Integrated Anticipatory Control of Road Networks

A game-theoretical approach

Henk Taale

5th November 2008

Integrated Anticipatory Control of Road Networks

A game-theoretical approach

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op vrijdag 5 december 2008 om 12:30 uur

door

Hendrik TAALE

wiskundig ingenieur

geboren te Middelharnis

Dit proefschrift is goedgekeurd door de promotor:

Prof.dr. H.J. van Zuylen

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr. H.J. van Zuylen	Technische Universiteit Delft, promotor
Prof.dr.ir. S.P. Hoogendoorn	Technische Universiteit Delft
Prof.dr.ir. B. De Schutter	Technische Universiteit Delft
Prof.dr. M. Smith	University of York
Prof.dr. N.H. Gartner	University of Massachusetts Lowell
Prof.dr.ir. E.C. van Berkum	Universiteit Twente

"But the path of the just is as the shining light,

that shineth more and more unto the perfect day.

The way of the wicked is as darkness:

they know not at what they stumble."

— Proverbs 4:18-19

Preface

Honour to whom honour is due, there is One whom I should thank the most: the LORD God, Creator and Saviour. He provided me with health and opportunities to fulfil this task. I am thankful for that and want to acknowledge His name. Although the task was quite a big one, considering I started off in 1999 and finished nine years later, there was always Sandra to encourage me and to keep me focussed. That I finished this thesis, is for a large part due to her continuous support, love and faith in me. I am also grateful to my sons Rick and Frank, who provided me during these years with the necessary distractions, mostly consisting of soccer, basketball, computer games, etc. Thanks guys, you compensated the "lack of friends" ;-). Thanks go to my parents, who made it possible for me to study and get acquainted with science and the beauty and fun of it. That I could continue this acquaintance, was also not possible without the support of my parents-in-law. Also thanks to the rest of my family, friends and colleagues within and outside Rijkswaterstaat, who persisted in asking me when I was ready with my thesis and therefore contributed to it significantly.

Especially, I would like to thank professor Henk van Zuylen. He asked me to start this research in the first place and during all these years he kept faith in me and provided me with suggestions to explore new research directions and to find new ways solving problems. Also thanks to all my colleagues, fellow PhD students and room-mates (I don't know why, but I had quite a few) of the faculty. When I did my research at the university for one day in the week, I always enjoyed the hospitality, the company and the games of table tennis. Besides that: Theo and Serge, thanks for reading drafts; Femke, Serge and Victor, thanks for the proof in Appendix B.

Finally, I would like to express my thanks and gratitude to the Centre for Transport and Navigation (formerly known as the AVV Transport Research Centre) of Rijkswaterstaat for permitting me to start this work, for allowing to do it partially during working hours and for giving me enough trust and time to finish it. I am sure that Rijkswaterstaat can profit from its results.

Contents

List of Figures

List of Tables

Notation

This section lists the symbols used. They are grouped into the general categories of indices, sets and variables. The variables are related to the control part, the dynamic traffic assignment (DTA) model and the dynamic network loading (DNL) model. The symbols are presented in an alphabetical order, first in normal case, then the capitals. Greek symbols are placed in between the normal characters, based on the position in the Greek alphabet. After the symbols a list of abbreviations is given.

Indices:

a	:	index for a link
d	:	index for a destination
i	:	index for a maximum conflict group and for an incoming link in the DNL model
j	:	index for an iteration in the assignment and for an outgoing link in the DNL model
k	:	index for a time period
k'	:	index for a time period
m	:	index for a movement of a signal controlled intersection
m'	:	index for a movement of a signal controlled intersection
n	:	index for a node
o	:	index for an origin
r	:	index for a route
t	:	index for a time step
t'	:	index for a time step

Sets:

$\mathcal{A}$:   set of links
$\mathcal{A}^0$:   set of normal links
$\mathcal{A}^c$:   set of controlled links
$\mathcal{A}^r$:   set of roundabout links
$\mathcal{A}^p$:   set of priority links
$\mathcal{A}_i^n$:   set of incoming links of node n
$\mathcal{A}_j^n$:   set of outgoing links of node n
$\mathcal{A}_t^{crit}$: set of critical links for time step t
$\mathcal{D}$:   set of destinations
$\mathcal{G}$:   space of feasible green times
Γ_r :   set of links belonging to route r
$\mathcal{K}$:   set of maximum conflict groups
$\mathcal{N}$:   set of nodes
O :   set of origins
$\mathcal{R}_a$:   set of routes to which link a belongs
$\mathfrak{R}$:   complete set of routes
$\mathfrak{R}^{od}$:   set of feasible routes between origin o and destination d
$\mathcal{T}$:   set of time periods
Ω :   space of feasible route flows

Control variables:

d :   delay for a movement [s]
g :   (effective) green time [s]
g' :   minimum (effective) green time [s]
$\mathbf{g}$: vector with green times for all controlled links and all time periods
λ :   ratio of green time and cycle time
q :   flow [veh/h]
s :   saturation flow [veh/h]
t^s :   start lag [s]
t^e :   end gain [s]
x :   degree of saturation
x'' :   maximum allowed degree of saturation
y :   flow ratio q/s
C :   cycle time [s]
G :   green time [s]
K :   maximum conflict group
L :   intersection lost time [s]
Y :   sum of the flow ratios for a maximum conflict group

DTA model variables:

β	:	positive parameter
γ	:	positive parameter
c_k^{rod}	:	travel costs of traffic departing during time period k from origin o to destination d using route r [s]
$\widehat{c}_k^{rod}$	:	perceived travel costs of traffic departing during time period k from origin o to destination d using route r [s]
δ_{ar}	:	link-coincidence matrix
ε	:	convergence error
ε^*	:	threshold value for the convergence error
ε_k^{rod}	:	random component of the travel costs of traffic departing during time period k from origin o to destination d using route r [s]
$\zeta^{(j)}$	:	parameter to smooth the flows for iteration j
η	:	small, positive parameter
f_k^{rod}	:	flow rate departing during time period k from origin o to destination d using route r [veh/h]
$\bar{f}_k^{rod}$	:	equilibrium flow rate departing during time period k from origin o to destination d using route r [veh/h]
$\mathbf{f}$	:	vector with route flows for all OD pairs, routes and time periods
$\bar{\mathbf{f}}$	:	vector with equilibrium route flows for all OD pairs, routes and time periods
Λ	:	matrix with elements following the standardized normal distribution $N(0,1)$
π_k^{od}	:	minimum travel costs of traffic departing during time period k from origin o to destination [s]
q_k^{od}	:	demand departing during time period k from origin o to destination d [veh/h]
ρ_k	:	contraction factor for time period k
θ	:	parameter for uncertainty in the knowledge of the travel time
ω	:	scaling factor
$\mathbf{C}$	:	link cost matrix
$\mathbf{C}^*$	:	adjusted link cost matrix
CF_k^{rod}	:	commonality (overlap) factor for route r of OD pair od and time period k
G	:	dynamic relative duality gap
L_k	:	Lipschitz constant for time period k
L_r	:	'length' of route r
L_{rs}	:	common 'length' of routes r and s
P_k^{rod}	:	probability to choose route r of OD pair od during time period k
PS_k^{rod}	:	path size overlap factor for route r of OD pair od and time period k
Z	:	total travel costs [veh.hours]

DNL model variables:

d_{at}	:	delay on link a at time step t [s]
d'_{at}	:	uniform delay on link a at time step t [s]
d''_{at}	:	overflow delay on link a at time step t [s]
d'''_{at}	:	initial queue delay on link a at time step t [s]
ϑ_a^{min}	:	minimum speed for link a [km/h]
ϑ_a^f	:	free flow speed for link a [km/h]
ϑ_a^c	:	speed at congestion for link a [km/h]
I'_{at}	:	travel time parameter for link a at time step t
I''_{at}	:	travel time parameter for link a at time step t
k_{at}	:	link dependent parameter for link a at time step t
κ_{at}	:	queue length on link a for time step t [veh]
$\overline{\kappa}_{a_i a_j t}$	:	part of queue length on link a_i related to link a_j for time step t [veh]
l_a	:	length of link a [m]
l_a^{crit}	:	critical length of link a [m]
$\overline{l}^{veh}$	:	the average length a vehicle occupies, including space between vehicles [m]
λ	:	integer for calculating the route costs obeying FIFO
$\mu_k^{na_i a_j}$	:	splitting rate for node n to distribute traffic from incoming link a_i to outgoing link a_j for time period k
p_a	:	number of lanes for link a
ρ_{p1}	:	parameter for priority capacity estimation [s]
ρ_{p2}	:	parameter for priority capacity estimation [s]
ρ_{p3}	:	parameter for priority capacity estimation [s]
ρ_{p4}	:	parameter for priority capacity estimation [s]
ρ_{r1}	:	parameter for roundabout capacity estimation [s]
ρ_{r2}	:	parameter for roundabout capacity estimation [s]
ρ_{r3}	:	parameter for roundabout capacity estimation [s]
ρ_{r4}	:	parameter for roundabout capacity estimation [s]
τ_{at}	:	travel time on link a at time step t [s]
$\widetilde{\tau}_a$	:	free flow travel time for link a [s]
$\widehat{\tau}_a$	:	extra travel time due to other traffic for link a [s]
φ_{at}	:	degree of saturation on link a at time step t [s]
s_a	:	number of time steps in which a change in inflow reaches the end of link a
s_a^p	:	p^{th} time step after a change in inflow of link a
u_{at}	:	inflow for link a at time step t [veh]
u'_{at}	:	restricted inflow for link a at time step t [veh]
u_{ct}	:	flow on the circulating lanes of a roundabout for time step t [veh/s]
u_{mt}	:	flow on major link m conflicting with priority link for time step t [veh/s]
v_{at}	:	outflow for link a at time step t [veh]

v'_{at}	:	corrected outflow for link a at time step t [veh]
$\widetilde{v}_{at}$	:	unconstrained outflow for link a at time step t [veh]
$v_{a_i a_j t}$	:	outflow for link a_i related to link a_j at time step t [veh]
$\overline{v}_{a_i a_j t}$	:	possible outflow for link a_i related to link a_j at time step t [veh]
$v'_{a_i a_j t}$	:	corrected outflow for link a_i related to link a_j at time step t [veh]
ν_{at}	:	travel time parameter for link a at time step t
w_{at}	:	difference between inflow and restricted inflow for link a at time step t [veh]
$\overline{w}_{at}$	:	difference between possible outflow $\overline{v}_{a_i a_j t}$ and outflow $v_{a_i a_j t}$ for link a at time step t [veh]
ψ_{at}	:	available space on link a at time step t [veh]
χ_{at}	:	number of vehicles on link a at time step t [veh]
z_{at}	:	parameter related to the initial queue for link a at time step t
Δ_h	:	length of the time step in hours [h]
Δ'_h	:	period of unmet demand in Δ_h [h]
Δ_s	:	length of the time step in seconds [s]
Q'_{ak}	:	capacity at the beginning of link a for time period k [veh/h]
Q''_{ak}	:	capacity at the end of link a for time period k [veh/h]
T	:	number of time steps per time period
T'	:	number of time periods
U_{nt}	:	inflow for node n at time step t [veh]
V_{nt}	:	outflow for node n at time step t [veh]
W_{nt}	:	difference between inflow and outflow for node n at time step t [veh]

Abbreviations:

AVV	:	Adviesdienst Verkeer en Vervoer
CONTRAM	:	CONtinuous TRaffic Assignment Model
DDUE	:	Deterministic Dynamic User Equilibrium
DNL	:	Dynamic Network Loading
DTA	:	Dynamic Traffic Assignment
DTM	:	Dynamic Traffic Management
DYNASMART	:	DYnamic Network Assignment Simulation Model for Advanced Road Telematics
ENETS	:	Equilibrium NETwork traffic Signal setting
FIFO	:	First In First Out
FLEXSYT	:	FLEXible traffic network Simulation studY Tool
GAOT	:	Genetic Algorithms for Optimisation Toolbox
HCM	:	Highway Capacity Manual
IBEC	:	International Benefits, Evaluation and Costs Working Group
ITS	:	Intelligent Transportation Systems
MARPLE	:	Model for Assignment and Regional PoLicy Evaluation

MPC	:	Model Predictive Control
MSA	:	Method of Successive Averages
NDP	:	Network Design Problem
OD	:	Origin-Destination
OPAC	:	Optimized Policies for Adaptive Control
PFE	:	Path Flow Estimator
PRODYN	:	PROgrammation DYNamique
RHODES	:	Real-Time Hierarchical Optimized Distributed Effective System
RMSPE	:	Root Mean Squared Percent Error
RTME	:	Regional Traffic Management Explorer
SATURN	:	Simulation and Assignment of Traffic to Urban Road Networks
SCATS	:	Sydney Coordinated Adaptive Traffic System
SCOOT	:	Split Cycle Offset Optimisation Technique
SDUE	:	Stochastic Dynamic User Equilibrium
STM	:	Sustainable Traffic Management
TCA	:	Traffic Control Architecture
TMA	:	Traffic Management Architecture
TRAIL	:	TRAnsport, Infrastructure and Logistics
TRANSYT	:	TRAffic Network StudY Tool
UTC	:	Urban Traffic Control
UTOPIA	:	Urban Traffic OPtimisation by Integrated Automation
VMS	:	Variable Message Sign

1

Introduction

One of the oldest research fields in traffic engineering is traffic control. From the early fifties of the previous century until now, traffic signal control has been the subject of numerous studies, real-life tests and operational implementations. In the past decades traffic signal control has evolved from a simple safety measure to an important dynamic traffic management tool. All this time, the question which control plan to use, has been an important one. In the beginning the focus was on the optimisation of local control plans. Later on, network approaches were developed and adopted. In almost all of the research done in the past, the current traffic situation, derived from measurements, was the basis for the development of control strategies. Only a few studies took into account the relation with behavioural aspects of road users, for instance route choice behaviour. In this thesis the behaviour of road users is the starting point. Of course, the situation in a network, at a certain time and on a certain location, is dependent on the amount of traffic, but this amount is the result of the choices travellers make: the choice of the travel mode, the choice of departure time and the choice of the route to take. The basic concept for our research is to use knowledge about the choice behaviour to improve control strategies.

The first section of this chapter describe the historical background and the context of this research (section 1.1), followed in section 1.2 by the problem description. The scope and objectives of the research are described in section 1.3. Section 1.4 lists the thesis contributions and section 1.5 gives an outline of the rest of the thesis.

1.1   Historical background

Several sources, for example Wikipedia (2002), mention the invention of the traffic signal in England in 1868. This traffic signal was manually operated by a police officer with a green and

red gas lamp and was used to enable members of parliament to cross London's busy streets. This traffic signal was granted only a short life, because it exploded less than a month after installation, wounding the police officer. After this event nothing happened in the field of traffic control for a long time. It was not until 1918 that the three-coloured traffic signal was introduced in London, operated manually. Britain's first automatic controlled traffic signals were installed on an intersection in Wolverhampton in 1927. The United States had introduced the automated signals earlier. Warren (1990) mentions that the first electrical traffic signal appeared in Salt Lake City in 1912, and the first traffic control system in Cleveland in 1914. Coordination started with six connected intersections in Salt Lake City in 1917. In the twenties of the previous century large cities such as New York were rapidly provided with traffic signals. This process was not so easy, because the meaning of the colours was different in New York from other cities. The first control equipment existed of manually operated signs on poles with the words 'STOP' and 'GO'. In 1916 the first traffic tower was erected. These towers had three lamps operated by police officers. At that time a red lamp meant that all traffic had to stop. An amber light meant cross-town traffic would have to stop, so that north and south-bound traffic could pass. And a green light would stop the traffic on north and south-bound lanes, so that cross-town traffic could proceed. In those days a driver could be fined for passing a green light! In 1924 the meaning and sequence of the colours was equalised with most other cities. Although, in Syracuse there is still a traffic signal which shows green above red, due to the Irish immigrants in that part of town (Wikipedia, 2002). In 1924 the electrification of the traffic signals was set to work. This was quite successful: in 1934 in New York City already 7,700 traffic signals were operational. In The Netherlands the first traffic signals appeared in the thirties. In 1930 the first traffic signal was installed in Eindhoven (near the Philips factory) and in 1936 the population of Amsterdam was incited to obey the newly installed traffic signals.

With the invention of the digital computer, new possibilities arose to improve existing traffic control strategies. Already in 1952 computers were used in Denver to choose the best strategy from a set of pre-defined control strategies, based on information from detectors. The first digital computer to control traffic appeared around 1959 in Toronto, a fact mentioned by Kelman (2004). With the development of computers the field of traffic control experienced a mushroom growth, because this development offered a lot of opportunities, not only in the control of the signals, but also in the research on new traffic control strategies.

The most simple local control strategy is fixed-time control: every movement gets a predefined amount of green time. Normally, the length of the green time is derived from the current traffic situation. With the computer and telematics developments and the increasing use of detection, more advanced traffic control is possible. Detection devices can count vehicles or detect vehicle presence, making actuated control possible. Actuated control uses detection to determine which movements have right of way and how long the green time will be. In The Netherlands actuated control is the most widespread form of traffic signal control. About 86% of all intersections has actuated control (Wilson, 1999). In the United States this percentage is much lower (about 50%), but increasing, because old equipment is rapidly replaced with new equipment, which is able to deal with semi (detection on side streets) or full actuated control (detection on all streets) (ITE, 1992).

With the computer also the possibilities in modelling became larger. The use of models was more accurate and faster than calculations by hand. The development of traffic models was a logical step. Noteworthy is that these models were used to design and optimise traffic control plans; first off-line, but later also on-line. TRANSYT is a well-known example of such a program (Roberston, 1969). TRANSYT can be used to coordinate traffic control in a network. It searches for all intersections in the network the control plan that optimises a certain performance index, typically delay and stops. TRANSYT is used off-line, e.g. to determine signal control plans for different periods of the day. It is also possible to use different control plans on-line. A selection from the available plans can be made manually or automatically, based on time or traffic conditions. These traffic control systems are called systems of the first generation (Gartner *et al.*, 1994). The second generation is based on another principle. These control systems calculate and implement control plans, based on actual measurements and model predictions, which minimise a certain objective function, usually a combination of delay and stops. A well-known system that uses this principle is the English SCOOT system (Hunt *et al.*, 1981). Third generation control systems are more hierarchical: they generate a control plan for the whole network, but that control plan can be adjusted locally, based on local conditions. The Italian UTOPIA, described by Di Taranto and Mauro (1989), is an example of such a system.

For decades traffic signal control was the most important traffic management measure, especially in urban areas. Since 25 years traffic management systems on motorways are deployed. In a wider context, traffic signal control and other traffic management systems are part of the Intelligent Transportation System (ITS). Using ITS, the goal is to improve the transportation system by making it more efficient and safer. Traditionally, traffic management is local: locally there is a problem and it is solved with a local traffic management measure, mostly without considering the effects on the rest of transportation system or other side effects. Also, in most cases, motorways and urban roads are operated and maintained by different road managers. In practise, these road managers are only responsible for their own part of the network and do not have the incentive to cooperate. There is also no integrated approach to traffic management. Therefore, in a network, traffic management is often incoherent, not optimised and different road managers can have different, sometimes conflicting and changing objectives. In The Netherlands this problem has been recognised and a structure for cooperation has been developed. Cooperation becomes even more urgent if one realises that a major part of the delay experienced by road users is suffered on the rural and urban roads and not on the motorway network. A structure for cooperation is given in the Dutch National Traffic Management Architecture (Rijkswaterstaat, 2001) and a stepwise method to support the cooperation is described in the Handbook Sustainable Traffic Management (Rijkwaterstaat, 2003). Both are summarised in appendix A.

1.2   Problem formulation

Traffic management and the behaviour of travellers influence each other. The two processes have different 'actors' who may have different goals. The road authority will try to achieve a network optimum and will try to manage traffic in such a way that this optimum is reached.

Tools for managing traffic are for example traffic signals, traffic information and ramp metering. The optimum can be a system optimum or a preferential treatment for certain user groups, e.g. public transport or pedestrians. Road users will search for their own optimum, e.g. the fastest, the most reliable or the cheapest way to travel from A to B, or a trade-off between those aspects.

Decisions taken by the road authority to manage traffic in such a way that it meets the objectives, have an influence on the possibilities for travellers to choose their preferred mode, route and time of departure. On the other hand, decisions taken by road users influence traffic flows and the situation in the network and therefore influence decisions of the road authority. For example: a change in traffic control may cause traffic volumes to change. If traffic control is modified such that congestion on a certain route disappears and delays on intersections, belonging to those route, decrease, traffic might be attracted from other links where congestion still exists or from links which are part of a longer route. This might have the consequence that queues, which originally disappeared, return. Delays may reappear at the original levels or worse (Van Zuylen, 2001). In those situations it is the question if there is still a net profit for the traffic system or society as a whole. The same question arises with respect to new traffic that may emerge as a consequence of shorter travel times, the so-called elastic demand or induced demand. Another example is the situation in which public transport gets priority in the traffic control plan. The delay for other road users may increase and thus force these road users to search for other routes in the network, other departure times or even other transport modes (Mordridge, 1997).

Mordridge (1997) makes a reasonable case, under certain conditions, for the idea that the improvement of the traffic condition for cars in a network with cars and public transport may cause a modal shift from public transport to the car, which at the end deteriorates the travel conditions for both modes. An example of this phenomenon is the city of Los Angeles where in the forties and fifties of the previous century electric cars were replaced by buses or where public transit lines were closed. This decline of the public transport, together with the construction of a lot of motorways, caused the car usage to grow considerably, leading to unsolvable congestion nowadays. An urban legend states that this is due to the automobile companies like General Motors, that bought public transport companies and let their affairs get in a mess (Schwarz, 1999).

The reaction of users to traffic management measures has been included in many models that simulate the dynamics of traffic flows in networks. The shift of traffic to other routes as a response to travel time changes and traffic information is a standard feature of most dynamic traffic assignment models, some take also the shift in departure time into account. Very few also give an estimate of modal shifts, the shift in the demand-supply equilibrium and re-allocation of activities. Traditionally, this belongs to the domain of the long-term strategic models, which are based on static assignments. The optimisation of traffic management and traffic information, taking into account the response of travellers, is still less common. This thesis tries to fill this gap partially, because it will deal with traffic control and route choice only.

