. Schadschneider. New kind of phase separation in a CA traffic model with anticipation. *Journal of Physics A: Mathematical and General*, (37):pages 3769–3781, 2004.

[Lav04] J. A. Laval. *Hybrid Models of Traffic Flow: Impacts of Bounded Vehicle Accelerations*. Ph.D. thesis, Graduate Division of the University of California, Berkeley, 2004.

[Lav06] J. A. Laval and C. F. Daganzo. Lane-changing in traffic streams. *Transportation Research Part B*, volume 40(3):pages 251–264, Apr 2006.

[Leb93] J. Lebacque. Les modèles macroscopiques de trafic. *Annales de Pont*, volume 67:pages 28–45, 1993. 3rd trim.

[Leb96] J. Lebacque. The Godunov scheme and what it means for first order traffic flow models. In J. Lesort, editor, *Proceedings of the 13th International Symposium on Transportation and Traffic Theory (ISTTT)*. Pergamon, Oxford, Nov 1996.

[Leb99] J. Lebacque and J. Lesort. Macroscopic traffic flow models: a question of order. In *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT)*, pages 3–25. Pergamon, Jerusalem, 1999.

[Lee52] T. Lee and C. Yang. Statistical theory of equations of state and phase transitions: II. Lattice gas and Ising model. *Physical Review*, volume 87(3):pages 410–419, Aug 1952.

[Lee99] H. Lee, H.-W. Lee, and D. Kim. Dynamic states of a continuum traffic equation with on-ramp. *Physical Review E*, volume 59(5):pages 5101–5111, May 1999.

[Lee02] K. Lee, P. Hui, D. Mao, B.-H. Wang, and Q.-S. Wu. Fukui-Ishibashi traffic flow models with anticipation of movement of the car ahead. *Journal of the Physical Society of Japan*, volume 71(7):pages 1651–1654, Jul 2002.

[Leo89] D. R. Leonard, P. Power, and N. B. Taylor. *CONTRAM: structure of the model*. TRL Report RR 178, Transportation Research Laboratory, Crowthorn, 1989.

[Ler00] G. Lerner, A. Hochstaedter, R. Kates, C. Demir, J. Meier, and A. Poschinger. The interplay of multiple scales in traffic flow: Coupling of microscopic, mesoscopic, and mocroscopic simulation. In *Proceedings of the 7th World Congress on Intelligent Transport Systems*. Turin, Nov 2000.

[LeV92] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser Verlag, Basel, 1992.

[Lev93] D. Levinson and A. Kumar. Integrating feedback into the transportation planning model: Structure and application. In *Proceedings of the 72nd Annual Meeting of Transportation Research Board*. Washington, D.C., Jan 1993.

[LeV01] R. J. LeVeque. Some traffic flow models illustrating interesting hyperbolic behavior. In *Minisymposium on Traffic Flow*. SIAM Annual Meeting, Jul 2001.

[LeV03] R. J. LeVeque. *CLAWPACK Version 4.2 User's Guide*. University of Washington, Dec 2003.

[Li90] W. Li and N. Packard. The structure of the elementary cellular automata rule space. *Complex Systems*, volume 4:pages 281–297, 1990.

[Li02] B. Li and B. De Moor. Dynamic identification of origin-destination matrices in the presence of incomplete observations. *Transportation Research B*, volume 36:pages 37–57, 2002.

[Li03] T. Li. Global solutions of nonconcave hyperbolic laws with relaxation arising from traffic flow. *Journal of Differential Equations*, volume 190:pages 131–149, 2003.

[Lig55] M. Lighthill and G. Whitham. On kinematic waves: II. A theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society*, volume A229, pages 317–345. 1955.

[Lim00] Q. Limited. *Paramics-online v3 – System Overview*. Technical report, Quadstone Limited, 16 Chester Street, Edinburgh EH3 7RA, Scotland, Nov 2000. (URL: http://www.paramics-online.com).

[Lim05] Y. Lim and B. Heydecker. Dynamic departure time and stochastic user equilibrium assignment. *Transportation Research B*, volume 39:pages 97–118, 2005.

[Lin95] W.-H. Lin and D. Ahanotu. *Validating the Basic Cell Transmission Model on a Single Freeway Link*. PATH technical note 95-3, Institute of Transportation Studies, University of California, Berkeley, 1995.

[Lin00] R. Lindsey and E. Verhoef. Congestion modeling. In D. Hensher and K. Button, editors, *Handbook of Transportation Modelling*, chapter 21. Elsevier Science, Ltd., 2000.

[Lin01] R. Lindsey and E. Verhoef. Traffic congestion and congestion pricing. In D. Hensher and K. Button, editors, *Handbook of Transport Systems and Traffic Control*, pages 77–105. Elsevier/Pergamon, Amsterdam, 2001.

[Lin03] M. Linauer and D. Leihs. Generating floating car data using GSM-network. In *Proceedings of the 10th World Congress and Exhibition on Intelligent Transport Systems and Services (ITSS03)*. Madrid, Spain, Nov 2003.

[Lin04] J. van Lint. *Reliable travel time prediction for freeways*. Ph.D. thesis, Technische Universiteit Delft, May 2004.

[Lin05] R. Lindgren. *Analysis of Flow Features in Queued Traffic on a German Freeway*. Ph.D. thesis, Portland State University, 2005.

[Lit87] R. Little and D. Rubin. *Statistical Analysis with Missing Data*. J. Wiley & Sons, New York, 1987.

[Liu02] H. X. Liu, X. Ban, B. Ran, and P. Mirchandani. An analytical dynamic traffic assignment model with probabilistic travel times and perceptions. *Transportation Research Record*, 2002.

[Lo99] H. K. Lo. A dynamic traffic assignment formulation that encapsulates the cell-transmission model. In A. Cedar, editor, *Transportation and Traffic Traffic Theory*, pages 327–350. Elsevier Science, 1999.

[Lo02] H. K. Lo and W. Szeto. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research B*, volume 36:pages 421–443, 2002.

[Lo04] H. K. Lo and W. Szeto. Modeling advanced traveler information services: static versus dynamic paradigms. *Transportation Research B*, volume 38(6):pages 495–515, Jul 2004.

[Lo05] H. K. Lo and W. Szeto. Road pricing modeling for hyper-congestion. *Transportation Research A*, 2005.

[Log03a] S. Logghe. *Dynamic modeling of heterogeneous vehicular traffic*. Ph.D. thesis, Katholieke Universiteit Leuven, Jun 2003.

[Log03b] S. Logghe and B. Immers. Heterogeneous traffic flow modelling with the lwr model using passenger-car equivalents. In *Proceedings of the 10th World Congress and Exhibition on Intelligent Transport Systems and Services (CD-ROM)*. ERTICO, ITS Europe, Madrid, Spain, Nov 2003.

[Log04] S. Logghe and F. Vanhove. *Het Belgische verkeer in cijfers*. 2004-01, Transport & Mobility Leuven, Jan 2004.

[Log05a] S. Logghe and F. Vanhove. *De betrouwbaarheid van Belgische verkeersdetectoren*. 2nd draft, Transport & Mobility Leuven, Mar 2005.

[Log05b] LogicaCMG. *Get moving faster with Mobile Traffic Services*, 2005.

[Log05c] LogicaCMG. *MTS Quality Assessment Results*, 2005.

[Lud98] J. Ludmann. *Beeinflussung des Verkehrsablaufs auf Straßen – Analyse mit dem fahrzeugorientierten Verkehrssimulationsprogramm PELOPS*. Ph.D. thesis, Institut für Kraftfahrwesen Aachen, 1998.

[Luk72] J. Luke. Mathematical models for landform evolution. *Journal of Geophysical Research*, volume 77:pages 2460–2464, 1972.

[Lus03] E. Lusk. *MPI-2: Extensions to the Message-Passing Interface*. Message Passing Interface Forum, Nov 2003.

[Mae01a] S. Maerivoet. *Advanced Computer Graphics using OpenGL*. Universitaire Instelling Antwerpen, Department of Mathematics and Computer Science, 2001.

[Mae01b] S. Maerivoet. *Het gebruik van microscopische verkeerssimulatie bij een onderzoek naar de fileproblematiek op de Antwerpse Ring*. Master's thesis, Universitaire Instelling Antwerpen, Jun 2001. Promotor: prof. dr. Serge Demeyer.

[Mae03a] S. Maerivoet, S. Logghe, B. D. Moor, and B. Immers. A comparison of a cellular automaton and a macroscopic model. In P. Bovy, S. Hoogendoorn, M. Schreckenberg, and D. Wolf, editors, *Proceedings of the Workshop on Traffic and Granular Flow '03*, pages 199–204. Delft University of Technology, Springer, 2004, Delft, The Netherlands, Oct 2003.

[Mae03b] S. Maerivoet and B. D. Moor. Development of an improved traffic cellular automaton model for traffic flows on a highway road network. In *Proceedings of the 10th World Congress and Exhibition on Intelligent Transport Systems and Services (CD-ROM)*. ERTICO, ITS Europe, Madrid, Spain, Nov 2003.

[Mae04a] S. Maerivoet. Efficient microscopic simulation of large scale highway traffic flows. In *23rd Benelux Meeting on Systems and Control*. Helvoirt, The Netherlands, Mar 2004.