1.3   Scope and research objectives

The work described in this thesis is restricted to a specific part of traffic management, namely traffic control. Traffic control consists of those measures with signals that road users have to comply with. Examples of such measures are traffic signal control and ramp metering. This in contrast with traffic information measures. Road users can use the information or not. Traditional traffic control is in most cases local and reactive: it responds to certain past and/or current traffic conditions. In the design of traffic control plans, past conditions determine the off-line or on-line implementation of traffic control plans. More precisely: in this thesis integrated anticipatory traffic control is studied. Integrated control means that the network is considered to be one multi-level network, consisting of motorways and urban roads. Anticipatory control means taking into account not only the current, but also the future traffic conditions. These future traffic conditions are related to the long term behaviour of road users, such as route choice and choice of departure time, and not to the short term driving behaviour, such as choice of lane or driving speed. Traffic control should be designed and optimised to anticipate this long term behaviour. If it is assumed that a modification in traffic control changes travel behaviour, it is necessary to anticipate this change. If delays are minimised, it should be done for the traffic volumes that will be present after the introduction of the optimised traffic control and not for the traffic volumes that existed before the optimisation. If the reaction of travellers is neglected in the optimisation of traffic control, the results may even be opposite to the desired improvement. Of course, it is possible to follow an interactive approach, where after each shift in traffic volumes, the control scheme is adjusted until equilibrium has been reached, or one may use self-adjusting traffic control. However, for certain examples it can be shown that the adjustment of traffic control, followed by a shift in traffic volumes, does not necessarily lead to a system optimum. It is even possible that the system oscillates between two or more states. This arises from the fact that the system optimum is not necessarily the same as the user optimum. The system optimum is good for the network as a whole, but can be disadvantageous for a part of the travellers in the network. This also holds if more than one road authority is involved. What is beneficial for one road authority, can be disadvantageous for the other one and vice versa. Cooperation between the road authorities and coordination of traffic control should overcome this problem. Therefore, we define the goal of our research as follows:

> The objective of this research is to optimise traffic management in such a way that the traffic system is at a certain, for all road authorities acceptable and prescribed optimum, taking into account the reaction of travellers.

As said before the reaction of travellers considered here has to do with 'long term' behaviour, such as the choice of mode, departure time and route. This as a result of choice of activity and destination. It is very complex to incorporate all these choices in one optimisation model for traffic control. For details about choice behaviour the reader is referred to the work of

Van Berkum and Van der Mede (1993). Also, we expect the largest effect on route choice, because that is the choice travellers consider first if something changes. Therefore, our work considers only route choice behaviour.

The main research question then becomes: how to control traffic taking into account route choice behaviour of road users? Related questions are:

- how to incorporate route choice in the optimisation of traffic control?

- what are the benefits of this anticipatory control approach?

- what are the benefits of cooperation between road authorities?

To be able to answer these research questions, a framework for anticipatory control has been formulated. Within this framework a new method for optimising traffic control has been developed and applied to several cases. Part of the framework is a dynamic network loading and a dynamic traffic assignment model. In the next section the achievements for these three topics are explained.

1.4   Thesis contributions

The contributions of this thesis can be grouped into the following five topics:

1. Method for optimisation of traffic control.

2. Dynamic traffic assignment model.

3. Dynamic network loading model.

4. Validation of the model.

5. General contribution.

1.4.1   Method for traffic control optimisation

In this thesis a new method for the optimisation of traffic control taking route choice into account (named anticipatory control), is developed. Using game theory, it can be shown that traditional traffic control is related to the Nash game or Cournot game (Nash game with two players), in which each player reacts on the moves of other players. Anticipatory control is related to the Stackelberg game, in which one or more players can anticipate the moves of other players if they have some knowledge about how players react. The best possible control strategy is called system optimum control. System optimum control can be seen as a Monopoly game, in which everyone chooses the moves that make the system profit the most. The anticipatory control

strategy described in this thesis, uses the whole time period under consideration (e.g. morning or evening peak) for the optimisation. Although, it uses a model to predict the behaviour of road users, it is different from Model Predictive Control (MPC). In MPC, for a certain time step within the time period considered, the state of the system is predicted for a certain time horizon and control measures are taken, based on this prediction. Then the next time step is considered and the process is repeated.

1.4.2 Dynamic traffic assignment

The topic of this thesis is not dynamic traffic assignment (DTA), but DTA is used in the framework for anticipatory control. Therefore, attention is paid to three different assignment methods: deterministic, stochastic and system optimal. Contributions for this specific part are the new contraction factor for the deterministic assignment. It is shown that using a dynamic contraction factor leads to faster convergence. For the stochastic assignment the smoothing factors and the convergence criterion are two aspects for improvement. It is shown that using adjusted MSA factors leads to faster convergence. For the convergence criterion two options were investigated: one based on flow and one based on the duality gap. Both criteria gave the same equilibrium, but the flow criterion showed an even faster convergence. These findings are all based on the results for one example network. Further research should reveal if these conclusions also hold for other examples. For the system optimum assignment in combination with system optimum control an evolutionary algorithm is used to solve both problems at once.

1.4.3 Dynamic network loading

The dynamic network loading model in this thesis uses travel time functions to propagate traffic through the network. New in this model is the combination of different travel time functions for different link types (normal, signal controlled, roundabout, priority), which makes it possible to simulate real-life networks with adequate accuracy. A special feature of the model is the treatment of the so-called critical links: short links which can be traversed in less than one time step. This is different and new compared to other models. Where other models use the free flow travel time as the criterion to determine the critical links, in our model the travel time itself is used and the determination of critical links and the adjustments for these links are done for every time step. The tests show that this way of handling critical links gives good results in terms of propagation of traffic and travel times.

1.4.4 Validation of the model

Newly developed models are rarely validated. At best the working of the model is shown for theoretical cases. In this thesis the model developed was validated, using real-life data for a motorway bottleneck and a real-life, medium-sized network. The contribution of the thesis for this part is the use of a dynamic OD estimation method, based on the work by Van Zuylen and

Willumsen (1981) and Van Zuylen (1981). The results after calibration are very good. The model predicts flows and travel times in good agreement with the measured values. The conclusion of the validation is that the dynamic traffic assignment model is capable of simulating medium-sized networks with good results in terms of flows and travel times.

1.4.5   General contribution

Macroscopic dynamic traffic assignment models bridge the gap between the long-term strategic planning models and the detailed microscopic simulation models. In The Netherlands that became an important issue, due to the introduction of the Handbook Sustainable Traffic Management and the accompanying tool: the Regional Traffic Management Explorer (RTME, see appendix A). Part of the RTME is a dynamic traffic assignment model, which is used to estimate the current situation in a regional network and to determine the effects of traffic management measures in that network. The DTA model in the RTME is the model described in this thesis. It is named MARPLE, which means Model for Assignment and Regional Policy Evaluation. Due to the integration in the RTME, MARPLE has become a full-fledged DTA model with a lot of useful features. Up to now, the RTME with MARPLE has been used in more than twenty projects and it is turning into the standard application in The Netherlands to be used for sustainable traffic management and planning of road works.

1.5   Thesis outline

In this section a brief description of the contents of the other chapters in this thesis and the relation between them, is given.

Chapter 2 summarises the literature on the subject of the combination of traffic assignment and traffic control. From the literature research topics are extracted, which need further research. The thesis focusses on some of these topics. For this further research a modelling framework is needed, which is described in *chapter 3*. In this framework three parts are included: the optimisation of traffic control, dynamic traffic assignment and dynamic network loading. In chapter 3 also the control part is specified, both for local control strategies as for the new anticipatory control strategy, including strategies if more than one control type or more than one road authority is involved.

Chapter 4 deals with the dynamic traffic assignment methods within the framework. Three assignment methods can be used: deterministic dynamic user optimum, stochastic dynamic user optimum and system optimum assignment. Besides that, important aspects such as route choice and convergence are treated in this chapter.

In *chapter 5* the dynamic network loading model is specified. Typical properties such as splitting rates, the propagation of traffic through the network and route travel times are dealt with in this chapter. The model is calibrated and validated in *chapter 6*. Not only for a small stretch of motorway with a bottleneck, but also for a medium-sized real network. Part of the calibration

consist of estimating the OD matrix. For this, the OD estimation method of Van Zuylen (1981) is extended into a dynamic version.

The anticipatory control strategy is tested for a number of small networks in *chapter 7*. These networks include cases with more than one type of control and more than one road authority.

Chapter 8 contains the conclusions, the contribution to the developments in this specific modelling field and directions for further research.

2

Literature related to anticipatory control

In this chapter we give an overview and short description of the available literature related to anticipatory control, as we defined it in section 1.3. We have striven for a complete overview of the literature on the combined traffic control and assignment problem. Literature related to route choice, traffic assignment or traffic control is reviewed only if it is necessary for our work.

The available literature can be classified in several ways. One could make a distinction between the different approaches to solve the combined traffic assignment and control problem, such as the iterative approach, the global optimisation approach and the bi-level programming approach. The iterative approach solves both problems separately and uses the outcome of one problem as the input for the other problem. In this iterative process the combined problem converges to a solution. The global optimisation approach solves both problems simultaneously and aims at a control policy that optimises globally, taking a user equilibrium into account as a constraint. The bi-level programming approach, which is used to solve the global optimisation problem in the more recent literature (e.g. Maher *et al.*, 2001; Clegg *et al.*, 2001), formulates the problem as two sub-problems and solves them on two levels. The difference with the global optimisation approach is that the user equilibrium is not considered as a constraint, but as a lower-level problem, which is solved separately. In practise the difference between the three methods is small, because for all the methods the two problems have to be solved in an iterative procedure.

Another classification can be made if the terminology of game theory is used. Game theory provides a framework for modelling a decision process with multiple decision makers ('players'), their interaction and their gain or loss. For an introduction to game theory the reader is referred to Fudenberg and Tirole (1993) and Başar and Olsder (1999). In her PhD thesis Joksimovic (2007) describes the basic concepts of game theory, consisting of the definitions (players and their interests, strategies and pay-offs, rules of the game and outcomes of the game),

the different game types and game concepts. In the case of the combined traffic assignment and control problem the road manager and the road user are the players and they make the decisions that influence the traffic process. Dependent on the level of information the information and the possibility of control the road manager has, three game concepts can be distinguished:

- The Nash-Cournot game, in which the players do not have knowledge about other player's strategies.

- The Stackelberg game, in which one or more players (the leaders) have knowledge about the strategies of the other players (the followers) and thus can anticipate the response of the followers on their decisions.

- The Monopoly game, in which one player can control all decision variables and thus creates an optimum for himself.

In recent literature about game theory, another game is mentioned: the inverse Stackelberg game. In this game the leader announces how he will react on decisions by the followers. The followers can anticipate this reaction. The leader can optimise his announcement, because he knows the response of the followers. Relatively little is known about inverse Stackelberg games: the theory is still in its infancy. Many phenomena are discovered by studying specific examples, such as the bi-level optimal toll design problem (Stankova *et al.*, 2006).

In this chapter the literature, related to anticipatory control, is reviewed in a more or less chronological order, because in that way the developments in this field are revealed and possible improvements can easily be detected. Another possibility would be that we use game theory terminology for the classification of the literature. But then we can still use this chronological order, because the sections 2.1 until 2.5, describing the iterative, global optimisation and bi-level solution approaches, all deal with Nash-Cournot games. In this type of game, players do not have knowledge about each other's strategy. Only the literature discussed in section 2.6 uses other types of games, such as the Stackelberg and Monopoly game.

The review of the literature is done using different sections to discuss different accents. In these sections some references are discussed and others are just mentioned. The discussions are necessary to come to conclusions and directions for further research (including the topic of this thesis), which are described in the final sections of the chapter.

2.1   Interaction between route choice and traffic control

More than thirty years ago Allsop (1974) was one of the first to consider the interaction between route choice and traffic control. In his paper he considered two separate research area: traffic control and traffic assignment. He proposed to integrate them by using a consistent mathematical formulation for both areas. A first step in that direction is a shared network definition. For that purpose, he developed a theoretical framework in which traffic assignment variables

are considered as functions of traffic control parameters. Some assumptions were made. For example: all major intersections have signal control, there is a fixed, static demand and travel times increase strictly monotone with the flow. Also for the control part some simplifying assumptions were made, such as the use of a fixed cycle time. The interaction between traffic assignment and control was illustrated by a simple example (see section 2.7 for the example networks). In this example the method of Webster (1958) to optimise traffic control was used.

Together with Charlesworth, Allsop developed an iterative approach, using the TRANSYT program for the calculation of the signal settings and the estimation of the relationship between link travel time and traffic flow (Allsop and Charlesworth, 1977). They found that for a certain artificial, but realistic network, two different initial assignments lead to two different solutions of the combined control and assignment problem. Charlesworth (1977) extended this work with some extra runs with different cycle times and found the same results.

The iterative approach was also used by Maher and Akçelik (1975, 1977) in their work on route control. The aim of route control is to spread the traffic throughout the network to optimise a performance index. Route control can consist of traffic signal control, physical traffic measures (e.g. ban turns or introduce one-way traffic) and route signs. Without route control, a user optimum is reached, which can be quite different from the system optimum. To illustrate this, they combined an assignment procedure and the calculation of signal settings, using the flows from the assignment, in an iterative process. Two simple networks were simulated and they concluded that route control could give higher savings than signal control alone, because it moves the network equilibrium from a user to a system optimum (Akçelik and Maher, 1977).

Gartner (1977) drew the same conclusion in his article on traffic equilibrium. He discussed the results obtained by Maher and Akçelik and added his own results for another example. He used the same iterative method: optimising signal settings for the current flow patterns, reassigning traffic for the current signal settings and repeating this until no changes occur in flow patterns or signal settings. He concluded his article by describing a method to incorporate route choice in a traffic control optimising program. This could be done by using the traffic flows as decision variables in the objective function of the signal optimisation program. In this way traffic flow variables are determined simultaneously with the traffic control variables.

A group of researchers of the Massachusetts Institute of Technology extended the iterative approach with mode choice and other forms of control like ramp metering (Gershwin *et al.*, 1978). They used existing programs for signal control optimisation and assignment and combined them with a modal split module. In an example network they investigated the effects of lane restrictions on route choice, mode choice and fuel consumption. The results were extended to emission of pollutants by Gartner *et al.* (1980).

2.2   A theoretical basis

Smith (1979b) did a lot of work to develop a sound theoretical basis for the problem. First, he showed with a simple example that the method of Webster to determine traffic signal settings (in

this case only green times, not the cycle time) does not maximise the use of network capacity, if route decisions are taken into account. He also described another method that increases the use of capacity. Assuming a convex solution space and using the definition of Wardrop (1952) for user equilibrium, he proved in another article that the combined assignment and control problem has a unique and stable solution (Smith, 1979a). Smith (1980) also developed a new control method (called P_0), which maximises the use of network capacity. This control method is discussed in detail and generalised to a set of control polices in two other articles (Smith, 1981c,b). The assignment problem has an unique equilibrium, as defined by Wardrop, if the link cost function is continuous and strictly increasing. The existence and uniqueness of an equilibrium is not guaranteed for realistic networks, because intersection control can be the cause of discontinuity and non-convexity (Smith, 1981a, 1982), but still there may exist an equilibrium and it may still be unique.

That the iterative approach does not necessarily leads to a system-optimal solution, was pointed out by Dickson (1981). For a simple example and assumptions about the travel times, he showed that the iterative procedure of calculating signal settings and reassigning traffic could lead to a decline in network performance instead of an improvement. It has to be mentioned though, that he did not use a very realistic travel time function: a linear function of the flow plus a term for the intersection delay, similar to the first term of Webster's delay formula.

Heydecker (1983) also raised the question whether a heuristic or iterative approach converges to a unique and stable solution. In his article it was proven that if the cost function increases strictly monotone with the flow and only depends on the flow on the link itself, a unique solution exists. But with a number of simple examples (T-junction and signal controlled junction) it was shown that these conditions are not always met. If this is the case, a heuristic approach is needed to come to a solution. Sheffi *et al.* (1983) proposed such an approach, which converges to a point in which traffic control and route choice are consistent, but not necessarily optimal. They took the optimisation of signal timings as a starting point and motivated that the difference between the user and system equilibrium is the real problem in finding a solution. Assuming user equilibrium they formulated the problem of finding optimal signal settings as a mathematical program. They compared two heuristic solution methods that find a (local) minimum and found that they do differ in green splits, but no so much in total delay.

Marcotte (1983) described the similarities between the network design problem (NDP), in which both capacity and flow are decision variables, and the optimal traffic signal settings problem. For the NDP, formulated as a mathematical optimisation problem, he studied an exact and heuristic solution algorithm. The exact solution algorithm is hard to use, because of the large number of constraints imposed on the problem, which makes it difficult to find a solution. The heuristic algorithm implies an iterative approach between the assignment of flows and the optimisation of capacities. Using the similarity between the NDP and the optimal traffic signal settings problem, he proved that a locally optimising control policy will stabilise at a consistent solution.

2.3   Traffic control strategies

According to Smith (1985) three solution methods for the combined traffic assignment and control problem are possible: the iterative approach, the integrated approach and a generalisation of the iterative approach taking some consequences of re-routeing into account. If the assumption is made that the best control strategy is the control strategy that in combination with the assignment minimises the maximum degree of saturation, then the control strategy P_0 is the result of this approach (see also section 3.2.3). In another paper Smith showed that P_0 complies with three conditions for an equilibrium (flow, queues and control) and that using P_0 simplifies calculating a solution (Smith, 1987).

P_0 was tested and compared with other control strategies in a simulation study (Smith *et al.*, 1987). For this study the combined simulation and assignment model for local networks SATURN was used (Van Vliet, 1982; Atkins, 2007). In the paper SATURN is combined with a signal setting optimisation module. With this simulation environment they compared three control policies: delay minimisation, Webster's method and P_0. For a realistic network in the UK the results were that Webster performed best (in terms of total delay in the network) in low congestion conditions and P_0 in high congestion conditions and that P_0 gave a more stable equilibrium. The equilibrium in this case is defined as the situation in which no changes in flows and green times can improve the situation in terms of total travel time in the network. Smith and Ghali (1990b,a) also reported results from simulation studies regarding the control strategy P_0. They showed for simple networks and also for more complex ones that P_0 gives less delay than other control policies, such as Webster and Delay Minimisation.

The superiority of P_0 is questioned by Van Vuren and Van Vliet (1992). Using the analytical and iterative approach with the simulation and assignment model SATURN, they investigated (among other things) the influence of networks and cost functions on the results for a number of control policies and realistic networks. They concluded that for realistic networks the delay minimisation policy performed best in decreasing delay, although it has some theoretical drawbacks. They also recommend to do further research on topics as realistic demands, cost functions and influence of route guidance.

Cantarella *et al.* (1991) used the ENETS procedure to come to a solution of the combined assignment and signal settings problem. ENETS stands for equilibrium network traffic signal setting. In this procedure the calculation of traffic signal settings is done in two steps: in the first step the settings for local junctions are calculated and in the second step the network coordination is determined. In an iterative process, in combination with an assignment model, the process converges to an equilibrium in terms of flow. They used this procedure on small test networks with good results.

Smith and Van Vuren (1993) discussed the point that traffic engineers do not take into account long-term effects (such as route choice) of traffic control plans. Allsop was the first to point that out and to propose a solution method. But his iterative approach can lead to oscillations. Therefore, a control strategy has to meet certain conditions. Unfortunately, most control strategies do not make a good use of the existing capacity. Therefore, a capacity maximising strategy,

such as P_0, is necessary to improve traffic flows. The article describes conditions for control policies, which ensure convergence of the algorithm, a proof of convergence, a method of classifying control policies and two new control policies. Different results are found for different cost functions.

A traffic dependent control strategy was described by Gartner and Al-Malik (1996). They use Webster's formula to calculate delay for signalised intersections. A simple network was simulated and results are given. They also showed that several solutions can exist. They extended this work with coordination and proposed to use TRANSYT for the coordination task (Gartner and Al-Malik, 1997).

The influence of the type of control on route choice is the topic of the work done by Van Zuylen and Taale (2000). This paper deals with the occurrence of instabilities in small networks. The microscopic model FLEXSYT-II- is used to simulate some simple examples (for information on FLEXSYT-II-, see Taale and Middelham (1995) and Taale and Middelham (1997)). They extend this work with more examples and showed that route choice indeed depends strongly on the type of control used (Taale and Van Zuylen, 2000). Furthermore, using a simulation framework, they studied anticipatory control: control traffic in such a way that route choice behaviour of road users is taken into account. For several small networks the simulation framework was used to determine the effects of uncertainty in route information (Taale and Van Zuylen, 2002), multiple user classes (Taale and Van Zuylen, 2003b) and anticipatory control (Taale and Van Zuylen, 2003a). The results showed that anticipatory control led to the lowest network delay in most of the cases, but not all. Especially, in networks in which route choice is not an issue (symmetric networks) or under-saturated networks, other control types might be better.

2.4   Integration with other management measures

Traffic control in relation with congestion pricing is discussed by Ghali and Smith (1993). After giving the results of a study with the dynamic assignment model CONTRAM (Leonard *et al.*, 1989; Taylor, 2003), which showed the positive effects of the control strategy P_0 for a large network, they discuss ways to handle congestion pricing. The problem is that the network cost function is convex only for networks with one bottleneck. It cannot be proven for general networks. Theoretical networks are used to show that this can be a problem.

An attempt to integrate traffic control for urban streets and freeways was made by Yang and Yagar (1994). They investigated the combined assignment and control problem for corridors. Although their assumptions were unrealistic (static demand, simple cost function, no control for urban junctions and urban network has spare capacity), they found interesting results on ramp metering. It was shown that in the corridor studied the ramp metering rates were most sensitive to variations in flows and that coordination between ramps is necessary to come to good results.

The same problem formulation and solution algorithm was also used for urban networks (Yang and Yagar, 1995). The solution algorithm is the interesting part. They formulate the combined assignment and control problem as a bi-level program in which the upper level deals with the

control problem and the lower level with the assignment problem. The assignment problem was formulated as a non-linear mathematical optimisation program with extra constraints for the controlled links.

An integrated control strategy for signal control and ramp metering with more realistic assumptions was also studied by Van Zuylen and Taale (2004). They studied a common situation with a ring road and a parallel arterial. The two road types are managed by two road authorities, each with their own control strategy. Several control control strategies are possible in this situation. The results show that separate or integrated anticipatory control gives less network delay than iterative reacting to the current situation.

An attempt to describe the interactions between driver information, route choice and signal control was made by Nihan *et al.* (1995). They showed that the existence of a unique equilibrium depends on the precision in the driver's perception. If drivers have enough information then more equilibria may exists, but some of these equilibria are unstable. It was found that the network yields the lowest total intersection delay when the equilibrium is such that all traffic and hence the major part of green time is assigned to only one of the two routes. So there is a trade-off between a network with minimum total delay, but no unique equilibrium, and a network with a unique equilibrium, but with higher total delay.

Explicit inclusion of signal green times and prices charged to traverse a route (public transport fares, parking charges, etc.) in a multi-modal, elastic, equilibrium transportation model was done by Clegg *et al.* (2001). In the article an algorithm is specified which, for a fairly general objective function, continually moves current traffic flows, green times and prices within the model toward locally optimal values while taking account of users responses. The results of applying a simplified form of the algorithm to a small network model with five routes and two signal controlled intersections are given.

2.5 Further research using a Nash-Cournot game

2.5.1 Iterative approach

A model to handle the combined traffic assignment and control problem is described by Meneguzzer (1995). He paid special attention to intersection delay modelling, including priority intersections. For a realistic network of 50 intersections three items were investigated: the convergence behaviour of the model, the impact of the maximum cycle time on the equilibrium solution and the uniqueness of the solution. He came to the conclusion that there can be problems with functions for intersection delay, because the uniqueness of an equilibrium solution cannot be guaranteed. The consequence is that different initial solutions can lead to different equilibria.