[Mae04b] S. Maerivoet. *Models in Aid of Traffic Management*. Seminar slides for 'Transportmodellen ter ondersteuning van het mobiliteits- en vervoersbeleid', May 2004. Federaal Wetenschapsbeleid, Belgium.

[Mae04c] S. Maerivoet. *Traffic: an Interplay between Models, Simulations, and Control Actions*. Slides presented at the Friday Seminar (Universiteit Antwerpen), Mar 2004.

[Mae04d] S. Maerivoet. *Traffic Cellular Automata*. Java software tested with JDK 1.3.1, 2004. (URL: http://smtca.dyns.cx).

[Mae04e] S. Maerivoet. Sustainability effects of traffic management systems. In *The Transport Science and Technology Congress, Athens, Greece, 2004*. Sep 2004. Slides presented at TransTec 2004.

[Mae04f] S. Maerivoet and B. De Moor. *Advancing density waves and phase transitions in a velocity dependent randomization traffic cellular automaton*. 03-111, Katholieke Universiteit Leuven, Oct 2004.

[Mae04g] S. Maerivoet and B. De Moor. Non-concave fundamental diagrams and phase transitions in a stochastic traffic cellular automaton. *The European Physical Journal B – Condensed Matter Physics*, volume 42(1):pages 131–140, Nov 2004.

[Mae05] S. Maerivoet and B. D. Moor. Cellular automata models of road traffic. *Physics Reports*, volume 419(1):pages 1–64, Nov 2005.

[Mae06] S. Maerivoet and B. D. Moor. *Sustainability Effects of Traffic Management Systems*. TR 06-033, Department of Electrical Engineering ESAT-SCD (SISTA), Katholieke Universiteit Leuven, Mar 2006. Final report for the DWTC PODO-II - CP/40 project.

[Mag00] L. Magne, S. Rabut, and J.-F. Gabard. Towards an hybrid macro-micro traffic flow simulation model. In *INFORMS Salt Lake City Spring 2000 Conference*. May 2000.

[Mah36] P. C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, volume 2(49), 1936.

[Mah04] B. P. Mahanti. *Aggregate Calibration of Microscopic Traffic Simulation Models*. Master's thesis, Massachusetts Institute of Technology, Sep 2004.

[Mah05] R. Mahnke, J. Kaupužs, and I. Lubashevsky. Probabilistic description of traffic flow. *Physics Reports*, volume 408:pages 1–130, Jan 2005.

[Mal03] K. Malone, M. Bliemer, C. Menard, and H. Versteegt. *Testing INDY for use as the Dynamic Traffic Assignment model in the Dutch National Model*. 03-7N-084-73431, TNO Inro, Delft, Jun 2003. ISBN 90-5986-015-2.

[Mal05] C. Mallikarjuna and K. Ramachandra Rao. Traffic flow modelling on highways using cellular automata: A review. In J. Bandyopadhyay and B. Maitra, editors, *Proceedings of International Conference on Structural and Transportation Engineering (START05)*, pages 912–919. Elite Publishing House, New Delhi, 2005.

[Man65] P. Mann. *Modelling Study of a Typical British City*, 1965.

[Man79] M. L. Manheim. *Fundamentals of Transportation Systems Analysis*. MIT Press, Cambridge, Massachusetts, 1979.

[Mar98] E. Marcucci. Road pricing: Old beliefs, present awareness and future research patterns. In *Proceedings of the International Symposium on Technological, Energy and Environmental Topics in Transport*. Palazzo delle Stelline, Corso Magenta, 63, Milano, Italy, Oct 1998.

[Mar02] J. E. Marca. *Activity-Based Travel Analysis in the Wireless Information Age*. Ph.D. thesis, University of California, Irvine, 2002.

[Mar03] F. Marchal. Integrating time of use in large scale traffic simulations. In *Proceedings of the 3rd Swiss Transport Research Conference (STRC)*. Monte Verità, Ascona, Mar 2003.

[Mat03] J. Matsukidaira and K. Nishinari. Euler-Lagrange correspondence of cellular automaton for traffic-flow models. *Physical Review Letters*, volume 90:page 088701, 2003.

[May90] A. D. May. *Traffic Flow Fundamentals*. Prentice Hall, Englewood Cliffs, New Jersey 07632, 1990. ISBN 0-13-926072-2.

[McN96] M. G. McNally. *An Activity-Based Microsimulation Model for Travel Demand Forecasting*. UCI-ITS-AS-WP-96-1, Institute of Transportation Studies, Center for Activity Systems Analysis, University of California, 1996.

[McN00a] M. G. McNally. *The Activity-Based Approach*. UCI-ITS-AS-WP-00-4, Center for Activity Systems Analysis, University of California, Dec 2000. In Hensher and Button (eds.), "Handbook of Transportation Modeling", Pergamon.

[McN00b] M. G. McNally. *The Four Step Model*. UCI-ITS-AS-WP-00-5, Center for Activity Systems Analysis, University of California, Dec 2000. In Hensher and Button (eds.), "Handbook of Transportation Modeling", Pergamon.

[Mes90] A. Messmer and M. Papageorgiou. METANET: A macroscopic simulation program for motorway networks. *Traffic Engineering and Control*, volume 31:pages 466–470, 1990.

[Mic01] E. Michler. Statistical analysis of floating car data systems. In *8th World Congress on Intelligent Transportation Systems*. Sydney, Sep 2001.

[Mik72] B. Mikhalkin, H. Payne, and L. Isaksen. Estimation of speed from presence detectors. *Highway Research Record*, volume 388:pages 73–83, 1972.

[Mil04] H. J. Miller. Activities in space and time. In P. Stopher, K. Button, K. Haynes, and D. Hensher, editors, *Handbook of Transport 5: Transport Geography and Spatial Systems*. Pergamon/Elsevier Science, 2004.

[Min99] M. M. Minderhoud. *Supported driving: impacts on motorway traffic flow*. TRAIL Thesis T99/4, Delft University of Technology, Delft, The Netherlands, 1999.

[Mit54] R. B. Mitchell and C. Rapkin. *Urban Traffic: A Function of Land Use*. Columbia University Press, New York, 1954.

[Mon64] E. Montroll and R. Potts. Car following and acceleration noise. In D. Gerlough and D. Capelle, editors, *An Introduction to Traffic Flow Theory*, pages 36–48. Highway Research Board, Washington, D.C., 1964. Special Report 79.

[Mor03] A. Moreira. Universality and decidability of number-conserving cellular automata. *Theoretical Computer Science*, volume 292:pages 711–721, 2003.

[Mos54] K. Moskowitz. Waiting for a gap in a traffic stream. *Proceedings of the Highway Research Board*, volume 33:pages 385–395, 1954.

[Mul02] A.-M. Mulligan and A. Nicholson. Uncertainty in traffic flow estimation using the moving-observer method. In *IPENZ Transportation Group Conference Papers*. 2002.

[Muñ00a] J. C. Muñoz and C. Daganzo. The bottleneck mechanism of a freeway diverge. *Transportation Research A*, volume 36(6):pages 483–505, 2000.

[Muñ00b] J. C. Muñoz and C. Daganzo. *Experimental Characterization of Multi-lane Freeway Traffic Upstream of an Off-ramp Bottleneck*. UCB-ITS-PWP-2000-13, Partners for Advanced Transit and Highways (PATH), Jan 2000.

[Muñ02a] J. C. Muñoz and C. Daganzo. Moving bottlenecks: A theory grounded on experimental observation. In M. Taylor, editor, *Proceedings of the 15th International Symposium on Transportation and Traffic Theory (ISTTT15)*, pages 441–462. Pergamon-Elsevier, 2002.

[Muñ02b] J. C. Muñoz and C. F. Daganzo. Fingerprinting traffic from static freeway sensors. *Cooper@tive Tr@nsport@tion Dyn@mics*, volume 1(1):pages 1–11, May 2002.

[Muñ03a] J. C. Muñoz and C. F. Daganzo. Structure of the transition zone behind freeway queues. *Transportation Science*, volume 37(3):pages 312–329, Aug 2003.

[Muñ03b] L. Muñoz, X. Sun, R. Horrowitz, and L. Alvarez. Traffic density estimation with the cell transmission model. In *Proceedings of the American Control Conference*, pages 3750–3755. Denver, Colorado, Jun 2003.

[Myu03] I. J. Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, volume 47:pages 90–100, 2003.

[Nag92a] K. Nagel and E. Raschke. Self-organizing criticality in cloud formation ? *Physica A: Statistical and Theoretical Physics*, volume 182(4):pages 519–531, Apr 1992.

[Nag92b] K. Nagel and M. Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I France*, volume 2(12):pages 2221–2229, Dec 1992.

[Nag93a] T. Nagatani. Self-organization and phase transition in traffic-flow model of a two-lane roadway. *Journal of Physics A: Mathematical and General*, volume 26:page 781, 1993.

[Nag93b] K. Nagel and H. J. Herrmann. Deterministic models for traffic jams. *Physica A*, (199):page 254, 1993.

[Nag93c] A. Nagurney. *Network economics: a variational inequality approach*. Kluwer Academic Publisher, Boston, USA, 1993.

[Nag94a] T. Nagatani. Traffic jam and shock formation in stochastic traffic-flow model of a two-lane roadway. *Journal of the Physical Society of Japan*, volume 63:page 52, 1994.