Lee and Hazelton (1996) proposed a stochastic optimisation method. They argued that a global optimisation method needs unrealistic assumptions and can guarantee only local equilibria. Therefore, a stochastic method is needed to go from a local to a global optimum. In their

paper they used simulated annealing to find this global optimum. Compared to the well-known Frank-Wolfe algorithm, the simulated annealing algorithm showed faster convergence to an equilibrium solution with less total delay.

The stochastic nature of the problem was taken into account in a paper by Watling (1997). He used the iterative approach to come to a stochastic equilibrium of the combined traffic assignment and control problem. He did this by using the concept of perceived costs and assuming a probability distribution. He studied three known control policies and found that for Webster's and the delay minimisation control policy multiple solutions may exist, which are stable and can persist for a long period of time. The control policy P_0 always gives one stable solution, but performs worse from a system point of view.

Abdelfatah and Mahmassani (1998, 2000) used the simulation model DYNASMART (Jayakrishnan *et al.*, 1994) to estimate path travel times and used these travel times to solve the combined problem. They tested two solution algorithms on a part of the Fort Worth network and compared these with a system optimal assignment. The conclusion was that joint optimisation of routing and signal settings can lead to an improvement of the network performance, but this improvement is smaller when demand increases. Their work was extended with vehicle-actuated control by Abdelghany *et al.* (1999). They did a number of experiments that included three types of coordination, one information strategy (using VMS) and an incident. The results showed the benefits of path-based coordination for both types of control. Path-based coordination favoured those intersections that are part of major corridors in the network.

The iterative approach was compared with three other solution algorithms by Lee and Machemehl (1998). They used simulated annealing and a genetic algorithm to overcome the problem of local solutions. Also a local search method was implemented. Three traffic networks known from literature were simulated. They found that the iterative approach does not contain good convergence properties, but it gives the smallest total travel time for the largest network. In another comparison they used three other local search methods, a fourth network and different demand levels (Lee and Machemehl, 1999). As much as 3520 simulations were run to obtain reliable results. The iterative approach came out the best in most of the realistic cases, leading to the smallest total travel time. For simple networks the numerical local search performed best.

2.5.2   Global optimisation approach

Results to the prejudice of the iterative approach were obtained by Wong *et al.* (1999). They compared the iterative approach with a global optimisation approach. This global optimisation approach uses a kind of hill-climbing scheme as in TRANSYT and the iterative approach uses TRANSYT itself. Both approaches use SATURN for the assignment. It came out that the hill-climbing scheme performed best if the results were compared with the total travel times of the base situation. An improvement of 4% was possible, whereas the iterative approach showed a 7% increase in total travel times.

Wong *et al.* (2001) also compared the combination of SATURN and TRANSYT with the method of successive averages and the Frank-Wolfe method to solve the assignment prob-

lem. The path-based traffic assignment formulation, together with the Frank-Wolfe solution algorithm, performed best in terms of total delay and convergence. Together with the signal optimisation part of TRANSYT, improvements of 15% in the performance index have been reported (Wong and Yang, 1999).

Cipriani and Fusco (2004) used an improved projected gradient algorithm to converge to a global optimum with user equilibrium as constraint. The improvement was found for a different calculation of the step size in the algorithm. They compared their method with a normal projected gradient algorithm and global search approaches, such as a hill-climbing scheme and a simulated annealing algorithm. The new algorithm showed a better performance in terms of total delay and calculation time.

2.5.3   Bi-level approach

Chiou (1997) used the bi-level formulation of the problem, in which the optimisation of the traffic control plan is the upper level problem, while the user equilibrium assignment is the lower level problem. The upper level problem includes also the optimisation of the offsets. The problem was solved using a gradient projection method and mixed search procedures for determining green times and offsets. It showed promising results for the same network as used by Allsop and Charlesworth. The results for this network were extended and some search methods were compared in another paper (Chiou, 1998) and an article in a journal (Chiou, 1999). The solution method was improved by combining a global search heuristic with a local optimal search method. The improved solution method was tested for two networks (Chiou, 2004).

The bi-level formulation in a dynamic context was used by Chen and Hsueh (1997). In the upper level the total intersection delays are minimised. In the lower level problem, the variational inequality approach is used to formulate the dynamic route choice model for user equilibria and an analytical model to propagate traffic through the network. For the solution, an iterative algorithm is used, containing Webster's delay formula for optimising the green times and the Frank-Wolfe algorithm for traffic assignment. The example showed that traffic responsive control (optimising green times taking route choice into account) is better than fixed-time.

Smith *et al.* (1998) used the bi-level formulation of the combined control and assignment problem to study a multi-modal network. They studied the effects of a bus gate and two bypasses for cars. The solution uses the steepest descent method. Another solution method of the bi-level program was called the cone projection method (Clegg *et al.*, 1999). Rather than an exact solution direction, this method gives a rough idea of the right direction to look for a solution and this can be advantageous for finding a global instead of a local optimum.

A stochastic equilibrium assignment model can be used as the basis for the local optimisation of signal settings. This approach is formulated and solved using a number of algorithms by Cascetta *et al.* (1998a). They also showed that the initial solution had no influence on the results for the network studied. Also other cost functions did not convergence faster or slower. Two

control policies were tested: Webster's and delay-minimisation. For congested networks delay-minimisation seems better. The bi-level optimisation of signal settings was formulated and solved in another paper (Cascetta *et al.*, 1998b). Again some solution algorithms are compared and it is shown that bi-level optimisation gives better results than local optimisation.

In two articles Ceylan and Bell (2004, 2005) also used the bi-level formulation of the problem. For the solution a framework with the Path Flow Estimator (PFE), a stochastic traffic assignment model (Bell *et al.*, 1997), and TRANSYT, in combination with a genetic algorithm to optimise signal controls, was developed. For the Allsop-Charlesworth network they showed that this framework gives better results than using only PFE and TRANSYT. This is due to the better optimisation procedure with genetic algorithms.

2.5.4   Real-time approaches

Hawas (2000) described an algorithmic framework developed for integrated real-time traffic assignment and signal control. The framework deals with a set of local controllers distributed in the network, and acts as a methodology in which assignment decisions and signal settings are adjusted on-line, cooperatively with other controllers and in accordance to the existing traffic conditions. The assignment decisions are based on a set of logical rules to evaluate (sub)paths in the network. The simulation experiment showed good results.

An attempt to use another real-time approach was made by Li and Shi (2003). They use a differential game to solve the combined assignment and control problem. In a differential game there are still a leader and followers, but now the problem is solved at the start of every time interval, instead of the complete simulation period. To solve the problem in a reasonable amount of time, a simulated annealing algorithm is used (Kirkpatrick *et al.*, 1983; Černý, 1985). The framework is tested for a simple network and compared with fixed-time control.

2.6   Research using a Stackelberg or Monopoly game

Gartner *et al.* (1980) were among the first to suggest that a traffic manager and road users can have different objectives and that each one tries to optimise his own objectives. Fisk (1984) explored this idea by comparing it with game theory. Game theory provides a framework for modelling interactions between (groups of) decision makers whose individual decisions determine the outcome of the game. In the article a distinction is made between a Nash non-cooperative game and a Stackelberg game. A Nash equilibrium, as the outcome of a Nash game, can be compared with the first principle of Wardrop in route choice (Wardrop, 1952), leading to a user equilibrium in route choice. In both equilibria the assumption is made that a player or road user has no information of other players' choices and in equilibrium no player or road user can improve his situation. In a Stackelberg game one player is the leader, who knows the response of the other players (followers) to his actions. As could be expected, the heuristic approaches to solve the Stackelberg game do not necessarily converge to the optimal solution.

The game theoretical approach was explored extensively by Chen (1998) in his PhD thesis. The goal of the thesis was to find an optimal, dynamic control strategy giving a system optimum that is consistent with a dynamic user equilibrium from the dynamic assignment. His research focussed on the development of a modelling framework in which control strategies and assignment can be combined in a dynamic setting. Therefore, a set of analytical models for the combined problem was formulated and new control strategies and new solution algorithms were developed. For the specification of the combined problem game theory was used.

Dynamic traffic assignment (DTA) was considered as a Nash non-cooperative game, resulting in a user equilibrium. The combined traffic assignment and control problem was formulated as a Cournot, Stackelberg or Monopoly game. In a Cournot game there are two players and each player does his moves independently from the other player. In fact, it is a Nash game with two players, in this case the road manager and the road users. In a Monopoly game the traffic manager controls both the traffic signals and the traffic flows and thus the outcome of this game is the system optimum.

Furthermore, several algorithms to solve these problems were developed and used in an example network. The solutions algorithms are more or less the same as for the bi-level global optimisation approach. Six control policies were tested and results were obtained using a linear queuing delay function and the HCM delay function. It appears that the Stackelberg control strategy is the best choice for practical applications in terms of total travel time and assumptions made.

The work of Taale and Van Zuylen described before, also uses the game theoretical framework. Road manager and road users are considered as players in a game who react and sometimes anticipate each other's moves. If more than one road manager is involved, the game becomes more complex with more game types and solutions. An example of a three player game is given in Van Zuylen and Taale (2004).

Varia and Dhingra (2004) use genetic algorithms to solve a Monopoly game, consisting of a combined dynamic assignment and signal optimisation problem. In the optimisation of control plans, attention is paid to the phase sequence. This is considered to be an important variable to obtain better results.

2.7   Examples

In most of the references reviewed the theory is illustrated by one or more examples. Most examples from the literature are shown in figure 2.1. The arrows in the picture show directions of travel. In these examples assumptions are made for a number of aspects. In table 2.1 an overview of the assumptions on the most important aspects is given. These aspects are: demand (S: static or D: dynamic), optimisation (G: only green times within a fixed cycle or C: also cycle time; O: offsets if a network is considered), calculation of delays (using a formula, such as Webster's or the formula used in TRANSYT, or a model such as SATURN or DYNASMART) and the types of traffic control used (SC: only signal control or also RM: ramp metering or VMS: variable message sign). A distinction is made in the approaches mentioned earlier.
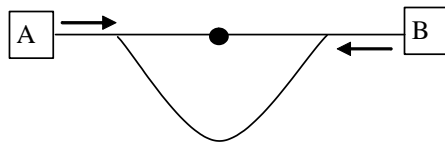
Allsop, 1974
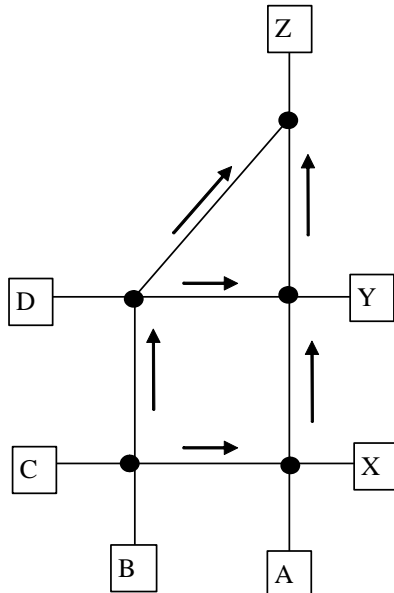
Maher and Akçelik, 1975 and 1977
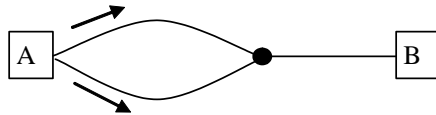
Smith, 1979, 1981
Van Vuren and Van Vliet, 1992
Nihan *et al*, 1995
Lee and Mechemehl, 1998
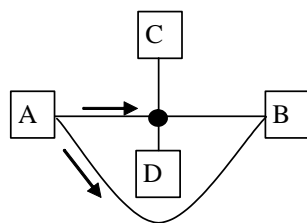
Dickson, 1981
Yang and Yagar, 1995
Taale and Van Zuylen, 2000, 2003
Maher *et al*, 2001
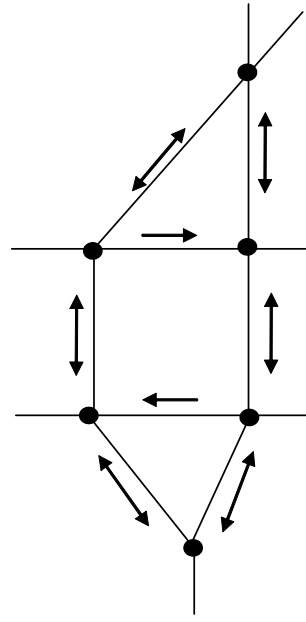
Allsop and Charlesworth, 1977
Charlesworth, 1977
Chiou, 1997, 1998 and 1999
Ceylan and Bell, 2004 and 2005
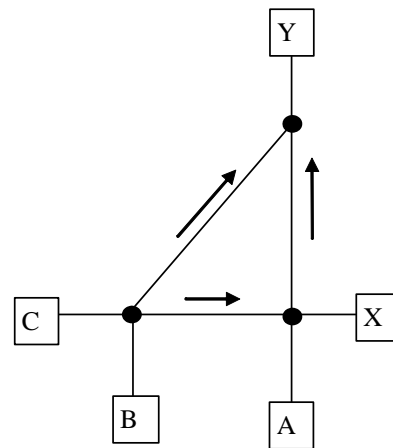
Maher and Akçelik, 1975 and 1977
Gartner, 1977
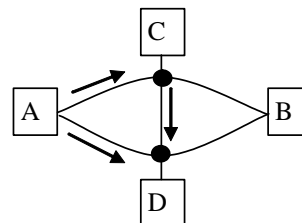
Lee and Mechemehl, 1998
Taale and Van Zuylen, 2000, 2003

Figure 2.1: Some example networks

Table 2.1: Summary of examples

Article	Demand	Optimisation	Delay/Travel time	Traffic control
Iterative approach				
Allsop (1974)	S	G, C	Webster	SC
Maher and Akçelik (1975)	S	G, C, O	Linear	SC
Allsop and Charlesworth (1977)	S	G, C, O	TRANSYT	SC
Charlesworth (1977)	S	G, C, O	TRANSYT	SC
Gartner (1977)	S	G, C, O	Webster	SC
Maher and Akçelik (1977)	S	G, C	Linear approximation	SC
Gershwin *et al.* (1978)	S	G	Webster/MITROP	SC, RM
Smith (1979b)	S	G	Webster	SC
Smith *et al.* (1987)	S	G	Kimber-Hollis	SC
Smith and Ghali (1990a)	S	G	Kimber-Hollis/Webster	SC
Cantarella *et al.* (1991)	S	G, O	Webster + platooning	SC
Van Vuren and Van Vliet (1992)	S	G	SATURN	SC
Smith and Van Vuren (1993)	S	G	BPR + Webster	SC
Meneguzzer (1995)	S	G, C	Akçelik	SC
Lee and Hazelton (1996)	S	G, C	Adjusted Webster	SC
Watling (1997)	S	G	Webster	SC
Abdelfatah and Mahmassani (1998)	D	G, C	DYNASMART	SC
Lee and Machemehl (1998)	S	G	Webster	SC
Abdelghany *et al.* (1999)	D	G, C	DYNASMART	SC, VMS
Lee and Machemehl (1999)	S	G	Webster	SC
Taale and Van Zuylen (2000)	D	G, C	FLEXSYT-II-	SC
Global optimisation approach				
Wong *et al.* (1999, 2001)	S	G	SATURN/TRANSYT	SC
Cipriani and Fusco (2004)	S	G	BPR + Webster	SC
Bi-level approach				
Yang and Yagar (1994)	S	G	Specific formulation	RM
Yang and Yagar (1995)	S	G	BPR	SC
Chiou (1997, 1998, 1999, 2004)	S	G, O	TRANSYT	SC
Chen and Hsueh (1997)	D	G	Webster/BPR	SC
Smith *et al.* (1998)	S	G	(unknown)	SC
Cascetta *et al.* (1998a,b)	S	G	Doherty/Webster/Robertson	SC
Maher *et al.* (2001)	S	G	BPR/Doherty	SC
Ceylan and Bell (2004, 2005)	S	G, C, O	TRANSYT	SC
Stackelberg game				
Chen (1998)	D	G	Linear/HCM	SC
Taale and Van Zuylen (2003a)	D	G, C	HCM	SC
Van Zuylen and Taale (2004)	D	G, C	Webster/HCM	SC, RM

From the table it can be seen that most of the time the application is restricted to examples with static demand, e.g. no demand profile is used. Also, the focus is on green time or green split optimisation. Optimisation of cycle time or off-sets is sometimes done. Integration with other traffic management measures is occasionally done, but the emphasis is on signal control, which is of course no surprise. A variety of delay models is used, from simple ones to complex simulation models. The 'ideal' combination of a dynamic demand profile, a complete signal plan optimisation, using realistic delay/travel time functions and integration of several traffic management measures, has not been found in the examples.

2.8   Concluding remarks

Although we have striven for a complete overview, it is possible that the literature discussed in the previous sections is not a complete list of all literature available on anticipatory control. But, we can state that it covers all relevant aspects. Important to note is that most articles and papers from the literature overview make some simplifying assumptions. Examples of these assumptions are a static demand profile, a fixed cycle time or simple travel time functions. Also, the research done is often restricted to traffic signal control only. Furthermore, most references deal with theoretical cases and use simple examples to illustrate the aspects described (see table 2.1 for an overview). Therefore, from the literature a number of relevant research topics for improvement can be extracted. These topics are listed and the contribution of this thesis to this topics is described.

- *Delay/travel time functions*. In most research the Webster two-term delay function is used as the cost function. In over-saturated conditions this function is not valid. So, cost functions which can deal with congestion, queues and spill back are needed. Another problem is that not much is known about delay functions for vehicle-actuated control. For example: Taale and Van Zuylen (2001) found that the Highway Capacity Manual (TRB, 1997) delay function was not in all cases suitable for Dutch traffic control circumstances. It performed well for fixed-time and simple vehicle actuated (VA) control in a simulation environment, but with complex VA control and real-life data the results were worse. The objective of this thesis is not to develop a better delay or travel time function, therefore the 'best' existing ones are used (see section 5.4).

- *Cycle time and off-set optimisation*. The cycle time is an important control parameter, but is neglected and fixed in most of the work described in literature. Cycle time optimisation should be integrated in the control problem and not only the green split optimisation. In this thesis the cycle time is a free variable, resulting from the sum of green times and lost time. Bounds are put on green times and not on the cycle time. Off-set optimisation is not studied in this thesis, because only simple intersections with two or three phase control are used.

- *Integrated networks and traffic management measures*. In the literature most of the attention has been paid to urban, signal controlled networks. Some work on integrated (urban

and inter-urban motorway) networks has been done (Yang and Yagar, 1994), but in most research the aspect of the different levels in the network and the integration of these levels has been neglected. Another aspect of integration is the integration of traffic management measures. Most of the research only considers traffic signal control, but there are few exceptions (Gershwin *et al.*, 1978; Gartner *et al.*, 1980; Ghali and Smith, 1993) and they show that the combined assignment and control problem can have great potential benefits, if control is extended to a mix of traffic management measures. In this thesis also ramp metering is considered, but it is relatively easy to integrate other traffic management measures as well.

- *User classes*. All reviewed literature considers only one travel mode: person cars. The exception is the work done by Smith *et al.* (1998). In transport systems also other modes of transport are important and relevant to traffic control, for example public transport priority at signal control junctions or waiting time for bicycles and pedestrians. The aspect of other travel modes should be thoroughly investigated. A successful attempt to include user classes in an assignment model was made by Bliemer (2001). In this work a distinction is made between the level of information of users. There are habitual users, which always choose the same route, users with a low level of information and users with a high level of information. In Taale and Van Zuylen (2003b) it is shown that the level of information has influence on the results that can be obtained with anticipatory control.

- *Travel choices*. So far, route choice is the only choice considered in all research. Of course, also other choices, such as travel mode and departure time, can be of influence on the traffic flows and thus on traffic control. This topic is not dealt with in this thesis, but is left for further research.

2.9   Summary

Three approaches to formulate and solve the combined traffic assignment and control problem are used in the literature:

- the iterative approach: using assignment and control models separately;

- the global optimisation approach: using integrated assignment and control models, where assignment is used as a condition to solve the control problem;

- the bi-level approach, also using integrated assignment and control models, but now the control problem is the upper level problem and the assignment problem is the lower level problem.

In almost all references the last two approaches are solved with a heuristic iterative solution algorithm. In terms of game theory this means that all approaches are examples of the Nash-Cournot game, in which players do not have knowledge about each other's strategies. Only

a few references use a Stackelberg game in which the leader has knowledge about the choice strategy of the followers and therefore route choice of road users is anticipated in the control problem.

The literature discussed in this chapter is limited to traffic signal control and route choice. Aspects as other traffic control measures and other travel choices are rarely taken into account. In most references a static demand is assumed and the optimisation of the traffic control is restricted to green times only. How delay is calculated varies: often a delay formula, such as Webster's or the TRANSYT formula, is used, but also simulation models, such as SATURN or DYNASMART, are employed. It must be emphasised that most delay formulas that are used, are not capable of handling over-saturated conditions and therefore lack practical relevance, because route choice in a network depends strongly on the network conditions, level of congestion and the occurrence of spill back.

A final conclusion from the literature is that relevant research topics include delay and travel time functions, cycle time optimisation, integration with other traffic management measures and integration of networks. This thesis focusses on these topics.

3

Framework for Anticipatory Control

From the literature review in the previous chapter it was clear that existing research contains several topics which need more attention. To be able to investigate these topics, a general framework is needed. This framework should include the possibility to deal with things as traffic control optimisation, traffic management measures and network types. This general framework is sketched in figure 3.1. It contains the following basic steps: after initialisation, control plans are optimised (g_{new}); the optimised control plans are used in the dynamic network loading model to determine new travel costs (c_{new}) and these are used to do a new assignment resulting in new flows (f_{new}). This process is repeated until convergence is reached.

In this chapter attention is paid to the several types of traffic control strategies related to route choice that have been developed throughout the years. In the first section the basic principles and definitions of traffic control are described. After that, the local control strategies most used in relation with route choice are defined and finally the framework for anticipatory control is specified in more detail. Details about the dynamic network loading model and the dynamic assignment model are given in subsequent chapters. Not all the topics mentioned in section 2.8 are dealt with in this thesis, but the framework, developed for local control, network control and anticipatory control, gives the possibility to extend it in the future, taken into account all the topics mentioned.

3.1   Basic principles and definitions

Traffic control is about managing conflicting traffic movements at an intersection. A movement is defined as the trip from one arm of the intersection to the other. Movements (left turn, right
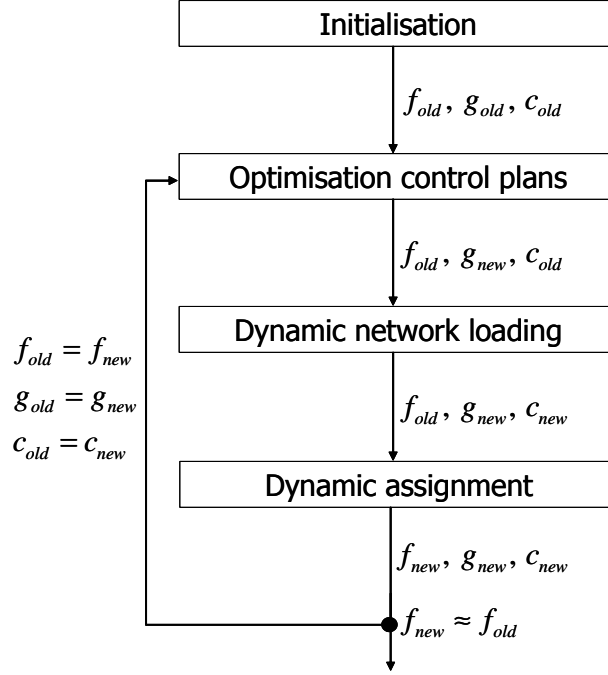
Figure 3.1: General framework

turn or straight ahead) can be controlled together in one phase or separately. It can also be done in different ways: using priority rules (give way or roundabout) or using a traffic signal. If traffic signals are used, a control strategy or control plan is needed, which defines the amount of green time for every movement or phase of the intersection and the starting point of the green time within the control cycle (off-set). Timings in traffic control plans can be fixed or dependent on the presence of vehicles or the amount of traffic approaching the intersection.

For a signalised intersection with a fixed-time control strategy, the cycle time and green times are the variables that influence the delay of vehicles and these variables form the control plan. For the control plan of a signalised intersection certain characteristics can be defined. In this section the definitions of Akçelik (1989) will be followed closely. If G_m is the green time for movement m, then the effective green time g_m can be defined as $G_m - t_m^s + t_m^e$, where t^s is the start lag and t^e the end gain. In the remainder of this thesis we assume that $t_m^s = t_m^e \; \forall m$ and this implies that $g_m = G_m$. Another useful parameter is the flow ratio. If q_m is the flow and s_m is the saturation flow for movement m, then the flow ratio y_m is defined as q_m/s_m. The flow ratio can also be considered as the fraction of the time that is needed for movement m to handle all traffic for that movement, but only if the arrival pattern of traffic has no variation. The parameter which relates the flow ratio to the green time ratio (g_m/C, where C is the cycle time) is called the degree of saturation. This degree of saturation x_m (i.e. the volume to capacity ratio) is defined as

$$x_m = \frac{q_m C}{s_m g_m} = \frac{y_m C}{g_m}. \tag{3.1}$$

For an intersection certain movements have a conflict with each other, which means that they are not allowed to receive the green light at the same time. A set of movements conflicting with

each other is called a maximum conflict group, if no other movement, which has a conflict with all other movements in the set, can be added to the set. In other words: "a maximum conflict group is a group of mutual conflicting streams (movements) that cannot be expressed as a subset of a larger conflict group" (Muller and De Leeuw, 2006). Within each maximum conflict group an internal lost time L can be determined for a certain sequence of movements. The internal lost time is the sum of the inter green times plus the sum of the start lags and minus the sum of the end gains of all movements in the set. The inter green time depends on the sequence of the movements and therefore another sequence of movements can give another internal lost time. For the calculation of the optimal cycle time the sequence of movements, within a maximum conflict group, with minimum internal lost time is used.

In designing a (fixed-time) control strategy three conditions are important:

[1] the available cycle time must be larger or equal than the time needed to handle all traffic;

[2] the green time of every movement must be longer than or equal to the minimum green time;

[3] the degree of saturation of every movement must be less than or equal to the desired maximum degree of saturation.

We consider the set $\mathcal{K}$ of maximum conflict groups of all movements of an intersection. For every maximum conflict group $K_i \in \mathcal{K}$, the internal lost time L_i and the intersection load Y_i is known, where Y_i is the sum of the flow ratios of all movements within K_i:

$$Y_i = \sum_{m \in K_i} y_m = \sum_{m \in K_i} \frac{q_m}{s_m}. \tag{3.2}$$

According to condition [1], the following must hold for the cycle time C_i

$$C_i \geq \frac{L_i}{1 - Y_i}. \tag{3.3}$$

If we distribute the available time in the cycle proportional to the flow ratio of the movements, then the green time g_m for movement m is given as

$$g_m = \frac{y_m}{Y_i}(C_i - L_i), \tag{3.4}$$

with the condition that $g_m \geq g'_m$, where g'_m is the minimum green time (in fact the minimum effective green time, which is the minimum green time plus end gain minus start lag), then condition [2] leads to

$$C_i \geq \max_m \left\{ L_i + g'_m \frac{Y_i}{y_m} \mid m \in K_i \right\}. \tag{3.5}$$

Finally, condition [3] gives

$$C_i \geq \max_m \left\{ \frac{x''_m L_i}{x''_m - Y_i} \mid m \in K_i \right\}, \tag{3.6}$$

where x_m'' is the maximum allowed degree of saturation for movement m, which is also given. Condition 3.6 holds under the assumption that $x_m'' > Y_i$ for all movements m, otherwise negative cycle times would be possible. Because condition 3.6 is the same as condition 3.3, but more strict if $x_m'' < 1$ for all $m \in K_i$ (a dream for every traffic engineer), the combination of all three conditions leads to the following contraint for the minimum cycle time

$$C_i \geq \max \left[\max_m \left\{ L_i + g_m' \frac{Y_i}{y_m} \mid m \in K_i \right\}, \max_m \left\{ \frac{x_m'' L_i}{x_m'' - Y_i} \mid m \in K_i \right\} \right]. \tag{3.7}$$

Due to this constraint, it is possible that the cycle time becomes very large. This is caused by movements with a very low flow ratio y_m. If the cycle time is bounded by a maximum, it is possible that for a certain movement the minimum green time is not met (equation 3.4). A method to correct this problem is described by Taale (1995). This analytical method does not simply increase the green time for that particular movement, but adjusts the flow rates virtually, such that all minimum and maximum bounds are met. Muller and De Leeuw (2006) describe a method to determine the best maximum conflict group, with the lowest cycle time and highest flexibility. In real life, for congested intersections, it is possible that due to an increasing Y_i the cycle time becomes too long and traffic safety can suffer from that, because drivers, cyclists and pedestrians are not likely to wait too long to cross an intersection. In that case the cycle time is maximised to a certain value, which can differ per country and intersection type.

In the next sections the control strategies, used in this thesis, are described. They are divided into local strategies, which only apply to single intersections, and the new anticipatory control strategy, which anticipates route choice in a network. Besides that, a section is devoted to network control systems. For the local strategies we do not describe actuated or adaptive control strategies, because these types of control use vehicle detection, something not included in the macroscopic traffic models, used for planning purposes. There are some developments to represent actuated controllers in a macroscopic model, but that is left for further research.

3.2 Local control strategies

In this section the local control strategies used in this thesis are discussed. They serve as a kind of benchmark to which the anticipatory strategy is related to. Anticipatory control should be better than these local strategies, in terms of delay or other network indicators.

3.2.1 Fixed-time control

With fixed-time control the green times and cycle time are fixed during the entire period. Of course, the choice of a different set of green times will lead to different results. For this work, it was chosen to use the green times, which are the 'best' for the busiest time period, which is also done in traffic engineering practise. The calculation of the green times for this period is done with the Webster formulas, which are described in the next section.

3.2.2 Webster

In the late fifties of the previous century Webster (1958) published his famous report on the optimisation of fixed-time traffic control. In his work Webster did a theoretical analysis and carried out a lot of simulations to derive a formula for the average delay due to signal control. The simulations were carried out with a simple model for arrivals and departures of vehicles. Webster used the delay formula to specify a general, optimal fixed-time control plan. He found that the general formulas for an optimal cycle time and the accompanying green times (for a two-phase controller with $Y < 0.7$) are

$$C = \frac{1.5L + 5}{1 - Y},$$
(3.8)

where C is the cycle time for the intersection (sec), L the lost time (intergreen) per cycle (sec) and Y is the sum of the flow ratios y_m for each phase m of the maximum conflict group K' chosen:

$$Y = \sum_{m \in K'} y_m = \sum_{m \in K'} \frac{q_m}{s_m},$$
(3.9)

where q_m and s_m are the flow (veh/h) and saturation flow (veh/h) for movement m. If we take the green time to be proportional to the flow rate, the green time g_m (sec) becomes:

$$g_m = \frac{y_m}{Y}(C - L).$$
(3.10)

If we replace g_m in equation 3.1 with 3.10, we obtain

$$x_m = \frac{YC}{C - L}.$$
(3.11)

The degree of saturation is independent of any movement variable, which means that the Webster control strategy equalises the degree of saturation. The parameters 1.5 (scale factor) and 5 (extra term in seconds) are based on simulations and minimise delay. Van Zuylen (1980) showed that other coefficients than 1.5 and 5 in formula 3.8 can give better results for intersections with more phases and larger Y. Nevertheless, in this thesis formulas 3.8 and 3.10 are used to calculate new cycle times and new green times for every time period and every intersection in the network for the flows entering that intersection. This means that the control plan changes due to changing route flows. In the algorithm used, minimum and maximum bounds for the green times are taken into account, according to the algorithm described in Taale (1995). This algorithm does not simply change the green times to meet the boundaries, but also changes the other green times of the movements in the maximum conflict group to maintain the balance.

3.2.3 Smith's P_0

According to Smith (1985) three solution methods for the combined traffic assignment and control problem are possible: the iterative approach, the integrated approach and a generalisation

of the iterative approach taking the control strategy into account. The control strategy P_0 is the result of this approach. Smith (1987) showed that the P_0 control strategy complies with three conditions for equilibrium (flow, queues and control) and that using P_0 simplifies calculating a solution. It has capacity maximising properties, because it tries not to equalise the delays for every conflicting movement, but the product of delay and saturation flow. So, for a simple intersection with two conflicting movements 1 and 2, P_0 results in

$$d_1 s_1 = d_2 s_2, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.12)$$

where d_m is the delay for movement m. For more complex intersections, this can be generalised into

$$d_m s_m = d_{m'} s_{m'}, \quad \forall m, m' \in K', \ m \neq m', \qquad\qquad\qquad\qquad (3.13)$$

where K' is the chosen maximum conflict group, as defined in section 3.1. The P_0 control strategy favours movements with higher capacity and thus attracts traffic to higher capacity roads.

In this thesis the P_0 control strategy, described by equation 3.13, is implemented as a minimisation problem. For every intersection and every time period the product of delay and saturation flow, for all movements in the maximum conflict group, is equalised and minimised. The saturation flow is known and the delay is estimated with the HCM 2000 delay formulas, which are described in section 5.4.2.

3.2.4   Local optimisation with genetic algorithms

Genetic algorithms (GA) are part of the larger family of evolutionary algorithms. According to Bäck (1996), evolutionary algorithms mimic the process of natural evolution, the driving process for the emergence of complex and well adapted organic structures, by applying variation and selection operators to a set of candidate solutions (population) for a given optimisation problem. To optimise, it must be known what 'optimal' is. Furthermore, the boundaries of the solution space must be known. An evolutionary algorithm starts with a certain population of individuals, sometimes random, sometimes based on knowledge of the problem. Every individual has a 'fitness', calculated with the fitness function. Using recombination techniques and mutations new individuals are created. A selection mechanism selects the individuals with the best fit for the new population. This process (named 'generation') is repeated until a stop criterion is met, which can be: distance to the real optimum, number of generations, number of function evaluations, etc. Evolutionary algorithms can be classified into genetic algorithms, evolutionary programming, evolution strategies and genetic programming. These types of evolutionary algorithms differ in the way the individuals are represented and the operators used. More details can be found in Bäck (1996), Eiben (2002) and Eiben and Smith (2003).

In the past some simulation experiments with genetic algorithms have been carried out to optimise signal control plans. See for example Foy *et al.* (1992), Hadi and Wallace (1993), Hadi

and Wallace (194), Montana and Czerwinski (1996) and Clement and Anderson (1997). From these experiments it can be concluded that using genetic algorithms to find optimal timing plans for intersections, which adapt to the actual traffic situation, is a promising idea. For the Dutch situation, with a lot of local, advanced vehicle actuated traffic signal control, the use of evolutionary algorithms to adapt the maximum green times (one of the important parameters of vehicle actuated control) related to the changing conditions, was also studied by Taale *et al.* (1998), Taale (2000) and Taale (2002), with fairly good results.

For this thesis a real valued genetic algorithm was used, implemented as a MATLAB$^{\circledR}$ toolbox by Houck *et al.* (1995), which is named the Genetic Algorithms for Optimisation Toolbox (GAOT). First, initial, random solutions are defined in a certain domain. Then, by means of random mutations within these solutions, random combinations of solutions and selection of the best solutions, a pre-defined performance (fitness) function is optimised. For the performance of solutions delay is chosen as the indicator. In order to find the best control plan for the traffic signals at each intersection separately, the total delay is minimised. For the current traffic conditions, the delay is estimated for each candidate control plan with the HCM 2000 delay formulas (see section 5.4.2). Genetic algorithms do not guarantee an optimal solution, but they will approach it fairly close, dependent on the number of generations and the size of the population. An advantage of genetic algorithms is the larger possibility than for standard optimisation methods that local optima are avoided; a disadvantage is the calculation time needed.

3.2.5 Ramp metering

For ramp metering a capacity-demand feed-forward algorithm is used. From the downstream capacity on the motorway $s_{downstream}$ and upstream flow on the motorway $q_{upstream}$ the number of vehicles allowed to enter the motorway is defined as the difference between these two values. Using a fixed cycle time C, the green time of the metering signal can be calculated with

$$g = \frac{s_{downstream} - q_{upstream}}{s_{on-ramp}} C, \tag{3.14}$$

where $s_{on-ramp}$ is the saturation flow of the on-ramp. By varying the green time between a minimum value and the cycle time, the on-ramp can be metered. If the green time is equal to the cycle time, no metering is applied.

3.2.6 Framework for local control

To be able to study the local control strategies in combination with traffic assignment, an integrated framework was developed (see figure 3.2). After initialisation the route flows, green times and route travel times are known (f_{old}, g_{old} and c_{old} respectively). The first step is to optimise the control plans with one of the local strategies described in the previous sections, resulting in new green times g_{new}. After that a dynamic network loading (DNL) is performed, with the existing route flows f_{old} and new control plans g_{new} as input. The DNL is a simulation
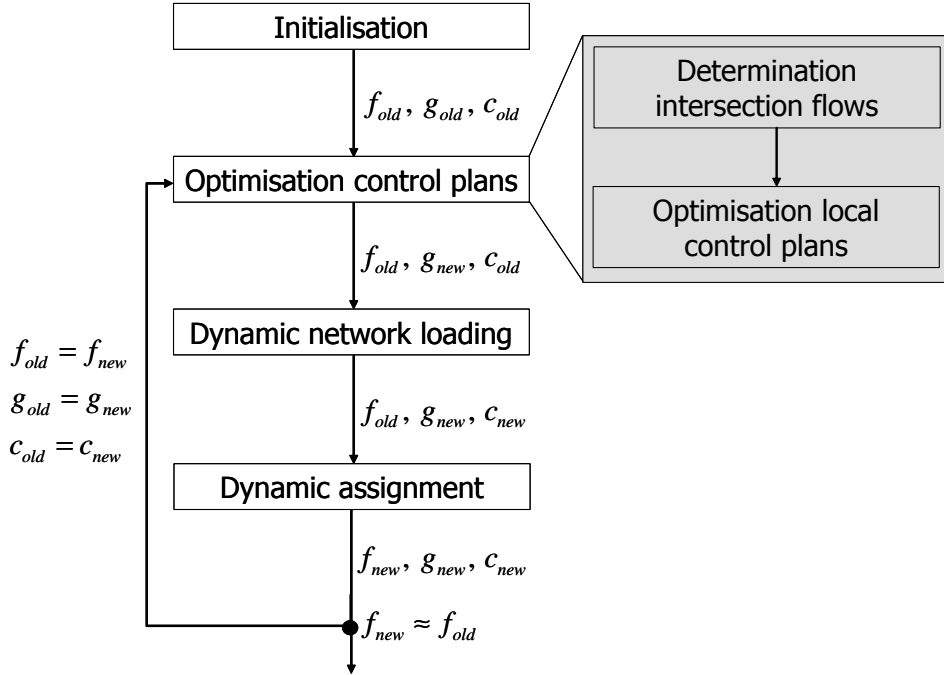
Figure 3.2: Framework for local control

of the traffic situation to determine link flows and speeds and route travel times and derive from that new route travel costs c_{new}. Then new route flows f_{new} are determined with a dynamic traffic assignment (DTA), based on the calculated route costs c_{new}.

The dynamic network loading model and the dynamic assignment model are described in chapter 5 and 4 respectively. Before that, the framework for anticipatory control is described in the next section.

3.3   Network control systems

The control strategies discussed in the previous sections are all local strategies: for one intersection only. In the past also a number of network control strategies were developed. For specification of fixed-time control plans TRANSYT (Roberston, 1969) is an example of a tool, well-known and widespread. More real time and adaptive examples of these network strategies are SCATS (Lowrie, 1982) and SCOOT (Hunt *et al.*, 1981), nowadays used around the world (www.scoot-utc.com). Research during the last decades on this topic led to other systems such as OPAC (Gartner, 1983), the French system PRODYN (Henry and Farges, 1989), the Italian UTOPIA, which is tested and used in some cities in Europe and is described in Di Taranto and Mauro (1989), and RHODES, an American development by Head *et al.* (1992). SCOOT and SCATS are used around the world, but the other systems are only implemented in some cities for testing and are still in their development stage. An extended overview of local and network control strategies is given by Van Katwijk (2008). In his thesis he also develops a new adaptive control strategy based on a multi-agent approach.

3.4   Anticipatory control

All these systems have in common that they react on the traffic flows in the network. Current traffic flows are detected and, based on various algorithms, signal control plans are activated. Sometimes predictions are made, but generally speaking these predictions are related to the arrival pattern at the halt line. This research focusses on predictions on a more higher level: the traffic flows in a network itself. Is it possible to anticipate the expected traffic flows? To be able to do that, game theory is used (see also section 2.6). A good introduction to game theory is provided by Fudenberg and Tirole (1993) and Başar and Olsder (1999).

3.4.1   Game theoretical approach

The traffic control engineer is considered to be one player and the road users are the other players. If it is assumed that all road users have the same choice behaviour, they can be considered as one player. If we assume that only route choice is relevant, then we have two players, each with one set of decision variables. The traffic engineer has the signal settings and the road users have route choice. A Nash game (sometimes also called Cournot game if only two players are involved) is played when both players react on each other's moves: the traffic engineer sets the signal control plans, the road users travel and select routes based on their individual preferences and the experienced travel times. This game results in the Nash equilibrium, which is defined as the situation in which no player can benefit by changing his strategy, while the other players keep theirs unchanged (similar to Wardrop's definition of equilibrium). In every iteration the road authority tries to minimise the total travel costs Z defined by

$$\min_{\mathbf{g}} Z_{\mathbf{g}} = \sum_{o,d}\sum_{r}\sum_{k} c_k^{rod}(\mathbf{g},\mathbf{f}) f_k^{rod}, \ \ \mathbf{g} \in \mathcal{G}, \ \ \mathbf{f} \in \Omega, \ \ o \in O, \ \ d \in \mathcal{D}, \ \ r \in \mathfrak{R}^{od}, \ \ k \in \mathcal{T}. \ \ (3.15)$$

Whereas the road users strive for an equilibrium, which can be defined by

$$\sum_{o,d}\sum_{r}\sum_{k} c_k^{rod}(\bar{\mathbf{f}})(f_k^{rod} - \bar{f}_k^{rod}) \geq 0, \ \ \forall \mathbf{f} \in \Omega, \ \ o \in O, \ \ d \in \mathcal{D}, \ \ r \in \mathfrak{R}^{od}, \ \ k \in \mathcal{T}, \ \ \ \ \ (3.16)$$

where

- $\mathbf{g}$ is the vector with green times for all controlled links and all time periods k, the cycle time is not a variable, because for every intersection it is the result of the sum of the green times and the lost time and for a metered link the cycle time is fixed;

- $\mathbf{f}$ is the vector with route flows for all OD pairs, routes and time periods;

- $\bar{\mathbf{f}}$ is the vector with equilibrium route flows for all OD pairs, routes and time periods;

- o is an origin;

- d a destination;

- r a possible route;

- k is the departure time interval;

- f_k^{rod} the route flow between o and d for route r departing during time interval k;

- c_k^{rod} the travel costs for the route flow between o and d for route r departing during time interval k, including delay caused by congestion or blocking back;

- O is the set of origins;

- $\mathcal{D}$ is the set of destinations;

- $\mathfrak{R}^{od}$ the set of feasible routes between o and d;

- $\mathcal{T}$ is the set of time periods;

- $\mathcal{G}$ the space of feasible green times;

- Ω the space of feasible route flows.

The framework sketched in figure 3.2, together with the local control strategies from section 3.2 result in this equilibrium.

In a Nash game every player reacts on the moves of the other player(s). If one of the players has information on the reaction of the other player(s), he can anticipate the moves of the other player(s) and benefit from this knowledge. For the traffic control engineer, one way to anticipate is to predict choice behaviour and to use this prediction to adjust the signal control plans. In this way, he anticipates the reaction of the road users to his 'move'. In this thesis choice behaviour is restricted to route choice, but can be extended with departure time choice or other choices. In game theory this anticipation is called a Stackelberg game. The traffic engineer sets the signals and takes route choice into account. The road users choose their route based on experienced travel costs, normally represented by travel times. Based on the current and anticipated flows, the traffic engineer changes the signal plans and again the road users react to this by changing route and this procedures repeats until an equilibrium is reached.