[Nag94b] K. Nagel. Life-times of simulated traffic jams. *International Journal of Modern Physics C*, volume 5(3):pages 567–580, 1994.

[Nag94c] K. Nagel and S. Rasmussen. Traffic at the edge of chaos. In R. A. Brooks and P. Maes, editors, *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, page 222. 1994.

[Nag95a] K. Nagel. *High-speed microsimulations of traffic flow*. Ph.D. thesis, Universität zu Köln, Mar 1995.

[Nag95b] K. Nagel and M. Paczuski. Emergent traffic jams. *Physical Review E*, volume 51(4):pages 2909–2918, Apr 1995.

[Nag95c] K. Nagel and M. Schreckenberg. *Traffic Jam Dynamics in Stochastic Cellular Automata*. LA-UR 95-2132, Los Alamos National Laboratory, 1995.

[Nag96] K. Nagel. Particle hopping models and traffic flow theory. *Physical Review E*, volume 53(5):pages 4655–4672, May 1996.

[Nag98a] K. Nagel. Experiences with iterated traffic microsimulations in dallas. In D. Wolf and M. Schreckenberg, editors, *Proceedings of the Workshop on Traffic and Granular Flow '98*. Springer Verlag, 1998.

[Nag98b] K. Nagel, M. Rickert, P. M. Simon, and M. Pieck. The dynamics of iterated transportation simulations. In *Proceedings of the 3rd Triennial Symposium on Transportation Analysis (TRISTAN)*. San Juan, Porto Rico, Jun 1998.

[Nag98c] K. Nagel, P. Stretz, M. Pieck, S. Leckey, R. Donnelly, and C. L. Barrett. *TRANSIMS traffic flow characteristics*. Technical report, Los Alamos National Laboratory, Jun 1998.

[Nag98d] K. Nagel, D. E. Wolf, P. Wagner, and P. Simon. Two-lane traffic rules for cellular automata: A systematic approach. *Physical Review E*, volume 58(2):pages 1425–1437, Aug 1998.

[Nag01] K. Nagel and M. Rickert. Parallel implementation of the TRANSIMS microsimulation. *Parallel Computing*, volume 27(12):pages 1611–1639, 2001.

[Nag02a] K. Nagel. Distributed intelligence in large scale traffic simulations on parallel computers. In *Collective Cognition: Mathematical Foundations of Distributed Intelligence*. Santa Fe Institute, Aug 2002.

[Nag02b] K. Nagel. Traffic networks. In S. Bornholdt and H. Schuster, editors, *Handbook on Networks*. Jan 2002.

[Nag03a] K. Nagel, P. Wagner, and R. Woesler. Still flowing: old and new approaches for traffic flow modeling. *Operations Research*, volume 51(5):pages 681–710, 2003.

[Nag03b] A. Nagurney. Influence of Beckmann, McGuire, and Winsten's Studies in the Economics of Transportation. In *Proceedings of the 50th North American Regional Science Association Meeting*. Philadelphia, Pennsylvania, Nov 2003.

[Nag04] K. Nagel and B. Raney. Transportation planning II: Complex systems applications for transportation planning. In M. Koll-Schretzenmayr, M. Keiner, and G. Nussbaumer, editors, *The Real and Virtual Worlds of Spatial Planning*, chapter 16. Springer-Verlag, Berlin, 2004.

[Nag05] K. Nagel and P. Nelson. A critical comparison of the kinematic-wave model with observational data. In H. S. Mahmassani, editor, *Proceedings of the 16th International Symposium on Transportation and Traffic Theory (ISTTT16)*. University of Maryland, Jul 2005.

[Nam96] D. Nam and D. Drew. Traffic dynamics: method for estimating freeway travel times in real time from flow measurements. *Journal of Transportation Engineering*, volume 122:pages 185–191, 1996.

[Nas51] J. Nash. Non-cooperative games. *Annals of Mathematics Journal*, volume 54:pages 286–295, 1951.

[Nas82] C. Nash. A reformulation of the theory of optimal congestion taxes: a comment. *Journal of Transport Economics and Policy*, volume 26:pages 295–299, 1982.

[Nel95] P. Nelson. A kinetic model of vehicular traffic and its associated bimodal equilibrium solutions. *Transport Theory and Statistical Physics*, volume 24(383), 1995.

[Nel98] P. Nelson and A. Sopasakis. The Prigogine-Herman kinetic model predicts widely scattered traffic flow data at high concentrations. *Transportation Research B*, volume 32(8):pages 589–604, Nov 1998.

[Nes02] R. van Nes. *Design of multimodal transport networks - A hierarchical approach*. Ph.D. thesis, Technische Universiteit Delft, Delft, The Netherlands, Sep 2002.

[Neu48] J. von Neumann. The general and logical theory of automata. In L. A. Jeffress, editor, *Cerebral Mechanisms in Behavior*, pages 1–41. John Wiley, New York, Sep 1948. Paper presented at the Hixon Symposium.

[Neu99] L. Neubert, L. Santen, A. Schadschneider, and M. Schreckenberg. Single-vehicle data of highway traffic: A statistical analysis. *Physical Review E*, volume 60(6):pages 6480–6490, Dec 1999.

[New63a] G. F. Newell. Instability in dense highway traffic, a review. In *Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow*, pages 73–83. OECD, London, 1963.

[New63b] G. F. Newell. Theories of instability in dense highway traffic. *Operations Research Society Japan*, volume 5:pages 9–54, 1963.

[New82] G. F. Newell. *Applications of Queueing Theory*. Chapman & Hall, 2nd edition, 1982. ISBN 0412245000.

[New93a] G. F. Newell. A simplified theory of kinematic waves in highway traffic, part 1: General theory. *Transportation Research B*, volume 27(4):pages 281–287, 1993.

[New93b] G. F. Newell. A simplified theory of kinematic waves in highway traffic, part 2: Queueing at freeway bottlenecks. *Transportation Research B*, volume 27(4):pages 289–303, 1993.

[New93c] G. F. Newell. A simplified theory of kinematic waves in highway traffic, part 3: Multi-destination flows. *Transportation Research B*, volume 27(4):pages 305–313, 1993.

[New96] G. F. Newell. Theory of highway traffic flow 1945-1965. *Transportation Research A*, volume 30(6):page 479, Nov 1996.

[New98] G. F. Newell. A moving bottleneck. *Transportation Research B*, volume 32(8):pages 531–537, 1998.

[New99] G. F. Newell. Delays caused by a queue at a freeway exit ramp. *Transportation Research B*, volume 33:pages 337–350, 1999.

[New02a] G. F. Newell. Memoirs on highway traffic flow theory in the 1950s. *Operations Research*, volume 50(1):pages 173–178, Jan 2002.

[New02b] G. F. Newell. A simplified car-following theory: a lower order model. *Transportation Research B*, volume 36(3):pages 195–205, Mar 2002.

[Ngo03] D. Ngoduy, S. Hoogendoorn, and H. van Zuylen. Cross-comparison of numerical schemes for macroscopic traffic flow models. *Transportation Research Record*, 2003.

[Nih87] N. L. Nihan and G. A. Davis. Recursive estimation of origin-destination matrices from input-output counts. *Transportation Research B*, volume 21(2):pages 149–163, 1987.

[Nis99] K. Nishinari and D. Takahashi. A new deterministic CA model for traffic flow with multiple states. *Journal of Physics A: Mathematical and General*, volume 32:pages 93–104, 1999.

[Nis01a] K. Nishinari. Euler and Lagrange representation of traffic models. In *Proceedings of the Workshop on Traffic and Granular Flow '01*. Nagoya University, Japan, Oct 2001.

[Nis01b] K. Nishinari. A Lagrange representation of cellular automaton models of traffic flow. *Journal of Physics A: Mathematical and General*, volume 34:pages 10727–10736, 2001.

[Nis03a] K. Nishinari, D. Chowdhury, and A. Schadschneider. Cluster formation and anomalous fundamental diagram in an ant trail model. *Physical Review E*, volume 67(3):pages 6120–1–6120–11, Mar 2003.

[Nis03b] K. Nishinari, M. Treiber, and D. Helbing. Experimental time gap distributions and erratic scattering in heterogeneous flow. *Physical Review E*, volume 68(6):page 7101, 2003.

[Nis03c] K. Nishinari, M. Treiber, and D. Helbing. Interpreting the wide scattering of synchronized traffic data by time gap statistics. *Physical Review E*, volume 68(6):page 7101, Dec 2003.

[Nis04] K. Nishinari, M. Fukui, and A. Schadschneider. A stochastic cellular automaton model for traffic flow with multiple metastable states. *Journal of Physics A: Mathematical and General*, volume 37:pages 3101–3110, 2004.

[Nis05] Y. Nishimura, T. Cheon, and P. Šeba. *Metastable Congested States in Multisegment Traffic Cellular Automaton*, Jul 2005.

[Oht01a] H. Ohta. Probing a traffic congestion controversy: density and flow scrutinized. *Journal of Regional Science*, volume 41:pages 659–680, 2001.

[Oht01b] H. Ohta. Probing a traffic congestion controversy: response to comment. *Journal of Regional Science*, volume 10:pages 695–699, 2001.

[Orr02] P. Orrick. Gordon Newell memorialized at TRB. *ITS Review Online*, volume 1(1), Jul 2002.

[Ort01] J. Ortuzar and L. G. Willumsen. *Modelling Transport*. Wiley, 3rd edition, Oct 2001. ISBN 0-471-86110-3.