To anticipate on route choice the traffic engineer has to predict it in the optimisation procedure. To be able to do that, the framework of figure 3.1 is extended to the framework shown in figure 3.3. In the optimisation of control plans four steps are needed: generate a certain control plan by whatever method, run a simulation with a dynamic network loading model to see how traffic propagates through the network with this control plan, based on these results run a dynamic traffic assignment to obtain a new route flow distribution and again run the dynamic network loading model to come to a final evaluation of the control plan.
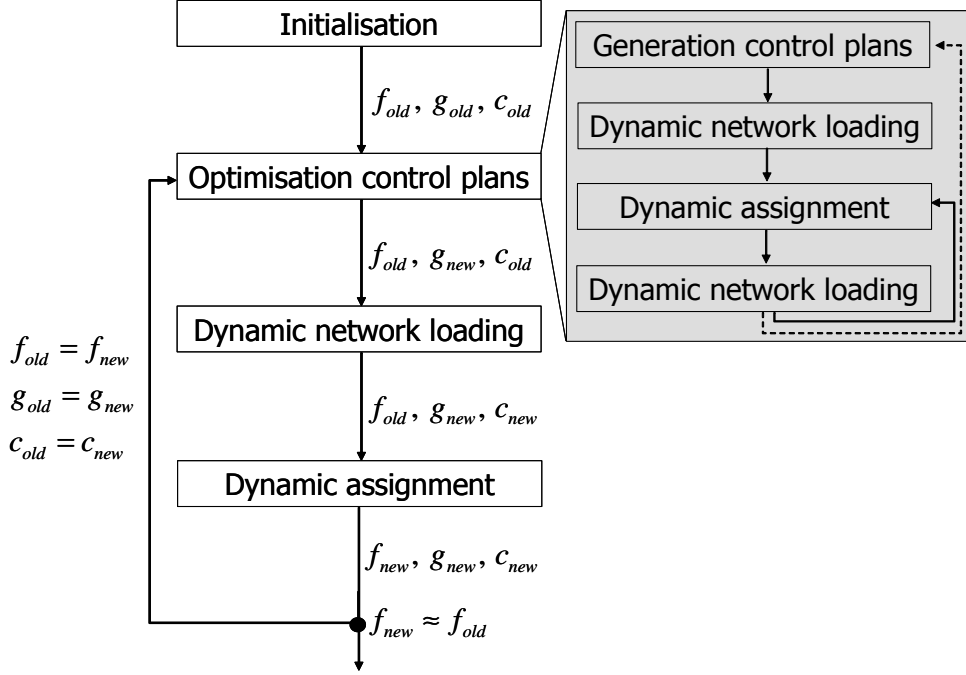
Figure 3.3: Framework for anticipatory control

3.4.2 Bi-level formulation

The complete framework represents a bi-level optimisation problem. In the upper level problem the traffic manager tries to minimise the total travel costs Z defined by

$$\min_{\mathbf{g}} Z_{\mathbf{g}} = \sum_{o,d}\sum_{r}\sum_{k} c_k^{rod}(\mathbf{g},\bar{\mathbf{f}})\overline{f}_k^{rod}, \quad \mathbf{g} \in \mathcal{G}, \ \bar{\mathbf{f}} \in \Omega, \ o \in O, \ d \in \mathcal{D}, \ r \in \mathfrak{R}^{od}, \ k \in \mathcal{T}. \quad (3.17)$$

subject to

$$\sum_{o,d}\sum_{r}\sum_{k} c_k^{rod}(\bar{\mathbf{f}})(f_k^{rod} - \overline{f}_k^{rod}) \geq 0, \quad \forall \mathbf{f} \in \Omega, \ o \in O, \ d \in \mathcal{D}, \ r \in \mathfrak{R}^{od}, \ k \in \mathcal{T}, \quad (3.18)$$

where all variables have the same meaning as in equation 3.15. The bar above the route flows $\mathbf{f}$ means that the route flows are in equilibrium, which is the solution to the lower level dynamic traffic assignment problem. So the road manager performs his optimisation for network flows that are constrained by the requirements of the user equilibrium. There are various ways to specify the lower level problem: for example as an optimisation problem or using the variational inequality approach. The complete variational inequality formulation is given in chapter 4.

3.4.3 Solution method

For the optimisation problem in the upper level (equation 3.17), a solution method has been sought. A large amount of literature on local and global optimisation methods exists. For an

overview, the reader is referred to Bliek *et al.* (2001) and Neumaier (2004). Numerous analytical and heuristic methods have been tested. Due to the nature of the problem, the number of variables to optimise and the fact that a function evaluation consists of a combined DNL, DTA and DNL run, an analytical approach would become very complex and is therefore not very suitable. That is why a heuristic approach is chosen, which uses as less function evaluations as possible. A workable method is the evolution strategy (ES) with covariance matrix adaptation (CMA-ES), as described by Hansen and Ostermeier (1996), Hansen and Ostermeier (2001), Hansen *et al.* (2003) and Hansen (2006). Evolution strategies belong to the larger family of evolutionary algorithms, just like genetic algorithms, and primarily use mutation and selection as operators (see Bäck (1996) and also section 3.2.4). In a (μ, λ)-ES, a population of λ individuals are mutated from μ parents. From the $\mu + \lambda$ individuals, the best μ individuals are selected to be the parents for the next generation. In our case, an individual is a feasible combination of green times of all intersections and time periods (a control plan). Every individual is evaluated using traffic assignment and simulation as shown in figure 3.3.

Using this heuristic method can take a lot of CPU time, especially if within the optimisation the assignment process iterates towards equilibrium. To decrease CPU time, the number of iterations of the assignment within the evaluation process of the control plans can be limited. In the next section this assumption is tested.

3.4.4   Predicting ahead

In figure 3.3 the optimisation of control plans contains two loops. The dashed line represents the generation of a new control plan based on the outcome of the assignment loop. The continuous line represents the iterative assignment process for a particular control plan within the optimisation loop. This can be considered as predicting one or more days ahead in a day-to-day route choice process. The number of iterations in this loop represent the number of days ahead for which a prediction is made to optimise the control plan. In the ideal case this loop should iterate until equilibrium is reached, because we want to determine the best control plan for the equilibrium situation. For computational reasons this could not be feasible and maybe also not necessary. To test this assumption, we use the networks of figure 3.4.
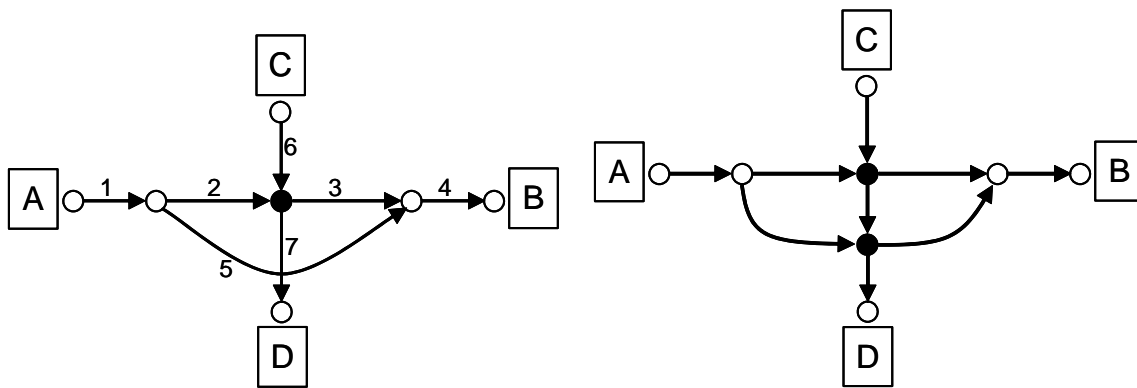
Figure 3.4: Two networks to test prediction quality in optimisation

More details of these networks can be found in section 7.1 and section 7.2. For this two simple cases two scenarios were run: one scenario for which in the optimisation process only one iteration of the assignment is calculated and a scenario for which in the optimisation process the assignment is iterated until equilibrium is reached. The results for the total delay in the network and the CPU time is shown in figure 3.5. The graph shows that, for these networks, using an equilibrium assignment in the optimisation process is somewhat better for the final result (7% and 3% less delay), but requires far more calculation time (8 times and 5 times more). Based on these results, we have chosen to use one iteration in the optimisation process for the simulations in this thesis.
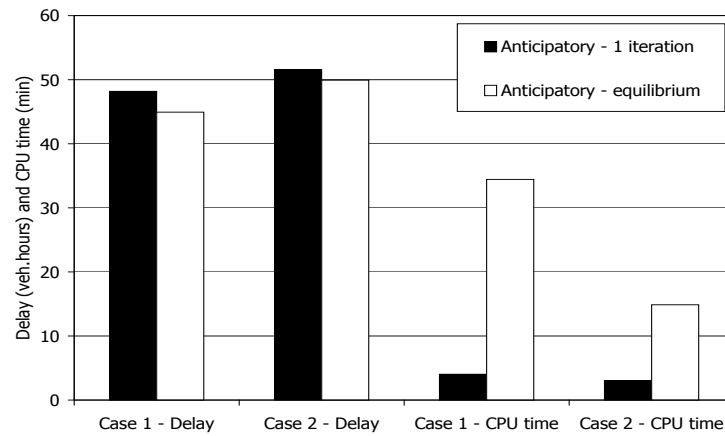
Figure 3.5: Total delay and CPU time for two cases

3.4.5   Stochastic variation

An evolutionary algorithm is a stochastic process. Dependent on the random initialisation or the random seed, different solutions can be found. To test this and to compare the CMA-ES with the genetic algorithm GAOT, the cases of figure 3.4 have been used. Details of these networks can be found in chapter 7. The results for 10 random starts concerning the total delay in equilibrium and the total number of function evaluations to reach that solution, are shown in table 3.1. Compared with genetic algorithms, the CMA-ES finds a solution with less function evaluations and CPU time and the average solution is a little bit better than for the GAOT algorithm, with a smaller standard deviation.

3.5   System optimum control

The system optimum control strategy is not really a practical one, but it is a kind of benchmark, useful to compare with other control strategies, because it represents the best that can be achieved if the traffic manager has total control, also on the route choice of the road users.

Table 3.1: Comparison of GAOT and CMA-ES

	Case 1				Case 2			
	Total delay		# func evals		Total delay		# func evals	
Run	GAOT	CMA-ES	GAOT	CMA-ES	GAOT	CMA-ES	GAOT	CMA-ES
1	47.94	47.73	4496	1428	115.02	106.34	4999	1430
2	49.04	47.70	4927	1632	123.41	104.78	5293	1760
3	48.35	46.36	3820	1632	105.94	114.38	5433	2090
4	47.44	47.16	3563	1632	104.81	107.09	5023	1760
5	46.40	46.46	4245	1530	107.84	103.30	6143	1210
6	50.38	50.65	3314	1734	105.67	104.96	5479	1540
7	48.48	48.19	4026	1326	104.51	105.56	5257	1650
8	49.27	48.77	4265	1530	106.22	104.37	6288	1430
9	45.71	47.54	4737	1632	117.72	103.84	4345	1650
10	49.51	46.52	4514	1734	110.78	103.65	5030	1430
Avg.	48.25	47.71	4191	1581	110.19	105.83	5329	1595
SD	1.36	1.23	487	123	6.13	3.07	537	232

In some cases the road manager is able to impose route choice to the road users by regulations (one-way streets and prohibition of turning movements). The system optimum strategy can give guidance for those situations. In game theory the system optimum strategy is called a monopoly game. The optimisation problem is defined as

$$\min_{\mathbf{g},\mathbf{f}} Z = \sum_k \sum_{o,d} \sum_r f_k^{rod} c_k^{rod}(\mathbf{g},\mathbf{f}), \ \ \mathbf{g} \in \mathcal{G}, \ \mathbf{f} \in \Omega, \ o \in O, \ d \in \mathcal{D}, \ r \in \mathfrak{R}^{od}, \ k \in \mathcal{T}, \ \ (3.19)$$

where again all variables have the same meaning as for equation 3.15. To solve this problem analytically (for example using marginal costs) is difficult (if not impossible) due to the non-linearities involved (Chen, 1998). Therefore, the evolution strategy with covariance matrix adaptation is used also to solve this minimisation problem (see section 3.4.3). In this case a member of the population is a vector with route flows (not necessarily in equilibrium) and green times for every time period. Again, ten simulations are run and the minimum value for the total delay is used.

3.6 Extension to three players

Until now we have assumed there are two players in the game of traffic control and assignment: the road authority and the road user who want to optimise their own objectives. This can easily be extended to a game with three players: for example two road authorities and the road user. If we assume that one player (the leader) knows how the other players (the followers) will respond to any decision he may make, we have a game that is known as a Stackelberg game. If, on the contrary, we assume that the players do not know each other's strategy, they will each optimize their own objective function assuming that the other player remains unchanged. The equilibrium that can be achieved in that situation is that no player can improve his objective

function by changing his own decision without cooperation of the other player. This equilibrium is called the Nash equilibrium.

In the case of the Nash equilibrium, when the road authorities have no a-priori knowledge of each other objectives, we can see the iterative equilibrium as follows:

[1] forced by a certain traffic situation, the first road authority takes some traffic measure

[2] the road users react and the effects in terms of flows on his infrastructure is observed by the second road authority

[3] the second road authority takes measures to optimize the traffic conditions on his roads

[4] road users react again and after a shift of routes a new user equilibrium emerges

[5] the first road authority observes the shifted flows and adjusts his control measures etc.

The reality might be a little more complicated, because the control measures taken by a road authority are often based on the present traffic conditions. After the realisation of the measures traffic flows may change in the whole network, which makes an adjustment of the control measures necessary on both levels of infrastructure. However, as shown e.g. by Taale and Van Zuylen (2003a), it is possible to optimize control measures taking in advance already into account the reaction of the road users. It is not unlikely that, if control measures are taken without taking into account the reaction of the road users, that the state that emerges after the adaptation process is worse than the initial state before the optimization.

In the situation of an urban road infrastructure with two authorities with their own objectives and drivers that optimize their own routes, the following strategies can be distinguished:

[1] The motorway and municipal authorities optimize their objectives independently, the drivers react and both authorities adjust their strategies if necessary, until an equilibrium is reached. This is an example of the Nash equilibrium.

[2] The motorway authority optimizes his objectives, while he anticipates the reaction of the drivers. The municipal road authority does the same. They do not know each other's strategy.

[3] The motorway authority knows the reaction of the drivers and the municipal road authority and optimizes his strategy, anticipating their response. This is a Stackelberg game with the motorway authority as leader.

[4] Similar to 3, but now with the municipal road administrator as leader who optimizes his control anticipating the reaction of drivers and the motorway authority.

[5] Both road administrators cooperate and define an objective for the whole road infrastructure, e.g. the minimization of the total delay. They optimize their traffic control while they anticipate the reaction of the drivers.

[6] Based on the shared objective for the whole network, the drivers are forced to take the route that optimizes the network.

Of course, there are some alternatives in between, apart from these 6 alternatives. Most interesting is, however, the question whether cooperation brings much advantage compared to single actor optimization and whether the Stackelberg solution is much better than the Nash equilibrium. There is a lot of discussion in several countries about the need to have a single road authority responsible for the traffic in a region or that the existing road administrators can do their work individually, with the assumption that coordination emerges automatically due to the interaction between the different sub-networks. The framework described in section 3.4 will also be used to investigate this problem. This is realised by dividing the network into sub-networks and using the method for one or all of the sub-networks, dependent on the strategy. But first the modules in the framework will be described in more detail, the assignment methods in chapter 4 and the dynamic network loading module in chapter 5.

3.7   Summary

In this chapter we have described a framework for anticipatory control, using the concepts of game theory. First, some basic control principles and definitions were given and some well-known local control strategies have been defined. For anticipatory control it is assumed that the road authority knows the reaction of the road users and anticipates this reaction, using his control possibilities. This anticipation consists of predicting the route choice behaviour. Ideally, this would be the route choice behaviour in equilibrium, but to decrease calculation time, a prediction of route choice for the next day is used.

In this thesis a new method for the optimisation of traffic control, taking route choice into account, is developed. Using game theory, it can be shown that traditional traffic control is related to the Nash game or Cournot game, in which each player reacts on the moves of other players. Anticipatory control is related to the Stackelberg game, in which one or more players can anticipate the moves of other players if they have some knowledge about how players react. System optimum control can be seen as a Monopoly game, in which everyone chooses the moves that make the system profit the most. New in this work is the use of a special type genetic algorithm to solve the optimisation problem. It is shown that for the cases studied this algorithm (the evolution strategy with covariance matrix adoption) performs better than the standard genetic algorithm. Also new is that the framework is developed in such a way that extension to more players (e.g. more road authorities with their own objectives) is possible.

4

Dynamic Assignment Methods

In the framework described in the previous chapter, traffic assignment plays an important part. Traffic assignment is concerned with the distribution of the demand among the available routes for every origin-destination pair. It is needed for the main loop of the framework, in which control strategies are evaluated, but also for anticipatory control, in which assignment is used to predict route choice to determine good traffic control plans. Traffic assignment is an important tool in current traffic research. Starting with the original four-step transport models, traffic assignment models have evolved into useful tools to study the effects of all kinds of planning and traffic management problems, also in a dynamic context. There is a large amount of literature concerning the development of traffic assignment models. For an overview of the relevant literature, the reader is referred to Chen (1999) and Bliemer (2001). In the sections below the methods used in this research are described. Much of it is also described in the references mentioned before. According to the classification of Chen (1999), a deterministic dynamic user optimal assignment, a stochastic dynamic user optimal assignment and a system optimum assignment are used. Dynamic, because traffic is a dynamic process and therefore travel demand changes during the time of day. In this thesis it is assumed there is no shift from travel demand from one time period to the other, so departure time choice is not included. Furthermore, the demand is inelastic, which means that the total demand for the whole assignment period does not change.

4.1 Deterministic dynamic user optimal assignment

A deterministic dynamic user optimal (assignment assumes all travellers have perfect knowledge of the traffic conditions they encounter during their trip. Therefore, it leads to the deter-

ministic dynamic user equilibrium (DDUE), which is defined in definition 4.1. In this definition discrete time intervals are used instead of continuous time. The choice for discrete time intervals is made to avoid the complexities involved solving the continuous time model.

Definition 4.1 *(DDUE) For each origin-destination (OD) pair, the route travel costs for all users travelling between a specific OD pair and departing during a specific time interval are equal, and less than (or equal to) the route travel costs which would be experienced by a single user on any unused feasible route for that time interval.*

We consider a network $\mathcal{M} = (\mathcal{N}, \mathcal{A})$ with nodes $\mathcal{N}$ and directed links $\mathcal{A}$. Let f_k^{rod} be the flow rate taking route $r \in \mathfrak{R}^{od}$ from origin $o \in O \subseteq \mathcal{N}$ to destination $d \in \mathcal{D} \subseteq \mathcal{N}$, departing during time interval k, where $\mathfrak{R}^{od}$ is the set of feasible routes between o and d. Let c_k^{rod} be the travel costs experienced by the travellers for this route and departure time interval and q_k^{od} the demand for this OD pair and departure time interval, then equilibrium definition 4.1 can be written as

$$\left(f_k^{rod} > 0 \implies c_k^{rod} = \pi_k^{od} \right), \quad \forall o, d, r \in \mathfrak{R}^{od}, k, \tag{4.1}$$

where

$$\pi_k^{od} \equiv \min_{r \in \mathfrak{R}^{od}} c_k^{rod}, \quad \forall o, d, k, \tag{4.2}$$

$$\sum_{r \in \mathfrak{R}^{od}} f_k^{rod} = q_k^{od}, \quad \forall o, d, r \in \mathfrak{R}^{od}, k. \tag{4.3}$$

To solve the deterministic dynamic user optimal assignment, such that condition 4.1 is satisfied, a variational inequality (VI) problem formulation can be used. The VI approach is chosen, because the equilibrium conditions can easily be transformed into a VI problem and the VI problem formulation allows for asymmetrical interactions in the cost functions. A route-based formulation for the discrete-time case (time is divided into periods) is as follows: find an $\bar{\mathbf{f}} \in \Omega$ such that

$$\sum_{o,d} \sum_{r \in \mathfrak{R}^{od}} \sum_k c_k^{rod}(\bar{\mathbf{f}})(f_k^{rod} - \overline{f}_k^{rod}) \geq 0, \quad \forall \mathbf{f} \in \Omega, \tag{4.4}$$

where Ω is defined as the set of all route flow patterns $\mathbf{f}$ satisfying the following constraints:

$$\sum_{r \in \mathfrak{R}^{od}} f_k^{rod} = q_k^{od}, \quad \forall o, d, k, \tag{4.5}$$

$$f_k^{rod} \geq 0, \quad \forall o, d, r \in \mathfrak{R}^{od}, k. \tag{4.6}$$

4.1.1 Solution algorithm

The general VI problem can be defined as follows: find an $\mathbf{x} \in X$ such that

$$\langle F(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \forall \mathbf{y} \in X \tag{4.7}$$

where $X \subset \mathbb{R}^n$, F is a mapping: $X \rightarrow \mathbb{R}^n$ and $\langle ., . \rangle$ is the standard inner product on $\mathbb{R}^n$. A general scheme for solving VI problems is given below. More details can be found in Nagurney (1993), also concerning the relation between VI and optimization problems. Let $\mathbf{h}$ and $\mathbf{g}$ be functions on solution vectors $\mathbf{x}$ and $\mathbf{y}$, which both belong to solution space X, then the general scheme is given in algorithm 4.1 (Bliemer, 2001).

Algorithm 4.1 *General scheme for solving VI problems*
Step 1:   *Determine an initial solution $\mathbf{x}^{(0)} \in X$. Set $j := 1$.*
Step 2:   *Construct a continuous function $\mathbf{h} : X \times X \rightarrow \mathbb{R}^n$ such that*
            $\mathbf{h}(\mathbf{x}, \mathbf{x}) = \mathbf{g}(\mathbf{x})$ for all $\mathbf{x} \in X$, and
            $\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}, \mathbf{y})$ is symmetric and positive definite for all fixed $\mathbf{x}, \mathbf{y} \in X$.
Step 3:   *Find a new solution $\mathbf{x}^{(j)}$ such that $\mathbf{h}(\mathbf{x}^{(j)}, \mathbf{x}^{(j-1)})^T (\mathbf{x} - \mathbf{x}^{(j)}) \geq 0$, $\forall \mathbf{x} \in X$.*
Step 4:   *If $\left\| \mathbf{x}^{(j)} - \mathbf{x}^{(j-1)} \right\| \leq \varepsilon$ for a certain $\varepsilon > 0$, then stop.*
            Otherwise, set $j := j + 1$ and return to Step 3.

Due to the restrictions posed on $\mathbf{h}(.,.)$ in step 2, finding a new solution in step 3 leads to a strictly convex optimization problem, which can be solved using well known optimization techniques (Bazaraa *et al.*, 1993). If the function $\mathbf{h}(\mathbf{x}, \mathbf{y})$ is chosen as

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{g}(\mathbf{y}) + \frac{1}{\rho}(\mathbf{x} - \mathbf{y}), \tag{4.8}$$

with the contraction factor $\rho > 0$, it leads to an algorithm, called the *projection method*. The associated optimization problem is the following quadratic programming problem

$$\mathbf{x}^{(j)} = \arg\min_{\mathbf{x} \in X} \sum_{k=1}^{n} \frac{1}{2} x_k^2 + \left(\rho g_k(\mathbf{x}^{(j-1)}) - x_k^{(j-1)} \right) x_k, \tag{4.9}$$

for which efficient and fast solution algorithms exist (Bazaraa *et al.*, 1993). In our research we use the route flow rates f_k^{rod} for the variable $\mathbf{x}$ and the route costs $\mathbf{c}_k^{rod}(\mathbf{f})$ for the function $\mathbf{g}(\mathbf{x})$. For algorithm 4.1 to converge, the route costs have to be strictly monotone in the route flows (Florian and Spiess, 1982). Section 5.4 gives the travel time functions which are used for the route costs. It can be seen that these functions are strictly monotone. To calculate the route costs based on the route flows, a dynamic network loading model is used. The simulation

period is split into smaller time steps and for every time step the dynamic route flows are loaded onto the network or propagated trough the network based on actual link travel times. After the simulation period is finished, route travel times are constructed and these can be used for calculating the route costs for every time period. The complete algorithm is algorithm 4.2.

The construction of a set of routes by enumeration is treated in section 4.4. The initial solution of the route flows $\mathbf{f}_k^{(0)}$ is based on the free flow travel times, including the uniform delay due to traffic signals (see chapter 5 for delay functions). The next two sections describe the choice of ρ and the convergence criterion.

Algorithm 4.2 *Deterministic dynamic user optimal assignment*

Step 1: *Construct a set of routes between every OD pair (see section 4.4).*

Step 2: *For each time period k determine an initial route flow solution $\mathbf{f}_k^{(0)} \in \Omega$.*
            Set $j := 1$.

Step 3: *Calculate route costs $\mathbf{c}_k^{(j)}(\mathbf{f}^{(j-1)})$ using a dynamic network loading model.*

Step 4: *For every time period k select a $\rho_k > 0$ such that $\rho_k < 1/L_k$ where L_k is the Lipschitz constant for the route cost function $\mathbf{c}_k(\mathbf{f})$ (equation 4.10)*

Step 5: *Calculate $\mathbf{f}_k^{(j)} \in \Omega$, $\mathbf{f}_k^{(j)} = \arg\min_{\mathbf{x} \in \Omega} \sum_{y=1}^{n} \frac{1}{2}x_k^2(y) + \left(\rho_k c_k(y) - f_k^{(j-1)}(y)\right) x_k(y)$.*

Step 6: *If convergence criterion is met, then stop.*
            Otherwise, set $j := j+1$ and return to Step 3.

4.1.2   Choice of ρ

In step 4 of algorithm 4.2 the contraction factor ρ_k is chosen such that $0 < \rho_k < 1/L_k$, where L_k is the Lipschitz constant for the route cost function $\mathbf{c}_k(\mathbf{f})$. The Lipschitz constant for function f is the smallest value $K \geq 0$ for which holds $|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$, for every (x_1, x_2) in the domain. Several choices for ρ_k are possible. The most obvious choice is to take $\rho_k =$ constant, but in practise choosing a good constant ρ_k, which leads to convergence, is difficult. In this research ρ_k is different from iteration to iteration and as close as possible to the Lipschitz constant to ensure maximum convergence speed. It is defined as

$$\rho_k^{(j)} = \frac{1}{L_k^{(j)}} - \eta = \left[\max \frac{\left|\mathbf{c}_k^{(j)} - \mathbf{c}_k^{(j-1)}\right|}{\left|\mathbf{f}_k^{(j)} - \mathbf{f}_k^{(j-1)}\right|} \right]^{-1} - \eta, \tag{4.10}$$

where j is the iteration number and η a small parameter, which is introduced to make sure that $\rho_k < 1/L_k$. Based on some test simulations the value of η is set to 0.01. To compare this choice
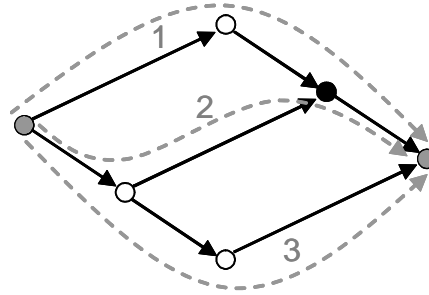
Figure 4.1: Small example network

for ρ with a constant ρ, a small example network was taken. The network consists of 7 links and 6 nodes and is shown in figure 4.1.

The network has one OD (origin-destination) pair (grey dots) with three routes (dashed grey lines) and one controlled intersection (black dot), with a fixed-time control plan. The simulation period for this example is divided into 4 quarters of an hour with increasing and decreasing demand. For this network and OD demand a deterministic dynamic user equilibrium (DDUE) was calculated for ρ = 79 and for the ρ as calculated with formula 4.10. The choice for ρ = 79 is based on the maximum ratio of the initial flows and the initial travel times. The DDUE is assumed to be reached if the convergence error is smaller than a threshold value. The convergence error represents the percentage of the demand for the OD pair that changes routes (see section 4.1.3). For the threshold value 0.1% was chosen. The results for the total delay in the network and the convergence error, as a function of the iteration, are shown in figure 4.2.

ρ is constant   ρ is variable
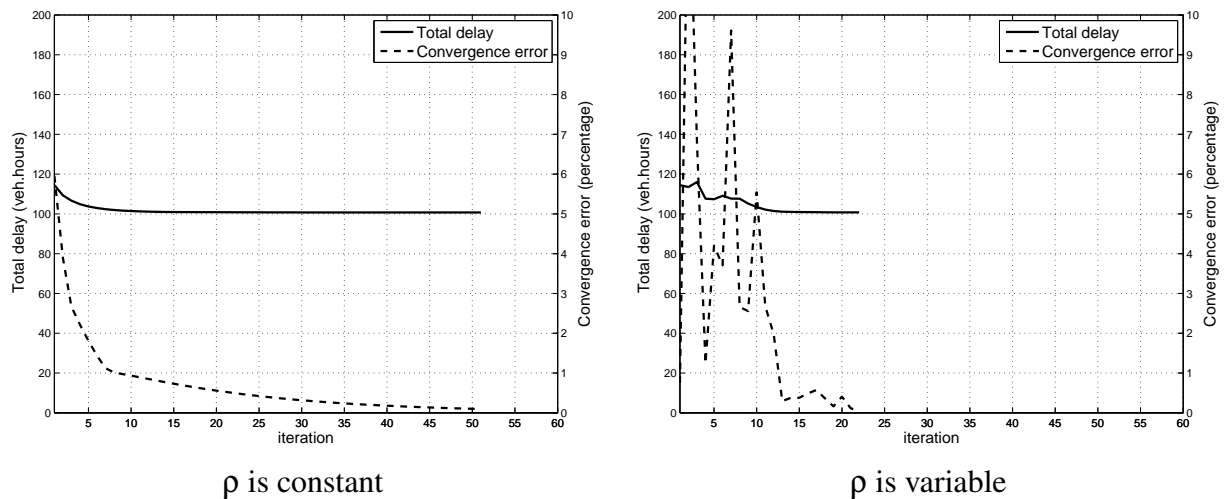
Figure 4.2: Results for different choices of ρ

The figure shows that for the example network with a variable ρ the convergence is much faster, although the convergence error is less smooth than for a constant ρ. For a constant ρ the equilibrium is reached in 51 iterations and for the variable ρ in 22 iterations. If the threshold value is decreased to 0.01%, the number of iterations for the constant ρ is 91 and for the variable

ρ 24. The results, in terms of route flows and route travel times, for the equilibrium situation with a convergence error of 0.1% are given in table 4.1.

Table 4.1: Flows and travel times for DDUE

		$\rho = 79$				$\rho = $ variable			
		Time period				Time period			
	Route	1	2	3	4	1	2	3	4
Flow	1	487.63	1029.31	630.13	254.60	487.55	1029.27	631.19	237.04
(veh/h)	2	784.39	1108.01	293.18	811.72	784.40	1107.96	293.85	836.18
	3	727.98	1862.68	1076.69	433.67	728.05	1862.77	1074.96	426.78
Travel time	1	7.99	9.13	8.67	7.70	7.99	9.13	8.68	7.68
(min)	2	7.99	9.13	8.67	7.67	7.99	9.13	8.68	7.68
	3	7.99	9.13	8.68	7.69	7.99	9.13	8.67	7.68

The table shows that the equilibrium route flows and travel times for both situations are almost equal and small differences are due to rounding errors. Only the fourth time period shows some larger differences in flows. These difference have disappeared if a convergence error of 0.01% is used. We can conclude that, for the example network, using a variable ρ gives faster convergence than a constant ρ. The equilibrium situation is the same, but is reached in far less iterations. Whether this also holds for other networks, is subject to further research.

4.1.3   Convergence

There are several possibilities to define convergence. One definition could be related to the equality of travel times on different routes for the same OD pair and time period. Another one could be defined by the difference between the travel times of two iterations. In this thesis the convergence error ε is defined as the maximum difference (over OD pairs and time periods) between the route flows of two iterations, calculated as a percentage of the demand of that OD pair during that time period:

$$\varepsilon = 100\% \cdot \max_k \max_{od} \max_r \frac{\left| f_k^{rod(j)} - f_k^{rod(j-1)} \right|}{q_k^{od}}. \tag{4.11}$$

It can be considered as the maximum shift in flow from one route to another, for a certain OD pair and time period. Convergence is achieved if the convergence error comes below a certain threshold value ε^*, which can be specified.

4.2   Stochastic dynamic user optimal assignment

For the stochastic dynamic user optimal assignment an error term is added to represent the perception error of the travellers concerning the travel costs. The stochastic dynamic user equilibrium (SDUE) can now be defined as in definition 4.2.

Definition 4.2 *(SDUE) For each origin-destination (OD) pair, any road user travelling between a specific OD pair and departing during a specific time interval, cannot improve his perceived route travel costs by unilaterally changing routes during that time interval.*

The perceived route costs $\widehat{c}_k^{rod}$ for OD pair *od*, route *r* and time period *k* can be represented by

$$\widehat{c}_k^{rod} = c_k^{rod} + \varepsilon_k^{rod}, \tag{4.12}$$

where c_k^{rod} are the real travel costs and ε_k^{rod} is the random component. If it is assumed that the random term is an independently and identically distributed Gumbel variate, then the multinomial logit (MNL) model is obtained (Sheffi, 1985). Given actual travel costs, the probability P_k^{rod} to choose route *r* for OD pair *od* and time period *k*, can then be described by

$$P_k^{rod} = \frac{e^{-\theta c_k^{rod}}}{\sum\limits_{s \in \Re^{od}} e^{-\theta c_k^{sod}}}, \quad \forall o, d, r \in \Re^{od}, k, \tag{4.13}$$

where $\theta > 0$ is a parameter that reflects the degree of uncertainty in the travel time knowledge of the road users. In the limit when θ approaches infinity, perfect knowledge is assumed and the deterministic user equilibrium solution is obtained. It is well known that the MNL model gives problems if routes overlap, among others explained by Ramming (2002). To overcome this problem several extensions of the MNL model have been developed. In the next sections two of them are discussed: the C-logit model and the Path size logit model.

4.2.1 C-logit model

The C-logit model was proposed by Cascetta *et al.* (1996). This model takes into account overlap in routes with the so-called commonality factor *CF*. There are several possibilities to define *CF*. In their work Cascetta *et al.* (1996) investigate four definitions. Based on some theoretical networks and the comparison of the outcome with other route choice theories, they conclude that the following specification is the best. This specification is given, for route *r* of OD pair *od* per time period *k*, by

$$CF_k^{rod} = \beta \ln \sum\limits_{s \in \Re^{od}} \left[\frac{L_{rs}}{\sqrt{L_r L_s}} \right]^{\gamma}, \quad \forall o, d, r \in \Re^{od}, k, \tag{4.14}$$

where L_r and L_s are the 'lengths' of routes *r* and *s* belonging to OD pair *od*. L_{rs} is the 'length' of the common links shared by routes *r* and *s* and β and γ are positive parameters, to be estimated or calibrated. 'Length' can be the physical length or the 'length' determined by travel costs. In this thesis free flow travel times are used. With this commonality factor and the known travel costs, the probability to choose route *r*, for OD pair *od* and time period *k*, and the corresponding flow *f*, are given by

$$P_k^{rod} = \frac{e^{-\theta c_k^{rod} - CF_k^{rod}}}{\sum\limits_{s \in \Re^{od}} e^{-\theta c_k^{sod} - CF_k^{sod}}}, \quad \forall o, d, r \in \Re^{od}, k, \tag{4.15}$$

$$f_k^{rod} = P_k^{rod} q_k^{od}, \quad \forall o, d, r \in \Re^{od}, k. \tag{4.16}$$

4.2.2   Path size logit model

A drawback of the C-logit model is the lack of theory on which the form of the commonality factor is based. In his PhD thesis Ramming (2002) introduces the path size term, as a promising way to specify a commonality factor. It is based on the discrete choice theory for aggregating alternatives, which has been used in other transportation settings such as destination choice. It leads to the path size logit route choice model (PS-logit), which is formulated as

$$P_k^{rod} \;=\; \frac{e^{-\theta c_k^{rod} + \ln PS_k^{rod}}}{\sum\limits_{s \in \Re^{od}} e^{-\theta c_k^{sod} + \ln PS_k^{sod}}}, \quad \forall o,d,r \in \Re^{od}, k, \tag{4.17}$$

$$PS_k^{rod} \;=\; \sum_{a \in \Gamma_r} \left(\frac{l_a}{L_r}\right) \frac{1}{\sum\limits_{s \in \Re^{od}} \left(\frac{L_r}{L_s}\right)^{\gamma'} \delta_{as}}, \quad \forall o,d,r \in \Re^{od}, k, \tag{4.18}$$

where most variables have the same meaning as in previous equations, Γ_r is the set of links belonging to route r, l_a is the length (distance or travel time) of link a, γ' is a parameter and δ_{as} is the link-coincidence matrix, which is 1 if link a belongs to route s and 0 if that is not the case. The PS-logit model shows a better fit on real life data for the Boston area than the C-logit model, but the results strongly depend on the parameter γ' and are only better than the results for the C-logit model if $\gamma' \geq 99$. A large value for γ' ($\gamma' = 40$) is also reported by Balakrishna (2002). Such large values of γ' cause computer accuracy problems if the algorithm is used, which is also noted by Hoogendoorn-Lanser *et al.* (2005) for route choice in multi-modal networks. Therefore, for the moment, we prefer and use the C-logit model.

4.2.3   Solution algorithm

To obtain the stochastic dynamic user equilibrium, algorithm 4.3 is used, which is comparable to algorithm 4.2, except for the calculation of flows and the smoothing of the flows, which is necessary to make the algorithm converge.

Algorithm 4.3 *Stochastic dynamic user optimal assignment*

Step 1:   *Construct a set of routes between every OD pair and calculate commonality factors with equation 4.14.*

Step 2:   *For each time period k determine an initial route flow solution* $\mathbf{f}_k^{(0)} \in \Omega$. *Set* $j := 1$.

Step 3:   *Calculate route costs* $\mathbf{c}_k^{(j)}(\mathbf{f}^{(j-1)})$ *using a dynamic network loading model.*

Step 4:   *Calculate new route flows* $\mathbf{f}_k^{(j)} \in \Omega$ *with equations 4.15 and 4.16.*

Step 5:   *Smooth route flows with* $\mathbf{f}_k^{(j)} = \mathbf{f}_k^{(j-1)} + \zeta^{(j)}(\mathbf{f}_k^{(j)} - \mathbf{f}_k^{(j-1)})$

Step 6:   *If convergence criterion is met, then stop.*
Otherwise, set $j := j+1$ *and return to Step 3.*

In step 5 the flows are smoothed. Normally, the method of successive averages (MSA) is used to smooth the flows ($\zeta^{(j)} = 1/j$). The convergence of the MSA is slow, because the step size ζ quickly becomes small and slowly decreases. To overcome this problem, a somewhat adjusted MSA is used, which is discussed in the next section.

4.2.4   Smoothing the flows

The step size for smoothing the flows is chosen in such a way that in the first few iterations the step size is larger than the step size for MSA, and is smaller in the next iterations to speed up convergence. To achieve that, an exponential term is added. Therefore, the adjusted step size is calculated as

$$\zeta^{(j)} = a_1 e^{-a_2 j} + \frac{a_3}{j}, \tag{4.19}$$

where j is the iteration number and a_1, a_2 and a_3 are parameters. According to Sheffi (1985), necessary conditions for algorithm 4.3 to converge are

$$\sum_{j=1}^{\infty} \zeta^{(j)} = \infty \text{ and } \sum_{j=1}^{\infty} \left(\zeta^{(j)} \right)^2 < \infty. \tag{4.20}$$

If a_1, a_2 and a_3 are chosen such that $a_1 > 0$, $a_2 > 0$ and $a_3 > 0$, these conditions are fulfilled (see appendix B). Figure 4.3 gives the value of the step size as a function of the iteration, compared with normal MSA, for the arbitrary choices $a_1 = 0.95$, $a_2 = 0.25$ and $a_3 = 0.05$.
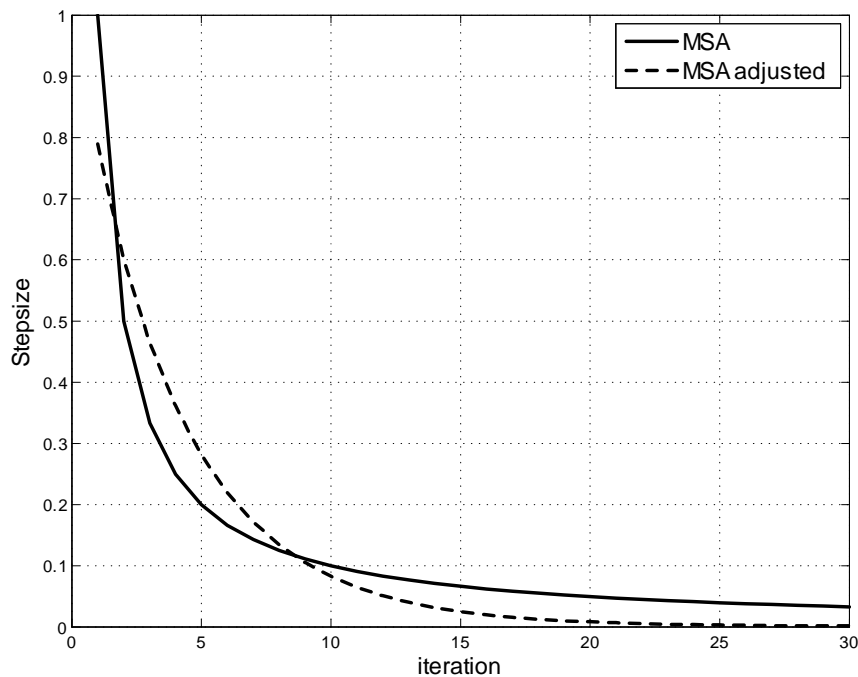
Figure 4.3: Step size values

Using the adjusted step size can make a difference in the number of iterations required, espe-
cially for small thresholds for the convergence error. For the example of figure 4.1, the results
for the different step size definitions, if the threshold for the convergence error is set to 0.01,
are shown in figure 4.4. The results show that for the network used, the convergence error for
normal MSA is smaller for the first 7 iterations, but then it takes more iterations to converge
than for the adjusted MSA: 47 iterations compared to 25 iterations. If larger threshold values
are used, the difference in number of iterations becomes smaller.

Normal MSA   Adjusted MSA
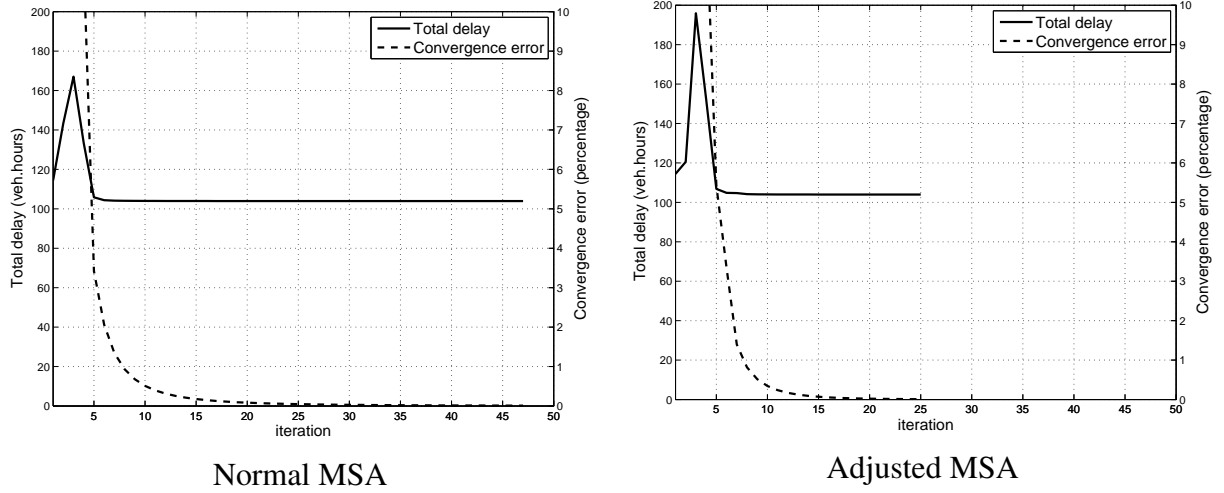
Figure 4.4: Results for different MSA types

4.2.5 Convergence

The criterion for convergence for stochastic assignments can be defined in different ways. Two
are used in this thesis and in the examples presented. The first criterion is the same as for
deterministic assignments (see section 4.1.3) and looks at the change in route flows for every
OD pair from one iteration to the next one. This change is normalised using the OD demand.
The maximum change over the OD pairs and time periods is compared with a threshold value.
In formula:

$$\max_{k} \max_{od} \max_{r \in \Re^{od}} \frac{\left| f_k^{rod(j)} - f_k^{rod(j-1)} \right|}{q_k^{od}} < \varepsilon^*, \tag{4.21}$$

where ε^* is the convergence threshold. Normally, values from 0.1%-5% are used in practise.
Bliemer and Taale (2006) state that, although this criterion is useful to see that at some point the
algorithm can be terminated, theoretically it does not guarantee that the algorithm has reached
a (stochastic) dynamic user equilibrium. This is due to the fact that applying MSA or other
averaging schemes yield smaller route flow changes in each iteration by default, so the use of
flows is not a good measure for convergence to a user equilibrium. A better measure is the

dynamic relative duality gap G, which is defined as

$$G = \frac{\sum\limits_{o,d} \sum\limits_{r \in \Re^{od}} \sum\limits_k f_k^{rod}(c_k^{rod} - \pi_k^{od})}{\sum\limits_{o,d} \sum\limits_k q_k^{od} \pi_k^{rod}}, \qquad (4.22)$$

where

$$\pi_k^{od} \equiv \min_{r \in \Re^{od}} c_k^{rod}, \quad \forall o,d,k. \qquad (4.23)$$

The duality gap gives the difference between the realised travel costs and the equilibrium travel costs. It goes to zero in case of a deterministic user assignment, because then all travellers use the route with minimum costs. In case of a stochastic assignment the gap function stabilises to a value larger than zero, because not all travellers use the route with the least costs. To which value the gap converges is not known beforehand. To be able to use it, we therefore monitor the relative change in the gap function after every iteration with

$$\frac{\left|G^{(j)} - G^{(j-1)}\right|}{G^{(j-1)}} < \varepsilon^*. \qquad (4.24)$$

If G converges, then also the relative change in G will converge. Tests with the small network from 4.1 and using $\varepsilon^* = 0.01\%$, show that both criteria lead to almost the same stochastic equilibrium, but using the dynamic relative duality gap takes more iterations to converge: 79 versus 25 iterations. The results are shown in figure 4.5 and table 4.2.