[Pac80] N. Packard, J. Crutchfeld, J. Farmer, and R. Shaw. Geometry from a time series. *Physical Review Letters*, volume 45:pages 712–716, 1980.

[Pah06] S. Pahaut and C. Sikow. History of thought and prospects for road pricing. *Transport Policy*, volume 13(2):pages 173–176, Mar 2006.

[Pal02] A. de Palma and F. Marchal. Real cases applications of the fully dynamic METROPOLIS tool-box: an advocacy for large-scale mesoscopic transportation systems. *Networks and Spatial Economics*, volume 2(4):pages 347–369, 2002.

[Pan05] S. Panwai and H. Dia. Comparative evaluation of microscopic car-following behavior. *IEEE Transactions on Intelligent Transportation Systems*, volume 6(3):pages 314–325, Sep 2005.

[Pap98] M. Papageorgiou. Some remarks on macroscopic traffic flow modelling. *Transportation Research A*, volume 32(5):pages 323–329, 1998.

[Par98] U. Parlitz. Nonlinear time-series analysis. In J. A. Suykens and J. Vandewalle, editors, *Nonlinear Modeling: Advanced Black-Box Techniques*, chapter 8, pages 209–253. Kluwer Academic Publishers Boston, Jun 1998. ISBN 0-7923-8195-5.

[Pas97] E. I. Pas and S. L. Principio. Braess' paradox: Some new insights. *Transportation Research B*, volume 31(3):pages 265–276, 1997.

[Pat94] M. Patriksson. *The Traffic Assignment Problem: Models and Methods*. VSP International Science Publishers, Utrecht, The Netherlands, 1994. ISBN 90-6764-181-2.

[Pat99] M. Patriksson. *Nonlinear programming and variational inequality problems: a unified approach*. Kluwer Academic Publisher, Dordrecht, The Netherlands, 1999.

[Pav75] S. Paveri-Fontana. On Boltzmann like treatments of traffic flow. *Transportation Research*, volume 9(225), 1975.

[Pay71] H. J. Payne. Models of freeway traffic and control. *Mathematical Models of Public Systems (Simulation Council Proceedings)*, volume 1:pages 51–61, 1971.

[Pay76] H. Payne, E. Helfenbein, and H. Knobel. *Development and testing of incident detection algorithms*. FHWA-RD-76-20, Federal Highway Administration, Washington, D.C., 1976.

[Pay78] H. J. Payne. *FREFLO: A macroscopic simulation model of freeway traffic*. Version 1 – User's guide, ESSOR, 1978.

[Pee01] S. Peeta and A. K. Ziliaskopoulos. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics,* volume 1:pages 233–265, 2001. Kluwer Academic Publishers, The Netherlands.

[Pee03] S. Peeta and T.-H. Yang. Stability issues for dynamic traffic assignment. *Automatica*, volume 39:pages 21–34, 2003.

[Pet62] C. A. Petri. *Kommunikation mit Automaten*. Schriften des IIM, Institut für Instrumentelle Mathematik, Bonn, Germany, 1962.

[Pet98] K. F. Petty, P. Bickel, M. Ostland, J. Rice, F. Schoenberg, J. Jiang, and Y. Ritov. Accurate estimation of travel times from single-loop detectors. *Transportation Research A*, volume 32(1):pages 1–17, 1998.

[Phi79] W. Phillips. Kinetic model for traffic flow with continuum implications. *Transportation Research Planning and Technology,* volume 5:pages 131–138, 1979.

[Pig20] A. C. Pigou. *The Economics of Welfare*. Macmillan and Co., 4th (1932) edition, 1920.

[Pip53] L. Pipes. An operational analysis of traffic dynamics. *Journal of Applied Physics*, volume 24(3):pages 274–281, 1953.

[Pip64] L. Pipes. Hydrodynamic approaches. In D. Gerlough and D. Capelle, editors, *An Introduction to Traffic Flow Theory*. Highway Research Board, Washington, D.C., 1964. Special Report 79.

[Pot04] A. Pottmeier, S. F. Hafstein, R. Chrobok, and M. Schreckenberg. *OLSIM: An Approved Traffic Information System*. Physics of Transport and Traffic, University of Duisburg-Essen, 2004.

[Pri60] I. Prigogine and F. C. Andrews. A Boltzmann like approach for traffic flow. *Operations Research*, volume 8(789), 1960.

[Pri71] I. Prigogine and R. Herman. *Kinetic Theory of Vehicular Traffic*. Elsevier, New York, 1971.

[Pri05] J. F. Price. *Lagrangian and Eulerian Representations of Fluid Flow*. Woods Hole Oceanographic Institution, Woodshole, Massachusetts, USA, 02543, Jan 2005.

[PTV05] PTV. *PTV Vision software suite (VISUM/VISEM/VISEVA/VISSIM)*. PTV AG – Traffic Mobility Logistics, 2005.

[Pus94] A. Pushkar, F. Hall, and J. Acha-Daza. Estimation of speeds from single-loop freeway flow and occupancy data using cusp catastrophe theory model. *Transportation Research Record*, volume 1457:pages 149–157, 1994.

[Raj96] N. Rajewsky, A. Schadschneider, and M. Schreckenberg. The asymmetric exclusion model with sequential update. *Journal of Physics A: Mathematical and General*, volume 29:pages 305–309, 1996.

[Raj98] N. Rajewsky, L. Santen, A. Schadschneider, and M. Schreckenberg. The asymmetric exclusion process: Comparison of update procedures. *Journal of Statistical Physics*, volume 92:pages 151–194, 1998.

[Rak96] H. Rakha, B. Hellinga, M. V. Aerde, and W. Perez. Systematic verification, validation and calibration of traffic simulation models. In *Proceedings of the 75th Annual Meeting of the Transportation Research Board*. Washington, D.C., Jan 1996.

[Rak98] H. Rakha, M. V. Aerde, L. Bloomberg, and X. Huang. Construction and calibration of a large-scale micro-simulation model of the Salt Lake area. *Transportation Research Record*, (1664):pages 93–102, 1998.

[Rak05] H. Rakha and W. Zhang. Estimating traffic stream space-mean speed and reliability from dual and single loop detectors. In *Transportation Research Board 84th Annual Meeting*, 05-0850. Washington, D.C., 2005.

[Ram27] F. P. Ramsey. A contribution to the theory of taxation. *Economic Journal*, volume 37:pages 47–61, Mar 1927.

[Ran03] B. Raney, N. Cetin, A. Völlmy, M. Vrtic, K. Axhausen, and K. Nagel. An agent-based microsimulation model of swiss travel – first results. *Networks and Spatial Economics*, volume 3:pages 23–41, 2003.

[Ras02] M. Rascle. An improved macroscopic model of traffic flow: Derivation and links with the lighthill-whitham model. *Mathematical and Computer Modelling*, (35):pages 581–590, 2002.

[Ray00] E. S. Raymond. *The Cathedral and the Bazaar*, 2000.

[Rec86a] W. Recker, M. McNally, and G. Root. A model of complex travel behavior: Part I – theory. *Transportation Research A*, volume 20(4):pages 307–318, 1986.

[Rec86b] W. Recker, M. McNally, and G. Root. A model of complex travel behavior: Part II - operational model. *Transportation Research A*, volume 20(4):pages 319–330, 1986.

[Rec95] W. Recker. The household activity pattern problem: General formulation and solution. *Transportation Research B*, volume 29(1):pages 61–77, 1995.

[Reu50] A. Reuschel. Fahrzeugbewegungen in der Kolonne bei gleichfoermig beschleunigtem oder verzoegertem Leifahrzeug. *Z Oesterr Ingr Architekt Vereines*, volume 95:pages 59–62,73–77, 1950.

[Ric56] P. I. Richards. Shockwaves on the highway. *Operations Research*, volume 4:pages 42–51, 1956.

[Ric96a] M. Rickert, K. Nagel, M. Schreckenberg, and A. Latour. Two lane traffic simulations using cellular automata. *Physica A*, volume 231:page 534, 1996.

[Ric96b] M. Rickert and P. Wagner. Parallel real-time implementation of large-scale, route-plan-driven traffic simulation. *Internation Journal of Modern Physics C*, volume 7(2):pages 133–153, 1996.

[Ric96c] M. Rickert, P. Wagner, and C. Gawron. Real-time traffic simulation of the German Autobahn network. In *Proceedings of the 4th Workshop on Parallel Systems and Algorithms PASA'96*. World Scientific Publishing Co. Pte. Ltd., 1996.

[Ric97] M. Rickert. *Traffic Simulation on Distributed Memory Computers*. Ph.D. thesis, Universität zu Köln, Dec 1997.

[Rod05] J.-P. Rodrigue, B. Slack, and C. Comtois. *The Geography of Transport Systems*. Routledge, London, 2005.

[Rom94] E. D. Romph. *A Dynamic Traffic Assignment Model – Theory and Applications*. Ph.D. thesis, Delft Technical University, Delft, The Netherlands, 1994.

[Ros01] S. Rosswog, C. Gawron, S. Hasselberg, R. Böning, and P. Wagner. Computational aspects in traffic simulation problems. *Future Generation Computer Systems*, volume 17(5):pages 659–665, Mar 2001.

[Rou84] P. J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, volume 79:pages 871–880, 1984.

[Rou87] P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.

[Rou99] P. J. Rousseeuw and K. V. Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, volume 41:pages 212–223, Aug 1999.

[Rou02] T. Roughgarden. *Selfish Routing*. Ph.D. thesis, Faculty of the Graduate School of Cornell University, May 2002.

[Rou03] T. Roughgarden. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, volume 67(2):pages 341–364, Sep 2003.

[Rou06] J. Rouwendal and E. T. Verhoef. Basic economic principles of road pricing: From theory to applications. *Transport Policy*, volume 13(2):pages 106–114, Mar 2006.

[Rub87] D. Rubin. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York, 1987.

[San99] L. Santen. *Numerical Investigations of Discrete Models for Traffic Flow*. Ph.D. thesis, Universität zu Köln, 1999.

[Sar00] P. Sarkar. A brief history of cellular automata. *ACM Computing Surveys*, volume 32(1):pages 80–107, Mar 2000. ISSN 0360-0300.

[Sau93] T. Sauer. Time series prediction by using delay coordinate embedding. In A. Weigend and N. Gershenfeld, editors, *Time Series Prediction: Forecasting the Future and Understanding the Past*, volume XV. Addison-Wesley, Jul 1993.

[Sch91] H. Schütt. *Entwicklung und Erprobung eines sehr schnellen, bitorientierten Verkehrssimulationssystems für Straßennetze*. Technical Report 6, Schriftenreihe der AG Automatisierungstechnik, T.U. Hamburg, Hamburg, 1991.

[Sch95] M. Schreckenberg, A. Schadschneider, K. Nagel, and N. Ito. Discrete stochastic models for traffic flow. *Physical Review E*, volume 51(4):pages 2939–2949, Apr 1995.

[Sch97a] A. Schadschneider. Analytical approaches to CA for traffic flow: Approximations and exact solutions. In M. Schreckenberg and D. Wolf, editors, *Proceedings of the Workshop on Traffic and Granular Flow '97*. Springer, 1997.

[Sch97b] A. Schadschneider and M. Schreckenberg. Car-oriented mean-field theory for traffic flow models. *Journal of Physics A,* volume 30:pages 69–75, 1997.

[Sch97c] A. Schadschneider and M. Schreckenberg. Traffic flow models with 'slow-to-start' rules. *Annalen der Physik,* volume 6(7):pages 541–551, 1997.

[Sch98] A. Schadschneider and M. Schreckenberg. Garden of Eden states in traffic models. *Journal of Physics A,* volume 31:pages 225–231, 1998.

[Sch99a] A. Schadschneider. The Nagel-Schreckenberg model revisited. *European Physical Journal B*, volume 10(3):pages 573–582, Aug 1999.

[Sch99b] J. Schafer. Multiple imputation: a primer. *Statistical Methods in Medical Research*, volume 8(1):pages 3–15, Mar 1999.

[Sch00] A. Schadschneider. Statistical physics of traffic flow. *Physica A,* (285):page 101, 2000.

[Sch01] M. Schreckenberg, R.Barlović, W. Knospe, and H. Klüpfel. Statistical physics of cellular automata models for traffic flow. In K. H. Hoffmann and M. Schreiber, editors, *Computational Statistical Physics,* pages 113–126. Springer, Berlin, 2001.

[Sch02a] A. Schadschneider. Traffic flow: A statistical physics point of view. *Physica A*, volume 313:pages 153–187, 2002.

[Sch02b] J. Schafer and R. Yucel. Missing data: our view of the state of the art. *Psychological Methods,* volume 7(2):pages 147–177, 2002.

[Sha03] T. Shaw. *Performance Measures of Operational Effectiveness for Highway Segments and Systems*. NCHRP-Synthesis 311, National Cooperative Highway Research Program, Transportation Research Board, Washington D.C., 2003.

[She85] Y. Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Inc., Englewood Cliffs, NJ 07632, 1985. ISBN 71998500.

[Shi99] Y. Shi. Self-organization in BML traffic flow model: Analytical approaches. *Communications in Theoretical Physics,* volume 31:pages 85–90, 1999.

[Sie] Siemens Nederland N.V. – Afdeling Traffic & Infra Solutions. *Siemens Lusdetector LD4*.

[Sie92] Siemens. *Voertuigtellingen - Functionele Analyse*. v1.00 T643, Siemens N.V., Aug 1992.

[Sim98a] P. Simon and H. Gutowitz. A cellular automaton model for bi-directional traffic. *Physical Review E,* volume 57(2):pages 2441–2444, 1998.

[Sim98b] P. Simon and K. Nagel. Simplified cellular automaton model for city traffic. *Physical Review E,* volume 58(2):pages 1286–1295, Aug 1998.

[Sim99] P. M. Simon, J. Esser, and K. Nagel. Simple queueing model applied to the city of portland. *International Journal of Modern Physics C,* volume 10(5):pages 941–960, 1999.

[Sim02] N. Simmons, G. Gates, and J. Burr. Commercial applications arising from a floating vehicle data system in europe. In *Proceedings of the 9th World Congress on Intelligent Transport Systems*. 2002.

[Sma82] K. Small. The scheduling of consumer activities: work trips. *American Economic Review*, volume 72(3):pages 647–679, 1982.

[Sma87] K. Small. A discrete choice model for ordered alternatives. *American Economic Review*, volume 55(2):pages 409–424, 1987.

[Sma92] K. A. Small. Using the revenues from congestion pricing. *Transportation*, volume 19(4):pages 359–381, 1992.

[Sma03] K. A. Small and X. Chu. Hypercongestion. *Journal of Transport Economics and Policy*, volume 37(3):pages 319–352, Sep 2003.

[Sma05] K. A. Small. Road pricing and public transit: Unnoticed lessons from london. *Access*, volume 26, 2005. University of California Transportation Center.

[Sme67] R. Smeed. Some circumstances in which vehicles will reach their destinations earlier by starting later. *Transportation Science*, volume 1:pages 308–317, 1967.

[Smi02] B. L. Smith, B. M. Williams, and R. K. Oswald. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C*, volume 10:pages 303–321, 2002.

[Smo62] R. Smock. An interative assignment approach to capacity restraint on arterial networks. *Transportation Research Board*, volume 156:pages 1–13, 1962.

[Smu89] S. Smulders. *Control of Freeway Traffic Flow*. Ph.D. thesis, University of Twente, Enschede, The Netherlands, 1989.

[Spi90] H. Spiess. Conical volume delay functions. *Transportation Science*, volume 24(2):pages 153–158, 1990.

[Sta01] J. Stada, S. Logghe, G. D. Ceuster, and B. Immers. Time-of-day modeling using a quasi-dynamic equilibrium assignment approach. In *Biennial Symposium on Transportation Analysis TRISTAN IV*. Sao Miguel, Azores, Portugal, Jun 2001.

[Stü06] G. Stüben. *autobahn.NRW – Expansion for motorway network*. Ministry of Building and Transportation of Nordrhein-Westfalen, Germany, Mar 2006.
(URL: `http://www.autobahn.nrw.de`).

[Sun02] S. Sundaram. *Development of a Dynamic Traffic Assignment System for Short-Term Planning Operations*. Master's thesis, Massachusetts Institute of Technology, Jun 2002.

[Sun03] Sun Microsystems, Inc. *JavaSpaces Service Specification v2.0*, 2003.

[Sze04] W. Szeto and H. K. Lo. A cell-based simultaneous route and departure time choice model with elastic demand. *Transportation Research B*, volume 38(7):pages 593–612, Aug 2004.

[Sze06] W. Szeto and H. K. Lo. Dynamic traffic assignment: Properties and extensions. *Transportmetrica*, volume 2(1):pages 31–52, 2006.

[Taa00] H. Taale, A. de Hoog, S. Smulders, and O. Tool. The results of a Dutch experiment with floating car data. In *Proceedings of the 9th IFAC Symposium on Control in Transporation Systems*, volume I. Braunschweig, Germany, Jun 2000.

[Tak81] F. Takens. Detecting strange attractors in turbulence. In D. Rand and L.-S. Young, editors, *Dynamical Systems and Turbulence*, pages 366–381. Springer, Berlin, 1981.

[Tak93] M. Takayasu and H. Takayasu. $1/f$ noise in a traffic model. *Fractals*, volume 1(4):pages 860–866, 1993.

[Tam03] C. M. Tampère, S. P. Hoogendoorn, and B. van Arem. Capacity funnel explained using the human-kinetic traffic flow model. In *Proceedings of the workshop on Traffic and Granular Flow '03*. T.U. Delft, The Netherlands, Delft, Oct 2003.

[Tam04a] C. M. Tampère. *Human-Kinetic Multiclass Traffic Flow Theory and Modelling*. TRAIL Thesis Series nr. T2004/11, TRAIL Research School, P.O. Box 5017, 2600 GA Delft, The Netherlands, Dec 2004.

[Tam04b] C. M. Tampère, B. van Arem, and S. P. Hoogendoorn. Gas kinetic traffic flow modelling including continuous driver behaviour models. *Transportation Research Record*, volume 1852:pages 231–238, 2004.

[Tam06] C. Tampère, J. Stada, B. Immers, E. Peetermans, and K. Organe. Identifying vulnerable sections in a national road network – application to the road network of Flanders. In *Seminar Infrastructure Reliability*. Technische Universiteit Delft, Jun 2006.