Flow criterion   Duality gap
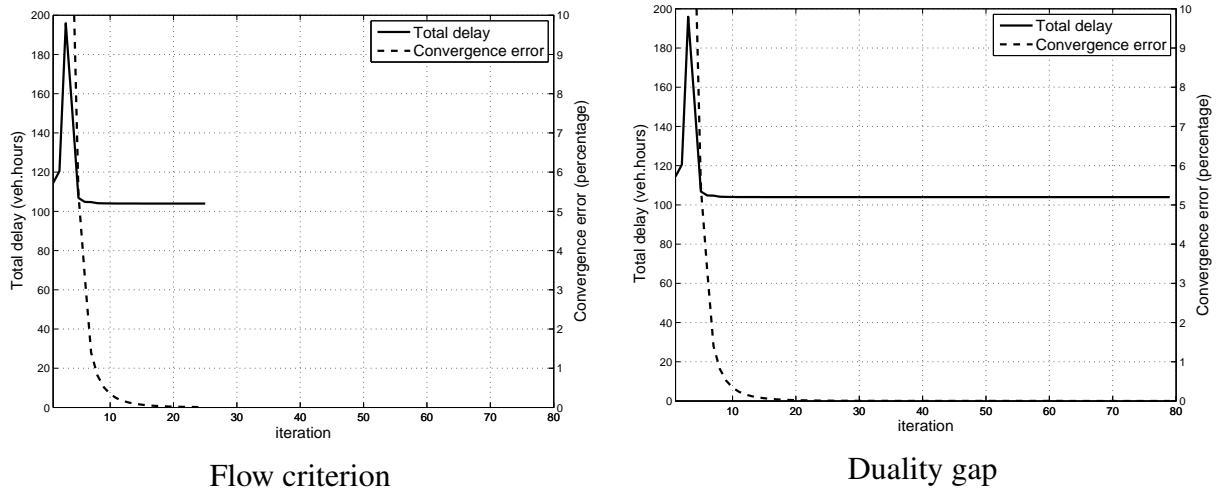
Figure 4.5: Results for using the flow criterion or the duality gap

The difference is caused by using the product of flow and travel time instead of only using flows. Because, for the example network, the flow criterion takes less iterations and leads to the same equilibrium, we choose the maximum change in route flows as the criterion for convergence. Further research should reveal if this was the right choice.

Table 4.2: Flows and travel times for DSUE

	Route	Flow criterion				Duality gap			
		Time period				Time period			
	Route	1	2	3	4	1	2	3	4
Flow	1	572.62	1086.58	572.90	442.36	572.61	1086.56	572.96	441.70
(veh/h)	2	731.77	1158.27	381.91	554.47	731.78	1158.26	381.67	556.54
	3	695.61	1755.18	1045.19	503.17	695.61	1755.19	1045.37	501.77
Travel time	1	8.10	9.42	9.08	7.84	8.10	9.42	9.08	7.83
(min)	2	7.84	9.30	9.40	7.53	7.84	9.30	9.40	7.53
	3	7.94	9.05	8.62	7.75	7.94	9.05	8.62	7.75

4.3   System optimum assignment

The system optimum is the situation where the travel costs are minimal for the system as a whole. Users choose their route in such a way that it is optimal for the system. This situation is in general not equal to a user equilibrium in which each user chooses the route which minimises his costs. The system optimum assignment can be defined as an optimization problem of finding the route flows $\bar{\mathbf{f}}$ such that

$$\bar{\mathbf{f}} = \arg \min_{\mathbf{f} \in \Omega} \sum_k \sum_{o,d} \sum_{r \in \mathfrak{R}^{od}} f_k^{rod} c_k^{rod}(\mathbf{f}) \tag{4.25}$$

This optimisation problem can be solved in various ways, which is, among other things, discussed by Peeta (1994). In our research, evolutionary algorithms are used to solve the problem, because not only the route flows, but also the signal settings should be optimised. We always use the system optimum assignment in combination with system optimum control (see section 3.5).

4.4   Enumeration of routes

A road user, travelling from an origin A to a destination B, uses a route through the network. He knows about the route for example by experience, through a road map, or a route planning program or device. Maybe he has knowledge of more routes from A to B, but he has chosen the route he uses for some reason. Possibly the route is the shortest or fastest one (in the perception of the road user) or could also be the most comfortable one. Route choice behaviour is still an important topic for research. Aspects as the learning process, among other things discussed by Bogers *et al.* (2006a), and the reliability, e.g. investigated by Bogers *et al.* (2006b), are open for further research.

Instead of finding the shortest routes continuously as part of the assignment and simulation process, it is assumed that the road user chooses a route from a predefined, enumerated set of routes. The advantage of this is that the routes can be computed in advance. The disadvantage is that the set of routes is fixed and cannot change during the assignment.

The enumeration of routes can be treated in several ways. Bliemer (2001) gives a short summary of some alternatives: all acyclic routes, the k-shortest routes, the essentially least cost routes, the most probable routes and efficient routes.

- *All acyclic routes*. All routes which contain no cycles are calculated and used. This method is computationally very expensive and also not very realistic, because most road users are not aware of all alternatives.

- *K-shortest routes*. This method gives the shortest route, the shortest route but one, etc., until *k* routes are found. It is intuitively better than the previous method. An example of such an algorithm is given by Eppstein (1998).

- *Essentially least cost routes*. This method is similar to the k-shortest routes method, but restricts the routes to a certain predefined bandwidth. Only routes that have a length within a certain bandwidth are included in the set of routes.

- *Most probable routes*. This is another method to restrict the set of routes, using Monte Carlo simulation. A number of times the link costs are randomly varied according to a certain distribution and the shortest route is calculated and added to the route set.

- *Efficient routes*. This method extends the notion of 'shortest route' to a multi criteria definition, which includes more than just travel time or travel costs.

It would require a separate thesis to describe all details and properties of these methods. The method used in this thesis is a combination of the third and fourth method described above. The number of routes for each OD pair is limited and the routes are bounded in length. The set is generated with a stochastic process. Properties of stochastic choice set generation are described by Li *et al.* (2007). With a Monte Carlo simulation the link costs are varied randomly, but within a certain bandwidth, using a scale factor:

$$\mathbf{C}^* = \mathbf{C}(1 + \omega |\Lambda|) \tag{4.26}$$

where $\mathbf{C}^*$ is the adjusted link cost matrix, $\mathbf{C}$ the original link cost matrix, ω a scaling factor and Λ a matrix of the correct size and with elements following the standardized normal distribution $N(0,1)$. The link cost matrix gives the costs to travel each link, normally represented by the costs to travel from one node to another. So, for every link a separate, extra cost term is added. With a 99.74% chance the elements of Λ lie between -3 and 3. If $\omega = \frac{2}{3}$ the length of the adjusted routes can never be longer than three times the original length with a chance of 99.74%. With the adjusted link costs the shortest route can be calculated and added to the set of routes, if it is shorter than the k-th shortest route and does not have too much overlap with existing routes in the set. This last condition is added to avoid routes that use the off-ramp and on-ramp instead of the motorway. The algorithm for the enumeration of routes is given in algorithm 4.4.

In the steps 2 and 3 a shortest route has to be calculated. This can be done with Dijkstra's or another algorithm. We have used the heap implementation of Dijkstra's algorithm of Bindel

(2004). This is a fast and available MATLAB implementation of the Dijkstra algorithm and could be easily integrated into the framework. Bliemer and Taale (2006) show that this algorithm can be used to generate routes for large networks in an acceptable amount of time.

Algorithm 4.4 *Enumerate routes*

Step 1:   *Choose number of wanted routes k and number of random drawings m^*.*

Step 2:   *For every OD pair, find the shortest route for the original cost matrix $\mathbf{C}$ and add this route to the set of routes $\mathfrak{R}^{od}$.*

Step 3:   *For $m = 1 : m^*$*
Draw the error matrix $\Lambda^{(m)}$.
Calculate the adjusted link costs with $\mathbf{C}^{(m)} = \mathbf{C}(1 + \omega |\Lambda^{(m)}|)$.
For every OD pair and the cost matrix $\mathbf{C}^{(m)}$, find the shortest route r^{odm}.

Step 4:   *For every OD pair and route r^{odm}:*
Check if r^{odm} is shorter than the k-th longest route in the set $\mathfrak{R}^{od}$.
Check if there is not too much overlap with the other routes in the set: for every route $s \in \mathfrak{R}^{od}$, the number of overlapping links divided by the minimum number of links in s or r^{odm} should be smaller than a threshold.
If these condition are fulfilled, add route r^{odm} to route set $\mathfrak{R}^{od}$

4.5   Summary

In this chapter the dynamic traffic assignment methods were discussed. Attention was paid to three different assignment methods: deterministic, stochastic and system optimal. The first two methods were described in detail with a problem description and a solution method. Special attention was paid to the contraction factor for the deterministic assignment. It was shown that for an example network using a new, dynamic contraction factor, based on the results of the current and previous iteration, leads to faster convergence. For the stochastic assignment the smoothing factors were an issue to elaborate on. It was shown, for the same example network, that using adjusted MSA factors led to faster convergence. For the convergence criterion of the dynamic stochastic user equilibrium assignment two options were investigated: one based on flow and one based on the duality gap. Both criteria gave the same equilibrium, but for the example network the flow criterion showed a faster convergence. The system optimum assignment is always used in combination with system optimum control. We use an evolutionary algorithm to solve both problems in one step. All assignment methods are route based. Therefore, a route set needs to be constructed. A construction process using enumeration was described.

5

Dynamic Network Loading Model

In the previous chapter it is explained how the OD flows are distributed across the available routes. The result is a set of route flows for every time period. To obtain information about the resulting link indicators, congestion in the network and travel times, the route flows are loaded onto the network. This is done with a deterministic dynamic network loading (DNL) model. The model is deterministic, because no random components are used, also not in the travel time functions. The model is dynamic, because the loading of the network evolves over time. At the time the author started with his research (way back in 1999) no DNL models were readily available, so it was decided to develop one, which could easily be used in the research framework.

5.1   General description of the DNL model

In general the model works as follows. At the origins traffic is loaded according to the demand profile. The traffic is propagated from origin to destination through the network link by link, based on link travel time functions. At decision nodes the traffic is distributed among the outgoing links according to the splitting rates determined by the route flows. Before traffic enters a link, it is checked whether or not this link has enough space. If not, (a part of) the traffic that wants to enter, is stored on the upstream link (blocking back, also for other traffic). The travel time functions used for the links are standard travel time functions from literature. They depend on the junction type the link connects to (standard junction, signalised junction, roundabout or priority junction). The link travel times are used to calculate the route travel times, which are input for the dynamic assignment model. A special feature of the model is the handling of short links. Normally, the length of the time step for the DNL is chosen such that traffic is not able

to traverse the shortest link in the network in one time step. In large networks this can give long calculation times. To circumvent this problem, short links are treated in a special way. In algorithm 5.1 the steps are shown and in subsequent sections the DNL model is described in more detail.

Algorithm 5.1 *Dynamic network loading model*
Step 1:   *Initialise all variables needed.*
Step 2:   *Calculate splitting rates μ_k^n for every node n and time period k.*
Step 3:   *Propagate traffic through the network using the following steps:*
            Determine the initial flow per link.
            Determine free flow travel time and capacity per link and time period.
            Divide each time period in a number of time steps.
            For every time step t:
            - Determine delay and travel time per link.
            - Calculate the outflow per link.
            - For every node determine the inflow.
            - Calculate the available space for every link.
            - For every node n determine the outflow with the splitting rates μ_k^n.
            - Determine the inflow for every link.
            - Determine links with blocking back and adjust the outflow for those links.
            - Calculate link indicators.
Step 4:   *Calculate delay and travel costs c_k^{rod} for every route r from origin o to destination d and time period k.*

Step 2 is discussed in section 5.2 and step 3 in section 5.3. A separate section (section 5.4) is devoted to the delay and travel time functions used. Section 5.5 describes how route travel times or costs are calculated (step 4). Finally, section 5.6 discusses how the model deals with short links.

5.2   Splitting rates

The network is a directed graph $(\mathcal{N}, \mathcal{A})$ with nodes $\mathcal{N}$ and directed links $\mathcal{A}$. Time is discretised in periods of a certain length. A route r consists of an ordered set of links $\Gamma_r = \{a \mid a \in \mathcal{A}\}$, which connects an origin o to a destination d and has a flow f_k^{rod} for every time period k. Every link a belongs to a set of routes $\mathcal{R}_a = \{r \mid r \in \mathfrak{R}\}$ and the flow on link a is determined by the route flows that use this link. For every node n and time period k the splitting rate $\mu_k^{na_i a_j}$ distributes the traffic of incoming link a_i to outgoing link a_j. The link a_i belongs to the set of

incoming links $\mathcal{A}_i^n$ of node n and $a_j \in \mathcal{A}_j^n$, the set of outgoing links. The splitting rate $\mu_k^{na_ia_j}$ is calculated as

$$\mu_k^{na_ia_j} = \frac{\sum\limits_{r \in \mathcal{R}_{a_i} \cap \mathcal{R}_{a_j}} f_{k'}^{rod}}{\sum\limits_{r \in \mathcal{R}_{a_i}} f_{k'}^{rod}}, \quad \forall n \in \mathcal{N}, \, a_i \in \mathcal{A}_i^n, \, a_j \in \mathcal{A}_j^n, \, k, \tag{5.1}$$

where $f_{k'}^{rod}$ is the flow (veh/h) for route r, belonging to OD pair od, and time period k'. Time period k' is the period in which the traffic should depart to arrive at node n in time period k. To determine k' the free flow travel time is used. It would be better to use the real travel time, but that would require that the travel times are already known, which is not the case, because the splitting rates are calculated first. Normally, an iterative process can be implemented by using the travel times from the previous iteration, but that would imply another loop, an increase of the calculation time and convergence issues. For now, free flow travel times are used. For origin nodes o the demand is distributed to the outgoing links with

$$\mu_k^{oa_j} = \frac{\sum\limits_{r \in \mathcal{R}_{a_j}} f_k^{rod}}{\sum\limits_{d} q_k^{od}}, \quad \forall o, \, a_j \in \mathcal{A}_j^o, \, k. \tag{5.2}$$

The splitting rates are used in the propagation of traffic through the network, which is described in the next section. To illustrate the calculation of splitting rates, we take a look at figure 5.1.
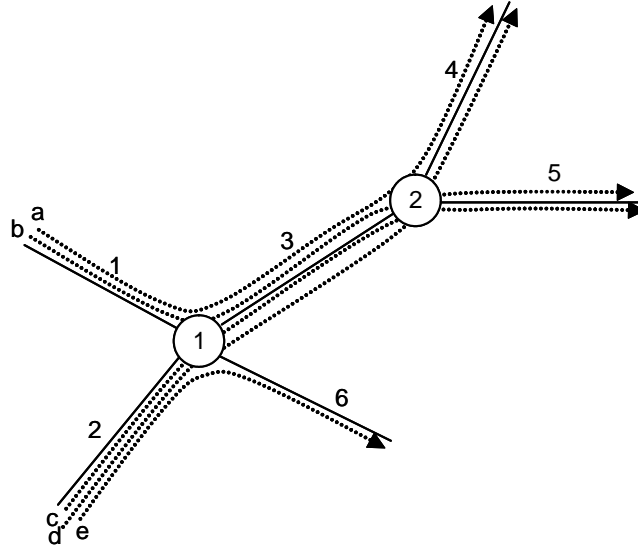
Figure 5.1: Part of a network with 5 routes

This figure shows part of a network with 2 nodes, 6 links and 5 routes using these links. The start and end of these routes is somewhere else in the network. The route flows for 3 periods are given in table 5.1. The table specifies the amount of vehicles departing from a zone and starting a certain route, during a certain period. Suppose we want to calculate the splitting rates

for period 3 and suppose that at node 1 traffic arrives, for route a from period 1, for route b from period 2, for route c from period 1, for route d from period 2 and for route e from period 3. For node 2 the same holds, because the nodes are not far apart. Then

$$
\begin{aligned}
\mu_3^{1-1-3} &= \frac{100+50}{100+50} = 1, \; \mu_3^{1-1-6} = \frac{0}{100+50} = 0, \\
\mu_3^{1-2-6} &= \frac{200+30}{200+30+125} = 0.6479, \; \mu_3^{1-2-6} = \frac{125}{200+30+125} = 0.3521, \\
\mu_3^{2-3-4} &= \frac{100+30}{100+50+200+30} = 0.3421, \; \mu_3^{2-3-5} = \frac{50+200}{100+50+200+30} = 0.6579.
\end{aligned}
$$

Table 5.1: Route flows for splitting rates

	Period		
Routes	1	2	3
a	100	150	100
b	40	50	60
c	200	100	25
d	10	30	60
e	125	125	125

5.3   Propagation of traffic

To capture the changes in demand, the simulation period (e.g. a morning peak period) is divided into T' time periods, with a length typically between 5 and 30 minutes. To propagate the traffic through the network, each time period is divided into T time steps. These time steps have a length Δ_s (s), which is typically between 5 and 30 seconds. It is assumed that during the time step the arrival rate of the vehicles is constant and the conditions on the link are uniform, which means that the density is equal everywhere on the link. For every time step, several traffic variables are determined in a specific order: degree of saturation for the links, link delay and travel time, link outflow, node inflow, available link space, link inflow, node outflow, corrected link outflow, queue length and link flow for the next time step.

The degree of saturation φ is demand divided by capacity. In this case, the demand is defined as the number of vehicles that can reach the end of the link within the time step. If it is assumed that flow is uniformly distributed along the link, φ can be calculated with

$$
\varphi_{at} = \frac{\Delta_s}{\tau_{a,t-1}} \cdot \frac{\chi_{a,t-1}}{Q''_{ak}\Delta_h}, \quad \forall a \in \mathcal{A}, t, \tag{5.3}
$$

where χ is the number of vehicles on the link (veh), τ the travel time (s), Q''_{ak} the capacity or saturation flow (veh/h) at the end of the link, for time period k to which time step t belongs, and Δ_h the length of the time step in hours ($\Delta_s/3600$). The capacity can vary in time due to signal control. Because, signal control varies only per time period k and gives an average ratio

of green time and cycle time, also the capacity only changes with period k and not with every time step t. Furthermore, for computational reasons $Q''_{ak} > 0$ to prevent division by zero. The degree of saturation is used to calculate the link delay travel times. The link delay d_{at} and travel time τ_{at} are dependent on the link type. The different link types and accompanying travel time functions are described in the next section. Based on the travel time, the link outflow v (veh) can be determined as

$$v_{at} = \max(u_{a,t-1}\frac{s_a^p}{s_a}, \frac{\Delta_s}{\tau_{at}}\chi_{at} + \kappa_{a,t-1}), \quad \forall a \in \mathcal{A}, t, \tag{5.4}$$

where u_{at} (veh) is the inflow for link a on time step t, s_a is a parameter to determine the number of time steps for link a and κ_a is the queue (veh) for link a. On the outflow the following constraints can be imposed

$$v_{at} \leq Q''_{ak}\Delta_h, \tag{5.5a}$$
$$v_{at} \geq 0. \tag{5.5b}$$

The unconstrained outflow $\widetilde{v}$ is the same as the normal outflow, given by 5.4, but, as the name indicates, without the constraints. In equation 5.4, the maximum with the (adjusted) inflow of the previous time step $u_{a,t-1}$ is needed to make sure the outflow is large enough in congested conditions, when the travel time τ_{at} is long. Otherwise, it takes too much time for the outflow to reach the level of the inflow if the inflow changes. To achieve that a change in inflow reaches the end of the link in the calculated number of time steps and is not spread out across the link (especially for long links), the inflow is adjusted with a factor to be certain that after s_a steps the outflow (at least) equals the inflow. The number of steps is determined by τ_{at}/Δ and s_a^p is the pth time step after a change in inflow. Of course the outflow cannot be larger than the capacity of the link. To show the difference without and with the adjustment, a small network was simulated, for which the results are shown in figure 5.2.

| Without adjustment | With adjustment |
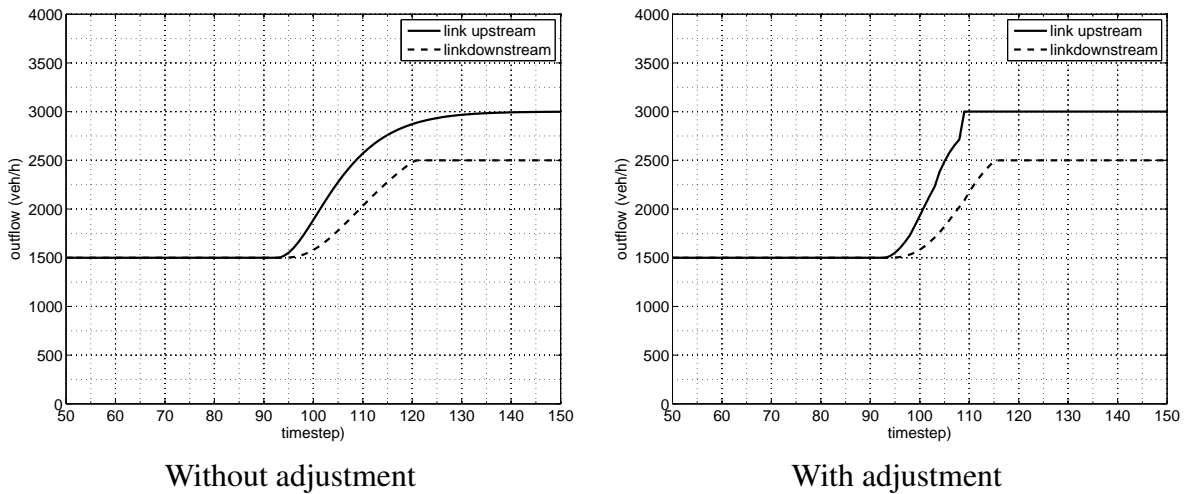
Figure 5.2: Outflows for two links with and without adjustment

The network has two links. At the end of the downstream link there is a bottleneck of 2500 veh/h. The inflow profile is 1500 veh/h for the first time period of 900 seconds and 3000 veh/h for the second time period. Figure 5.2 shows the outflow of both links for two situations: one where only the second term of the max function of equation 5.4 is used and one where both terms are used. From this figure it is clear that without the adjustment it takes more time steps (47 versus 16) to reach the required level of outflow. This is due to the spread of traffic across the link without the adjustment. The difference is larger when the travel time on the link increases.