[Tay03] N. B. Taylor. The CONTRAM dynamic traffic assignment model. *Networks and Spatial Economics*, volume 3:pages 297–322, 2003.

[Tha04] D. Thain, T. Tannenbaum, and M. Livny. Distributed computing in practice: The Condor experience. *Concurrency and Computation – Practice and Experience*, 2004.

[Thü26] J. H. von Thünen. *Der isolirte Staat in Beziehung auf Landwirthschaft und Nationalökonomie*, volume I. Hamburg, Perthes, 1826.

[Tim01] H. Timmermans. Models of activity scheduling behaviour. In *Proceedings of the 5th Workshop of The Nordic Research Network on Modelling Transport, Land-Use and the Environment*, pages 33–47. NorFA – The Nordic Academy for Advanced Study, Nynäshamn, Sweden, Sep 2001.

[Tim03] H. Timmermans. The saga of integrated land use-transport modeling: How many more dreams before we wake up ? In *Proceedings of the 10th International Conference on Travel Behaviour Research*. Lucerne, Switserland, Aug 2003.

[Tok96] T. Tokihiro, D. Takahashi, J. Matsukidaira, and J. Satsuma. From soliton equations to integrable cellular automata through a limiting procedure. *Physical Review Letters*, volume 76(18):pages 3247–3250, Apr 1996.

[Tre74] J. Treiterer and J. Myers. The hysteresis phenomena in traffic flow. In D. Buckley, editor, *Proceedings of the 6th Symposium on Transportation and Traffic Theory*, pages 13–38. 1974.

[Tre99] M. Treiber and D. Helbing. *Explanation of Observed Features of Self-Organization in Traffic Flow*, 1999. E-print cond-mat/9901239.

[Tre00] M. Treiber, A. Hennecke, and D. Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, volume 62:pages 1805–1824, 2000.

[Tre01] M. Treiber and D. Helbing. Microsimulations of freeway traffic including control measures. *Automatisierungstechnik*, volume 49:pages 478–484, 2001.

[Tre02] M. Treiber and D. Helbing. Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooperative Transportation Dynamics*, volume 1(1):pages 1–20, Feb 2002.

[Tre03] M. Treiber and D. Helbing. Memory effects in microscopic traffic models and wide scattering in flow-density data. *Physical Review E*, volume 68:page 046119, 2003.

[Tre05a] M. Treiber, A. Kesting, and D. Helbing. Delays, inaccuracies and anticipation in microscopic traffic models. *Physica A*, 2005.

[Tre05b] M. Treiber, A. Kesting, and D. Helbing. *Understanding widely scattered traffic flows, the capacity drop, platoons, and times-to-collision as effects of variance-driven time gaps*, Aug 2005. Submitted to Physical Review E.

[Tur52] A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London*, volume 237(641):pages 37–72, Aug 1952.

[Tur98] S. M. Turner, W. L. Eisele, R. J. Benz, and D. J. Holdener. *Travel Time Data Collection Handbook*. Report FHWA-PL-98-035, Texas Transportation Institute, The Texas A&M University System, Mar 1998.

[Tur99] D. L. Turcotte. Self-organized criticality. *Reports on Progress in Physics*, volume 62:pages 1377–1429, 1999.

[Und61] R. Underwood. Speed, volume, and density relationships – quality and theory of traffic flow. *Yale Bureau of Highway Traffic*, pages 141–188, 1961.

[Van00] N. Vandaele, T. V. Woensel, and A. Verbruggen. A queueing based traffic flow model. *Transportation Research D*, volume 5:pages 121–135, 2000.

[Var05] P. Varaiya. Reducing highway congestion: An empirical approach. *European Journal of Control*, volume 11:pages 301–309, 2005.

[Vel02] S. Velan and M. Florian. A note on the entropy solutions of the hydrodynamic model of traffic flow. *Transportation Science*, volume 36(4):pages 435–446, Nov 2002.

[Ven98] C. J. Venter. *The Timing of Activity and Travel Planning Decisions*. Ph.D. thesis, University of California, Berkeley, 1998.

[Ver98] E. T. Verhoef. Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing. In *Proceedings of the 38th Meeting of the European Regional Science Association*. Vienna, Austria, Aug 1998.

[Ver01] E. Verhoef. Probing a traffic congestion controversy: a comment. *Journal of Regional Science*, volume 41:pages 681–694, 2001.

[Ver03a] H. Versteegt, K. Malone, A. Martins, R. Castenmiller, and M. Bliemer. *Follow-up study on testing INDY for use as a Dynamic Traffic Assignment model in the Dutch National Model*. 03-7N-233-74021, TNO Inro, Dec 2003.

[Ver03b] H. Versteegt and C. Tampère. *PredicTime: State of the art and functional architecture*. 03-7N-024-73196, TNO INRO, The Netherlands, Feb 2003.

[Ver05a] S. Verboven and M. Hubert. LIBRA: a MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*, volume 75:pages 127–136, 2005.

[Ver05b] E. T. Verhoef. Speed-flow relations and cost functions for congested traffic: Theory and empirical analysis. *Transportation Research A*, 2005.

[Vic69] W. S. Vickrey. Congestion theory and transport investment. *American Economic Review*, volume 59:pages 251–260, 1969.

[Voe01] A. Voellmy, M. Vrtic, B. Raney, K. Axhausen, and K. Nagel. Status of a TRANSIMS implementation for Switzerland. *Networks and Spatial Economics*, Dec 2001.

[VVC03] Verkeerscentrum Vlaanderen, Departement Leefmilieu en Infrastructuur, Administratie Wegen en Verkeer, Vuurkruisenplein 20, 2020 Antwerpen. *MINDAT – Databank ruwe verkeersdata Vlaams snelwegennet – 2001 en 2003*, Feb 2003.

[Wad02] P. Waddell. Urbansim: Modeling urban development for land use, transportation and environmental planning. preprint. *Journal of the American Planning Association*, volume 68(3):pages 297–314, 2002.

[Wad04] P. Waddell and G. F. Ulfarsson. Introduction to urban simulation: Design and development of operational models. In D. A. Hensher, K. J. Button, K. E. Haynes, and P. Stopher, editors, *Handbook of Transport Geography and Spatial Systems*, volume 5 of *Handbooks in Transport*, chapter 13. Elsevier, Aug 2004.

[Wag97] P. Wagner, K. Nagel, and D. E. Wolf. Realistic multi-lane traffic rules for cellular automata. *Physica A*, volume 234:pages 687–698, 1997.

[Wah02] J. Wahle, R. Chrobok, A. Pottmeier, and M. Schreckenberg. A microscopic simulator for freeway traffic. *Networks and Spatial Economics*, volume 2:pages 371–386, 2002.

[Wal61] A. Walters. The theory and measurement of private and social cost of highway congestion. *Econometrica*, volume 29(4):pages 676–697, 1961.

[Wal04a] K. Waldeer. Numerical investigation of a mesoscopic vehicular traffic flow model based on a stochastic acceleration process. *Transport Theory and Statistical Physics*, volume 33(2):pages 31–46, 2004.

[Wal04b] K. Waldeer. A vehicular traffic flow model based on a stochastic acceleration process. *Transport Theory and Statistical Physics*, volume 33(2):pages 7–30, 2004.

[Wan96] B. Wang, Y. Woo, and P. Hui. Improved mean-field theory of two-dimensional traffic flow models. *Journal of Physics A*, volume 29:pages L31–L35, 1996.

[Wan97] R.-M. Wang. *An Activity-Based Trip Generation Model*. Ph.D. thesis, University of California, Irvine, 1997.

[Wan98] B.-H. Wang, Y.-R. Kwong, and P.-M. Hui. Statistical mechanical approach to Fukui-Ishibashi traffic flow models. *Physical Review E*, volume 57(3):pages 2568–2573, Mar 1998.

[Wan01] L. Wang, B.-H. Wang, and B. Hu. *A cellular automaton traffic flow model between the Fukui-Ishibashi and Nagel-Schreckenberg models*. Traffic Forum – Statistical Mechanics, Feb 2001.

[Wan03] Y. Wang, M. Papageorgiou, and A. Messmer. Motorway traffic state estimation based on extended Kalman filter. In *Proceedings of the European Control Conference ECC 03*. Cambridge, UK, Sep 2003.

[War52] J. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institution of Civil Engineers*, volume 1 of *2*. 1952.

[War54] J. Wardrop and G. Charlesworth. A method of estimating speed and flow of traffic from a moving vehicle. In *Proceedings of the Institution of Civil Engineers*, volume 3, pages 158–171. 1954.

[War05] P. Warffemius. Using the standard deviation of the travel time distribution as an indicator for valuing the reliability of travel time. In *ECTRI – European Conference of Transport Research Institutes, Young Researchers Seminar 2005*. National Institute for Coastal and Marine Management, The Hague, The Netherlands, May 2005.

[Web94] C. Webber and J. Zbilut. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, volume 76:pages 965–973, 1994.

[Weg96] R. Wegener and A. Klar. A kinetic model for vehicular traffic derived from a stochastic microscopic model. *Transportation Theory Statistical Physics*, volume 25(7):pages 785–798, 1996.

[Wer98] J. Werth. *Galilei-invariante Fahrzeugwechselwirkungen im Straßenverkehr*. Master's thesis, Duisburg University, Duisburg, Germany, 1998.

[Wer00] M. Wermuth, C. Sommer, and S. Wulff. TeleTravel Systems (TTS) – Telematic System for the Automatic Survey of Travel Behaviour. In *Proceedings of the 9th IFAC Symposium on Control in Transporation Systems*, volume I. Braunschweig, Germany, Jun 2000.

[Wes95] M. Westerman. *Real-Time Traffic Data Collection for Transportation Telematics*. Ph.D. thesis, Faculty of Civil Engineering, Delft University of Technology, Dec 1995.

[Whi74] G. B. Whitham. *Linear and Nonlinear Waves*. Wiley, New York, 1974.

[Wie74] R. Wiedemann. *Simulation des Straßenverkehrsflusses*. Technical report, Institutes für Verkehrswesen der Universität Karlsruhe, 1974.

[Win01] J. R. Windover and M. J. Cassidy. Some observed details of freeway traffic evolution. *Transportation Research A*, volume 35(10):pages 881–894, Dec 2001.

[Woe03] T. V. Woensel. *Models for uninterrupted traffic flows – A queueing approach*. Ph.D. thesis, Universiteit Antwerpen, Faculteit Toegepaste Economische Wetenschappen, UFSIA–RUCA, Jan 2003.

[Wol83] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, volume 55:pages 601–644, Jul 1983.

[Wol84a] S. Wolfram. Universality and complexity in cellular automata. *Physica D*, volume 10:pages 1–35, 1984.

[Wol84b] S. Wolfram, O. Martin, and A. Odlyzko. Algebraic properties of cellular automata. *Communications in Mathematical Physics*, volume 93:pages 219–258, Mar 1984.

[Wol99] D. E. Wolf. Cellular automata for traffic simulations. *Physica A*, volume 263:pages 438–451, 1999.

[Wol02] S. Wolfram. *A New Kind of Science*. Wolfram Media, Inc., 2002. ISBN 1-579-955008-8.

[Wöl05] M. Wölki, A. Schadschneider, and M. Schreckenberg. *Asymmetric exclusion processes with shuffled dynamics*, Sep 2005.

[Won02] S. Wong and G. Wong. An analytical shock-fitting algorithm for LWR kinematic wave model embedded with linear speed-density relationship. *Transportation Research B*, volume 36(8):pages 683–706, Sep 2002.

[Wu98] J. Wu, Y. Chen, and M. Florian. The continuous dynamic network loading problem: A mathematical formulation and solution method. *Transportation Research B*, volume 32(3):pages 173–187, 1998.

[Wue99] A. Wuensche. Classifying cellular automata automatically. *Complexity*, volume 4(3):pages 47–66, 1999.

[Yan52] C. Yang and T. Lee. Statistical theory of equations of state and phase transitions: I. Theory of condensation. *Physical Review*, volume 87(3):pages 404–409, Aug 1952.

[Yan97] Q. Yang. *A Simulation Laboratory for Evaluation of Dynamic Traffic Management Systems*. Ph.D. thesis, Massachusetts Institute of Technology, Jun 1997.

[Yan98a] H. Yang and H.-J. Huang. Principle of marginal-cost pricing: How does it work in a general road network ? *Transportation Research B*, volume 32(1):pages 45–54, 1998.

[Yan98b] H. Yang and Q. Meng. Departure time, route choice and congestion toll in a queueing network with elastic demand. *Transportation Research B*, volume 32(4):pages 247–260, 1998.

[Ype05a] I. Yperman, S. Logghe, and B. Immers. Strictly Pareto-improving congestion pricing in a multi-destination network. In *Proceedings of the 84th Annual Meeting of the Transportation Research Board*. Washington, D.C., 2005.

[Ype05b] I. Yperman, S. Logghe, C. Tampère, and B. Immers. *The Multi-Commodity Link Transmission Model for Dynamic Network Loading*, Nov 2005.

[Zah72] Y. Zahavi. Traffic performance evaluation of road networks by the α-relationship. *Traffic Engineering and Control*, volume 14(5–6), 1972.

[Zbi92] J. Zbilut and C. Webber. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, volume 171:pages 199–203, 1992.

[Zha97] X. Zhang and D. F. Jarrett. Stability analysis of the classical car-following model. *Transportation Research B*, volume 31(6):pages 441–462, 1997.

[Zha98] H. Zhang. A theory of nonequilibrium traffic flow. *Transportation Research B*, volume 32(7):pages 485–498, 1998.

[Zha99] H. Zhang. A mathematical theory of traffic hysteresis. *Transportation Research B*, volume 33:pages 1–23, 1999.

[Zha00] H. Zhang. Structural properties of solutions arising from a nonequilibrium traffic flow theory. *Transportation Research B*, volume 34:pages 583–603, 2000.

[Zha01a] H. Zhang. A finite difference approximation of a non-equilibrium traffic flow model. *Transportation Research B*, volume 34(4):pages 337–365, 2001.

[Zha01b] H. Zhang. A note on highway capacity. *Transportation Research B*, volume 35:pages 929–937, 2001.

[Zha02] H. Zhang. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research B*, volume 36(3):pages 275–290, Mar 2002.

[Zha03a] H. Zhang. Anisotropic property revisited – does it hold in multi-lane traffic ? *Transportation Research B*, volume 37(6):pages 561–577, Jul 2003.

[Zha03b] H. Zhang. Driver memory, traffic viscosity and a viscous vehicular traffic flow model. *Transportation Research B*, volume 37(1):pages 27–41, Jan 2003.

[Zha03c] H. Zhang. On the consistency of a class of traffic flow models. *Transportation Research B*, volume 37(1):pages 101–105, Jan 2003.

[Zha05] H. Zhang and T. Kim. A car-following theory for multiphase vehicular traffic flow. *Transportation Research B*, volume 39(5):pages 385–399, Jun 2005.

[Zus67] K. Zuse. Rechnender Raum. *Elektronische Datenverarbeitung*, volume 8:pages 336–344, 1967.

[Zus69] K. Zuse. Rechnender Raum. *Schriften zur Datenverarbeitung*, volume 1, 1969. Friedrich Vieweg & Sohn, Braunschweig, Germany.

[Zuy99] H. J. van Zuylen. *Het spel van de regels; Het richting geven aan chaotische maatschappelijke processen*. Technische Universiteit Delft, 1999.

List of publications

Journal papers

- Maerivoet S. and De Moor B., *"Non-concave fundamental diagrams and phase transitions in a stochastic traffic cellular automaton"*, in The European Physical Journal B – Condensed Matter Physics, vol. 42, nr. 1, pages 131–140, november 2004.
- Maerivoet S. and De Moor B., *"Cellular Automata Models of Road Traffic"*, in Physics Reports, vol. 419, nr. 1, pages 1–64, november 2005.

Conference papers

- Maerivoet S., Logghe S., Immers L.H., and De Moor B., *"A comparison of a cellular automaton and a macroscopic model"*, in Proceedings of the Workshop on Traffic and Granular Flow '03, Technical University of Delft, The Netherlands, october 2003.
- Maerivoet S. and De Moor B., *"Development of an improved traffic cellular automaton model for traffic flows on a highway road network"*, in Proceedings of the 10th World Congress and Exhibition of Intelligent Transport Systems and Services (ITSS03), Madrid, Spain, november 2003.

Internal reports

- Maerivoet S. and De Moor B., *"Advancing density waves and phase transitions in a velocity dependent randomization traffic cellular automaton"*, Internal Report 03-111, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), october 2004.
- Maerivoet S. and De Moor B., *"Graduate Students and the Art of Public Speaking"*, Internal Report 04-206, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), september 2005.
- Maerivoet S., Logghe S., Immers B., and De Moor B., *"Relating the dynamics of road traffic in a stochastic cellular automaton to a macroscopic first-order model"*, Internal Report 05-150, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), june 2005.
- Maerivoet S. and De Moor B., *"Traffic Flow Theory"*, Internal Report 05-154, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), july 2005.

- Maerivoet S. and De Moor B., *"Transportation Planning and Traffic Flow Models"*, Internal Report 05-155, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), july 2005.
- Maerivoet S. and De Moor B., *"Data Quality, Travel Time Estimation, and Reliability"*, Internal Report 06-030, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), march 2006.
- Maerivoet S. and De Moor B., *"Sustainability Effects of Traffic Management Systems"*, Internal Report 06-033, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), Final report to the DWTC, march 2006.
- Maerivoet S. and De Moor B., *"Dynamic Traffic Assignment based on Cellular Automata"*, Internal Report 06-058, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), march 2006.
- Maerivoet S. and De Moor B., *"Some Thoughts on Obtaining a PhD"*, Internal Report 06-100, ESAT-SCD (SISTA), K.U.Leuven (Leuven, Belgium), may 2006.

Abstracts

- Maerivoet S., Bellemans T., and De Moor B., *"A Case Study on Traffic Flow Modelling, Simulation and Control"*, IAP Study day, Louvain-la-Neuve, Belgium, November 26, 2002.
- Maerivoet S. and De Moor B., *"Modeling, Simulation and Control of Traffic Flows on a Highway Network"*, Faculty of Engineering PhD Symposium, Katholieke Universiteit Leuven, Leuven, Belgium, December 11, 2002.
- Maerivoet S. and De Moor B., *"Development of a Traffic Cellular Automaton Model for Highway Traffic"*, 22nd Benelux Meeting on Systems and Control, De Vossemeren, Lommel, Belgium, March 21, 2003.
- Maerivoet S., Logghe S., Immers L.H., and De Moor B., *"A Comparison of a Cellular Automaton and a Macroscopic Model"*, Workshop on Traffic and Granular Flow 2003, Technical University of Delft, The Netherlands, October 1–3, 2003.
- Maerivoet S. and De Moor B., *"Development of an Improved Traffic Cellular Automaton Model for Traffic Flows on a Highway Road Network"*, 10th World Congress and Exhibition on Intelligent Transport Systems and Services, Madrid, Spain, November 16–20, 2003.
- Maerivoet S. and De Moor B., *"Efficient Microscopic Simulation of Large Scale Highway Traffic Flows"*, 23rd Benelux Meeting on Systems and Control, Helvoirt, The Netherlands, March 17, 2004.
- Maerivoet S. and De Moor B., *"Sustainability Effects of Traffic Management Systems"*, The Transport Science and Technology Congress – TransTec04, Athens Hilton Hotel, Athens, Greece, September 1–5, 2004.

List of presentations

Lectures and seminars

- Maerivoet S., *"Traffic Flow Seminar I: An introduction to the modeling and simulation of traffic flows"*, SISTA-seminars, Katholieke Universiteit Leuven, Leuven, Belgium, June 6, 2002.
- Maerivoet S., *"Modelleren, simuleren en beheersen van verkeersstromen"*, HK05 – Gevallenstudies van datamining en industriële procescontrole, Katholieke Universiteit Leuven, Leuven, Belgium, October 9, 2002.
- Maerivoet S., *"Modelleren en simuleren van verkeersstromen"*, D266 – Economische Sectoren en Technologie, Katholieke Universiteit Leuven, Leuven, Belgium, March 17, 2003.
- Maerivoet S., and De Moor B., *"Development of a Traffic Cellular Automaton Model for Highway Traffic"*, 22nd Benelux Meeting on Systems and Control, De Vossemeren, Lommel, Belgium, March 21, 2003.
- Maerivoet S., *"Traffic Flow Seminar II: A Survey on the Qualitative Behaviour of Traffic Cellular Automata"*, SISTA-seminars, Katholieke Universiteit Leuven, Leuven, Belgium, April 10, 2003.
- Maerivoet S., and Vens C., *"Academic Oratory Skills - Mastering original and persuasive speech"*, Academic English, CLT/ILT, Katholieke Universiteit Leuven, Leuven, Belgium, May 16, 2003.
- Maerivoet S., *"Traffic and Chickens"*, SISTA Brainstorm Sessions, Katholieke Universiteit Leuven, Leuven, Belgium, June 16, 2003.
- Maerivoet S., *"Modelleren, simuleren en beheersen van verkeersstromen"*, HK05 – Gevallenstudies van datamining en industriële procescontrole, Katholieke Universiteit Leuven, Leuven, Belgium, October 29, 2003.
- Maerivoet S., and De Moor B., *"Development of an Improved Traffic Cellular Automaton Model for Traffic Flows on a Highway Road Network"*, 10th World Congress and Exhibition on Intelligent Transport Systems and Services, Madrid, Spain, November 16–20, 2003.
- Maerivoet S., *"Traffic: an Interplay between Models, Simulations, and Control Actions"*, Friday Seminar, University of Antwerp, Antwerp, Belgium, March 5, 2004.

- Maerivoet S., *"Verkeerskunde: een Samenspel van Modellen, Simulaties en Regelacties"*, D266 – Economische Sectoren en Technologie, Katholieke Universiteit Leuven, Leuven, Belgium, March 15, 2004.
- Maerivoet S., and De Moor B., *"Efficient Microscopic Simulation of Large Scale Highway Traffic Flows"*, 23rd Benelux Meeting on Systems and Control, Helvoirt, The Netherlands, March 17, 2004.
- Maerivoet S., *"Verkeer SISTA"*, Transport & Mobility Leuven (TML) seminar, Leuven, Belgium, April 28, 2004.
- Maerivoet S., *"Models in Aid of Traffic Management"*, Seminarie transportmodellen ter ondersteuning van het mobiliteits- vervoersbeleid in België, Brussels, Belgium, May 3, 2004.
- Maerivoet S., *"Traffic SISTA"*, SYSTeMS Dialog, University of Gent, Gent, Belgium, May 26, 2004.
- Maerivoet S., and De Moor B., *"Sustainability Effects of Traffic Management Systems"*, The Transport Science and Technology Congress – TransTec04, Athens Hilton Hotel, Athens, Greece, September 1–5, 2004.
- Maerivoet S., Vandewalle J., and De Moor B., *"Evalueren aan de hand van papers"*, Duo/Icto summercourse *"Doceren in het BaMa tijdperk"*, Katholieke Universiteit Leuven, Leuven, Belgium, September 16, 2004.
- Maerivoet S., *"Verkeerskunde: een samenspel van modellen, simulaties en regelacties"*, HK05 – Gevallenstudies van datamining en industriële procescontrole, Katholieke Universiteit Leuven, Leuven, Belgium, September 29, 2004.
- Maerivoet S., *"Verkeerskunde: een samenspel van modellen, simulaties en regelacties"*, D266 – Economische Sectoren en Technologie, Katholieke Universiteit Leuven, Leuven, Belgium, April 14, 2005.
- Maerivoet S., *"Real-Time Models of Road Traffic"*, PhD Doctoral Research Intermediate Evaluation, Katholieke Universiteit Leuven, Leuven, Belgium, April 29, 2005.
- Maerivoet S., *"Verkeerskunde: een samenspel van modellen, simulaties en regelacties"*, HK05 – Gevallenstudies van datamining en industriële procescontrole, Katholieke Universiteit Leuven, Leuven, Belgium, October 5, 2005.

Posters

- Maerivoet S., Bellemans T., and De Moor B., *"A Case Study on Traffic Flow Modelling, Simulation and Control"*, IAP Study day, Louvain-la-Neuve, Belgium, November 26, 2002.
- Maerivoet S., and Bart De Moor, *"Modeling, Simulation and Control of Traffic Flows on a Highway Network"*, Faculty of Engineering PhD Symposium, Katholieke Universiteit Leuven, Leuven, Belgium, December 11, 2002.
  Winner of the Best Poster Award.
- Maerivoet S., Logghe S., Immers L.H., and De Moor B., *"A Comparison of a Cellular Automaton and a Macroscopic Model"*, Workshop on Traffic and Granular Flow 2003, Technical University of Delft, The Netherlands, October 1–3, 2003.
- Maerivoet S., *"Modelleren, Simuleren en Regelen van Verkeersstromen"*, Dag van de Technologie, Katholieke Universiteit Leuven, Leuven, Belgium, March 14, 2004.

Project presentations

- Maerivoet S., and Bellemans T., *"PODO-II – CP/40: Sustainability Effects of Traffic Management Systems"*, DWTC kick-off meeting, Katholieke Universiteit Leuven, Leuven, Belgium, January 18, 2002.
- Maerivoet S., *"PODO-II – CP/40: Sustainability Effects of Traffic Management Systems"*, DWTC ontmoetingsdagen, Brussels, Belgium, March 21, 2002.
- Maerivoet S., et al, *"PODO-II – CP/40: Sustainability Effects of Traffic Management Systems"*, DWTC internal meeting with project consortium, Katholieke Universiteit Leuven, Leuven, Belgium, May 17, 2002.
- Maerivoet S., et al, *"PODO-II – CP/40: Sustainability Effects of Traffic Management Systems"*, DWTC internal meeting with the user committee, Katholieke Universiteit Leuven, Leuven, Belgium, June 13, 2002.
- Maerivoet S., et al, *"PODO-II – CP/40: Sustainability Effects of Traffic Management Systems"*, DWTC first intermediary evaluation, Brussels, Belgium, April 2, 2003.
- Maerivoet S., et al, *"PODO-II – CP/40: Sustainability Effects of Traffic Management Systems"*, DWTC meeting with the user committee, Katholieke Universiteit Leuven, Leuven, Belgium, December 12, 2003.

About the author

Sven Maerivoet was born on April 11, 1976 in Borgerhout, Belgium. He studied Computer Science at the University of Antwerp, Belgium, and received his Master's degree (Licentiaat) in July 2001. His Master's thesis was titled *"The use of microscopic traffic simulation in a study aimed at the congestion problems on Antwerp's ring road"*, in which he developed a microsimulation model for the propagation of traffic on motorways. It was obtained at the Department of Mathematics and Computer Science at the University of Antwerp, Belgium, under the supervision of prof. dr. Serge Demeyer.

Since september 2001, he has been working as a doctoral student on the subject of traffic flow modelling in the lab *"Signals, Identification, System Theory, and Automation"* (SISTA), at the Department of Electrical Engineering ESAT of the Katholieke Universiteit Leuven. His supervisors were prof. dr. ir. Bart De Moor and prof. ir. Ben Immers. He was the principal co-ordinator of a three-year project called *"Sustainability Effects of Traffic Management Systems"*, funded by a grant from the Federal Science Policy DWTC (Federal Office for Scientific, Technical, and Cultural Affairs). He also gave a number of lectures on the modelling, simulation, and control of traffic flows. His expertise mainly lies in the development of fast microsimulation models and numerical traffic data.