The inflow U_{nt} (veh) in node n for time step t is the sum of the outflows of the incoming links

$$U_{nt} = \sum_{a \in \mathcal{A}_i^n} v_{at}, \quad \forall n \in \mathcal{N}, t. \tag{5.6}$$

Before the flow is distributed to the outgoing links, the amount of vehicles these links can handle, is determined. The available space ψ (veh) on the link is calculated as

$$\psi_{at} = \frac{l_a p_a}{\overline{l}^{veh}} - \chi_{at}, \quad \forall a \in \mathcal{A}, t, \tag{5.7}$$

where l_a is the length of the link (m), p_a is the number of lanes and $\overline{l}^{veh}$ the average length a vehicle occupies, including the space between vehicles. The available space is a maximum for the link inflow, just like the link capacity. The inflow u (veh) for an outgoing link $a_j \in \mathcal{A}_j^n$ of a node n is defined as

$$u_{a_j t} = \sum_{a_i \in \mathcal{A}_i^n} (\widetilde{v}_{a_i t} - \kappa_{a_i, t-1}) \mu_k^{n a_i a_j} + \overline{\kappa}_{a_i a_j, t-1}, \quad \forall a_j \in \mathcal{A}_j^n, t, \tag{5.8}$$

where k is the time period to which time step t belongs, $\widetilde{v}_{a_i t}$ the unconstrained outflow and $\overline{\kappa}_{a_i a_j, t-1}$ that part of the queue which is related to outlink a_j. The link inflow is subject to the constraints

$$
\begin{align}
u_{at} &\leq \psi_{at}, \tag{5.9a} \\
u_{at} &\leq Q'_{ak} \Delta_h, \tag{5.9b} \\
u_{at} &\geq 0 \tag{5.9c}
\end{align}
$$

leading to a restricted link inflow $u'_{at} \leq u_{at}$. The capacity Q' (veh/h) is the capacity at the beginning of the link. This capacity can be different from the capacity at the end of the link Q'' due to signal timings or other restrictions. The total outflow V_{nt} (veh) from node n now becomes

$$V_{nt} = \sum_{a \in A_j^n} u'_{at}, \quad \forall n \in \mathcal{N}, t. \tag{5.10}$$

and due to the restrictions 5.9 imposed on u, this node outflow can be smaller than the node inflow U_{nt}. The difference $W_{nt} = U_{nt} - V_{nt}$ is the traffic that is not able to pass the node and

stays on the incoming links (blocking back, also for other traffic, e.g. traffic on an on-ramp can block the traffic on an upstream intersection) and therefore also the outflows on the incoming links have to be corrected. This is done only for the incoming links having a relation with the outgoing links causing the blocking back. For all outgoing links $a_j \in \mathcal{A}_j^n$ of a node n the amount of traffic w (veh), not able to pass the node is

$$w_{a_j t} = u_{a_j t} - u'_{a_j t}, \quad \forall a_j \in \mathcal{A}_j^n, \, t. \tag{5.11}$$

There are several methods to distribute this amount of traffic to the incoming links, e.g. uniform, proportional to the flow on the links, proportional to the outflow capacity Q'' or (which is often the same) proportional to the number of lanes. If it is assumed that traffic behaves in a polite manner, every lane gets its turn one by one. Therefore, the option to distribute the blocked traffic proportional to the incoming lanes, was chosen. First, the outflow of a link is split into the different directions

$$v_{a_i a_j t} = (\tilde{v}_{a_i t} - \kappa_{a_i,t-1}) \mu_k^{na_i a_j} + \overline{\kappa}_{a_i a_j, t-1}, \quad \forall a_i \in \mathcal{A}_i^n, \, \forall a_j \in \mathcal{A}_j^n, \, t, \tag{5.12}$$

where $\overline{\kappa}_{a_i a_j, t-1}$ is that part of the queue on the incoming link a_i which is heading for the outgoing link a_j. For every direction the possible outflow $\overline{v}_{a_i a_j t}$ is

$$\overline{v}_{a_i a_j t} = w_{a_j t} \frac{p_{a_i}}{\sum\limits_{\left\{ a_i | \mu_k^{na_i a_j} > 0 \right\}} p_{a_i}}, \quad \forall a_i \in \mathcal{A}_i^n, \, \forall a_j \in \mathcal{A}_j^n, \, t. \tag{5.13}$$

If the possible outflow of a link is larger than the outflow itself, there is room for other incoming links to fill the gap. The size of this available gap for outlink a_j is defined as

$$\overline{w}_{a_j t} = \sum\limits_{\left\{ a_i | \overline{v}_{a_i a_j t} > v_{a_i a_j t} \right\}} \left(\overline{v}_{a_i a_j t} - v_{a_i a_j t} \right), \quad \forall a_j \in \mathcal{A}_j^n, \, t. \tag{5.14}$$

The corrected outflow $v'_{a_i a_j t}$ of the incoming links $a_i \in \mathcal{A}_i^n$ to the outlink $a_j \in \mathcal{A}_j^n$ of a node n now becomes

$$v'_{a_i a_j t} = \begin{cases} \overline{v}_{a_i a_j t} + \overline{w}_{a_j t} \dfrac{p_{a_i}}{\sum\limits_{\left\{ a_i | \mu_k^{na_i a_j} > 0 \right\}} p_{a_i}} & v_{a_i a_j t} > \overline{v}_{a_i a_j t} \\[2em] v_{a_i a_j t} & v_{a_i a_j t} \leq \overline{v}_{a_i a_j t} \end{cases}, \quad \forall a_i \in \mathcal{A}_i^n, \, \forall a_j \in \mathcal{A}_j^n, \, t, \tag{5.15}$$

and the combined corrected outflow

$$v'_{a_i t} = \sum\limits_{a_j \in \mathcal{A}_j^n} v'_{a_i a_j t}, \quad \forall a_i \in \mathcal{A}_i^n, \, t. \tag{5.16}$$

Blocking back or a restricted capacity causes a (horizontal) queue on a link, which expands and shrinks in time. The queue plays an important role in the calculation of delay and travel time. These calculations take into account existing queues. The queue grows because of vehicles

arriving at the tail of the queue and it shrinks because vehicles leave the link. To avoid unnecessary queues and unrealistic delays, the growth is limited by the inflow. The length of the queue (veh) κ is given by

$$\kappa_{at} = \widetilde{v}_{at} - v'_{at}, \quad \forall a \in \mathcal{A}, t, \tag{5.17}$$

and is subject to the constraints

$$\kappa_{at} \quad \leq \quad \frac{l_a p_a}{\overline{l}_{veh}} - \psi_{at}, \tag{5.18a}$$

$$\kappa_{at} \quad \geq \quad 0. \tag{5.18b}$$

In reality it is possible that the front of the queue is already moving, while the back of the queue still moves backward due to the traffic demand. This can happen due to traffic signals or removing a capacity restriction. This 'shock wave' phenomenon is not modelled. The bottleneck always stays on the same location.

Finally, the number of vehicles χ on link a for the next time step is given by (conservation of flow)

$$\chi_{a,t+1} = \chi_{at} + u'_{at} - v'_{at}, \quad \forall a \in \mathcal{A}, t, \tag{5.19}$$

and the next time step is processed, starting with equation 5.3.

To initiate the algorithm the link flows and travel times have to be defined for the first time step. The initial number of vehicles for all links a is defined by

$$\chi_{a,0} = \frac{\tau_{a,0}}{\Delta_s} \sum_{r \in \mathcal{R}_a} f^{rod}_{k=1} \Delta_h, \quad \forall a \in \mathcal{A}. \tag{5.20}$$

The initial travel time $\tau_{a,0}$ is defined by

$$\tau_{a,0} = \widetilde{\tau}_a + \widehat{\tau}_a, \quad \forall a \in \mathcal{A}. \tag{5.21}$$

where $\widetilde{\tau}$ is the free flow travel time (s) and $\widehat{\tau}$ the extra travel time (s) due to other traffic. For the extra travel time component the second term of equation 5.22 is used. In the next section this equation and the other travel time functions are discussed.

5.4   Travel time functions

In the propagation of traffic, travel time functions are used to move traffic through the network. There is a large literature available on the theoretical and practical aspects of these functions for all sorts of links: controlled and uncontrolled. A good overview for controlled links is given

in Rouphail *et al.* (2002) and Viti (2006). For uncontrolled links the state-of-the-art report by Troutbeck and Brilon (2002) is a good source to start with.

In this thesis the focus will be on the description of the travel time functions used and not on the discussions which functions to choose or how to derive such a function. Four link types are distinguished: normal links ($\mathcal{A}^0 \subseteq \mathcal{A}$), controlled links ($\mathcal{A}^c \subset \mathcal{A}$, signal control or ramp metering), roundabout links ($\mathcal{A}^r \subset \mathcal{A}$) and priority links ($\mathcal{A}^p \subset \mathcal{A}$). In the next sections the choices for functions for these different link types are described. For all link types the calculated travel time is bounded with a maximum travel time. This maximum travel time is derived from a minimum speed $\vartheta_a^{\min}$, which is link specific.

5.4.1 Normal links

Normal links have no special characteristics, only the usual ones: length, number of lanes, saturation flow and free speed. The capacity at the end of these links Q_a'' is fixed and equal to Q_a'. The travel time function for these links has been derived from Akçelik (1991, 2003), because the Akçelik function has a good theoretical basis and is consistent with data, while the HCM function for this type of links has some problems for over-saturated condition (Akçelik, 2003). The Akçelik function uses Davidson's function (Davidson, 1966, 1978) as a basis and with the coordinate transformation technique (Akçelik, 1989) it becomes suitable for over-saturated conditions. The coordinate transformation technique, when applied to a steady-state curve derived from standard queuing theory, produces a time-dependent formula for delays. Steady-state delay models are asymptotic to the y-axis (i.e generate infinite delays) at a degree of saturation of 1. The coordinate transformation method shifts the original steady-state curve to become asymptotic to the deterministic over-saturation delay line. This is shown in figure 5.3.
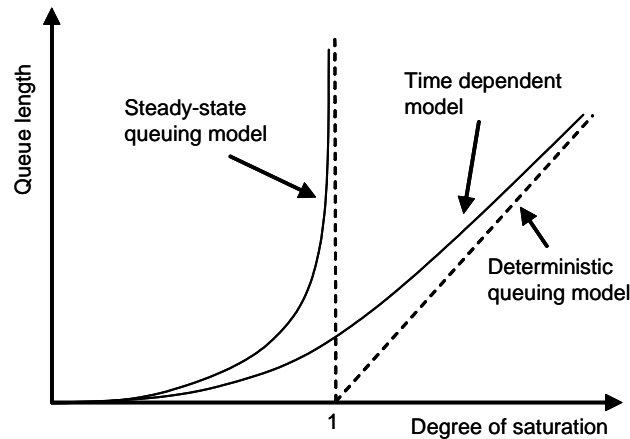
Figure 5.3: Principle of coordinate transformation technique

The coordinate transformation method makes the calculation of delay dependent on the analysis time period. This can be seen for example in equations 5.22 and 5.27, where the variable Δ_h is the length of the time period under consideration.

The Akçelik function calculates the travel time τ_{at} (s) as follows

$$\tau_{at} = \widetilde{\tau}_a + 0.9 l_a \Delta_h \left(z_{at} + \sqrt{z_{at}^2 + \frac{8 k_a \varphi_{at}}{Q_a'' \Delta_h} + \frac{16 k_a \kappa_{at}}{(Q_a'' \Delta_h)^2}} \right), \quad \forall a \in \mathcal{A}^0, t, \tag{5.22}$$

where $\widetilde{\tau}_a$ is the free flow travel time (s), l_a is the length of the link (m), Δ_h is the length of the analysis time period (h), φ_{at} is the degree of saturation defined by equation 5.3, κ_{at} is the length of the queue in vehicles defined by equation 5.17, z_{at} is a parameter, related to the initial queue and calculated as

$$z_{at} = \varphi_{at} - 1 + \frac{2 \kappa_{at}}{Q_a'' \Delta_h}, \quad \forall a \in \mathcal{A}^0, t, \tag{5.23}$$

and k_{at} is a link dependent parameter, which is related to other parameters through

$$k_a = \frac{2 Q_a'' (\frac{\vartheta_a^f}{\vartheta_a^c} - 1)^2}{\Delta_h (\vartheta_a^f)^2}, \quad \forall a \in \mathcal{A}^0, \tag{5.24}$$

where ϑ_a^f is the free flow speed (km/h) for link a and ϑ_a^c is the speed (km/h) at which free flow turns into congestion. It is obvious that $\widetilde{\tau}_a = 3.6 l_a / \vartheta_a^f$.

The speed at congestion ϑ_a^c is an input parameter and can be different for different free flow speeds. For a link with a length of 1000 meters, a capacity of 2000 veh/h, no initial queue and using a time period of 15 minutes, the travel times were calculated as a function of the degree of saturation φ and for different combinations of free speeds and speeds at congestion. These travel times are shown figure 5.4.
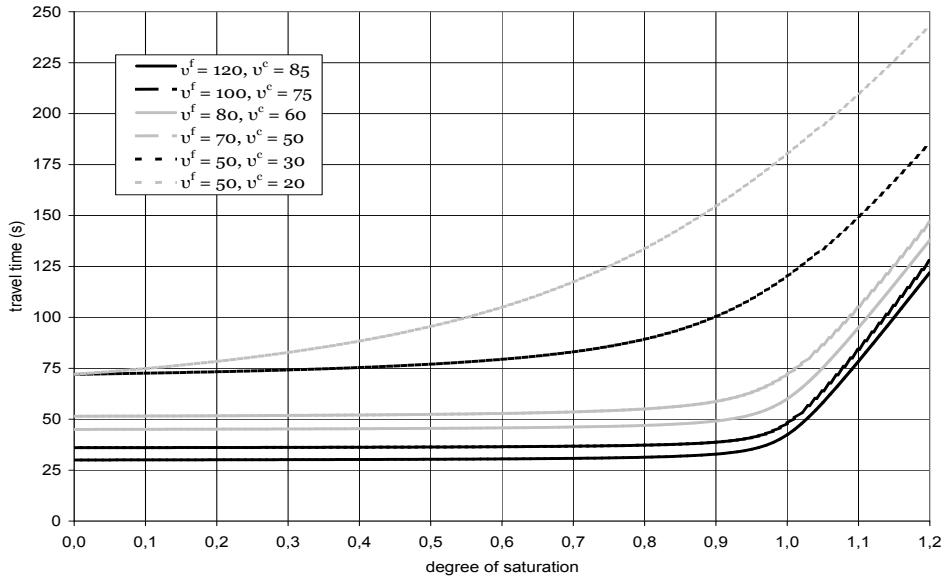
Figure 5.4: Travel times for combinations of free speed and congestion speed

5.4.2   Controlled links

The travel time function for controlled links (links ending in a controlled intersection) is derived from TRB (2000). In his PhD thesis, Viti (2006) shows that this function underestimates the delay, because it does not take into account random fluctuations in the arrival pattern and in the queue process. Nevertheless, we use the HCM2000 function, because at the time of developing the framework and running the simulations, the Van Zuylen-Viti function (Viti, 2006) was not available and presumably the results will not change that much.

The HCM2000 function (which is based on the work of Akçelik) for the travel time τ_{at} (s) is split into four parts

$$\tau_{at} = \widetilde{\tau}_a + d'_{at} + d''_{at} + d'''_{at}, \quad \forall a \in \mathcal{A}^c, t, \tag{5.25}$$

where $\widetilde{\tau}_a$ is again the free flow travel time, d' is the so-called uniform delay due to the signal, d'' the overflow delay due to stochastic variations in the demand and oversaturated conditions and d''' the delay due to the presence of an initial queue. The first part of the delay is defined as

$$d'_{at} = \frac{C_{ak}(1 - \lambda_{ak})^2}{2(1 - \min(1, \varphi_{at})\lambda_{ak})}, \quad \forall a \in \mathcal{A}^c, t, \tag{5.26}$$

where C_{ak} is the cycle time of the signal connected to link a and λ_{ak} is the ratio of the green time g_{ak} and cycle time C_{ak}. Note that the signal timings are changing per time period and not per time step. The second part of the delay is given by

$$d''_{at} = 900\Delta_h \left(\varphi_{at} - 1 + \sqrt{(\varphi_{at} - 1)^2 + \frac{8I'_a I''_a \varphi_{at}}{Q''_{ak}\Delta_h}} \right), \quad \forall a \in \mathcal{A}^c, t, \tag{5.27}$$

where I'_a is a parameter for a given arrival and service distribution (e.g. 0.5 for fixed-time signal control) and I''_a is a parameter for the ration of variance to mean of arrivals from upstream signals (e.g. 1.0 for Poisson arrivals). The capacity at the end of the link Q''_{ak} is given by $\lambda_{ak}Q'_a$. The third part of the delay is specified by

$$d'''_{at} = \frac{1800\kappa_{a,t-1}(1 + \nu_{at})\Delta'_h}{Q''_{ak}\Delta_h}, \quad \forall a \in \mathcal{A}^c, t, \tag{5.28}$$

where ν_{at} is a parameter defined by

$$\nu_{at} = \max\left(0, 1 - \frac{Q''_{ak}}{\kappa_{a,t-1}}(1 - \min(1, \varphi_{at})) \right), \quad \forall a \in \mathcal{A}^c, t, \tag{5.29}$$

and Δ'_h is the period of unmet demand in Δ_h, which is given by

$$\Delta'_h = \min\left(\Delta_h, \frac{\kappa_{a,t-1}}{Q''_{ak}(1 - \min(0.99, \varphi_{at}))} \right). \tag{5.30}$$

In these equations $\kappa_{a,t-1}$ is the length of the queue (veh) for link a and time step $t-1$. In the denominator of equation 5.30 the term $1 - \min(0.99, \varphi_{at})$ is used, instead of $1 - \min(1, \varphi_{at})$, to avoid possible division by zero.

Viti (2006) shows that a good estimation of the overflow delay, given by equation 5.28, is dependent of a good estimation of the overflow queue $\kappa_{a,t-1}$. His estimation consists of a linear and an exponential part, whereas the HCM2000 formulas only contain a linear. In our work the overflow queue is estimated with the inflow and outflow on a link (equation 5.17) and therefore, a separate estimation formula is not needed. But Viti shows that a simple flow balance is not enough, if the effects of stochastics on the expectation of the queue length are taken into account. Therefore, an improvement on this point is possible.

5.4.3 Roundabout links

For roundabout links a, belonging to the set of links $\mathcal{A}^r$ (links ending at a roundabout), the calculation of travel times is divided into two parts: the determination of the capacity of the link and, based on that capacity, the calculation of the travel time. The capacity of a link ending in a roundabout is dependent on traffic on the link and on the circulating traffic on the roundabout. Other external influences such as bicycles and pedestrians are not taken into account. The estimation of the capacity Q''_{at} (veh/h) is done with a formula derived by Wu (2001):

$$Q''_{at} = 3600 \frac{p_a}{\rho_{r3}} \left(1 - \rho_{r1} \frac{u_{ct}}{p_c} \right)^{p_c} e^{-u_{ct}(\rho_{r2} - \rho_{r1})}, \quad \forall a \in \mathcal{A}^r, t. \tag{5.31}$$

In this formula u_{ct} is the traffic flow in the circulating lanes (veh/s) for time step t, p_c is the number of circulating lanes, p_a the number of lanes on the link and ρ_{r1}, ρ_{r2} and ρ_{r3} are parameters (s), respectively the minimum gap, the zero gap and the move-up time. Between the parameters ρ_{r2} and ρ_{r3} a relation exists, which can be expressed as $\rho_{r2} = \rho_{r4} - \rho_{r3}/2$, where ρ_{r4} is another parameter, the critical gap. Details about the derivation of the formula and about the interpretation of the parameters can be found in Wu (2001).

The capacity of a link, approaching a roundabout, Q''_a, as a function of the circulating traffic flow u_c and for different lane configurations, is shown in figure 5.5. In this figure $p_a/p_c = 2/1$ means 2 lanes on the approaching link and 1 circulating lane on the roundabout itself. From the figure it is clear that the amount of circulating traffic on the roundabout has a large impact on the capacity of the link approaching the roundabout.

Based on the estimated capacity, the travel time can be calculated. The travel time function has been derived from Troutbeck and Brilon (2002) and is as follows

$$\tau_{at} = \widetilde{\tau}_a + \frac{1800(2 + \kappa_{a,t-1})}{Q''_{at}} + \frac{\Delta_s}{4} \left(\varphi_{at} - 1 + \sqrt{ \left(\frac{7200\kappa_{a,t-1}}{Q''_{at}\Delta_s} + \varphi_{at} - 1 \right)^2 + \frac{8\varphi_{at}}{Q''_{at}\Delta_s} } \right), \tag{5.32}$$

where all variables have the same meaning as in previous equations.

For different combinations of capacity Q''_a and initial queues κ_a, the travel times, for a link of 500 meters, a free speed of 50 km/h and a time period of 900 seconds, are shown in figure 5.6.
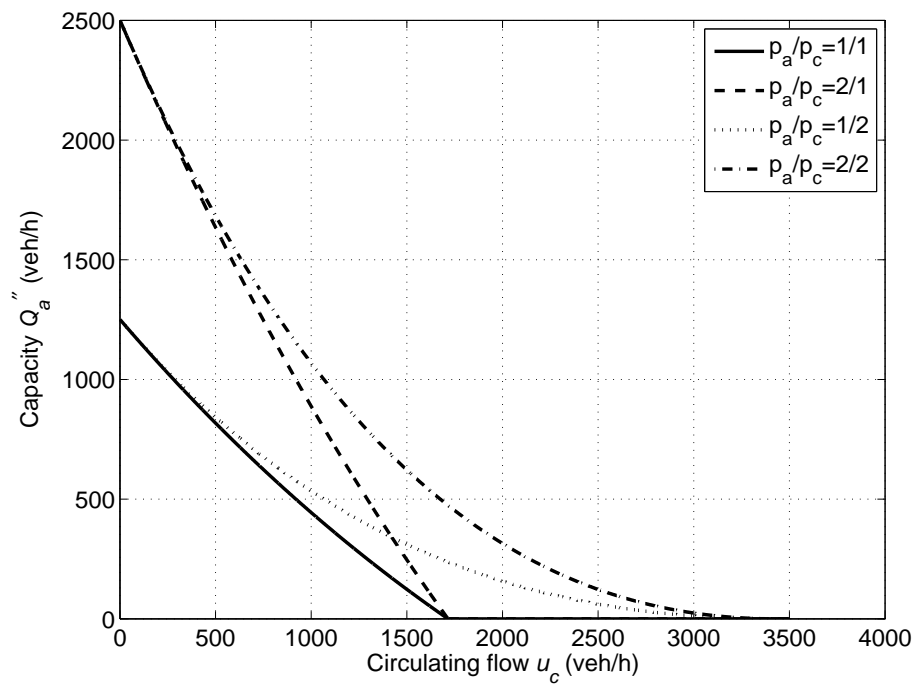
Figure 5.5: Capacity roundabout link
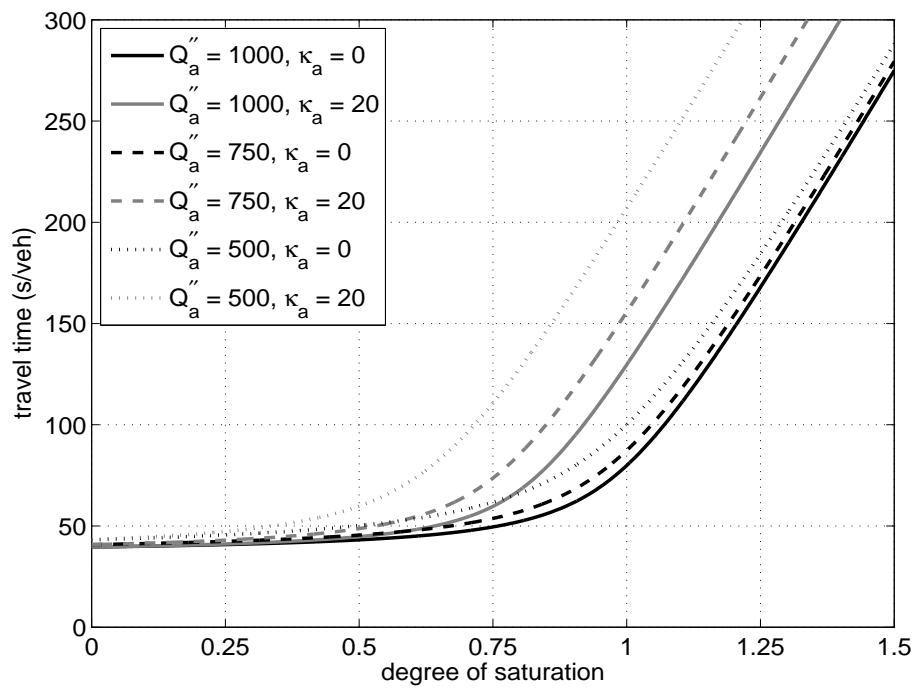
Figure 5.6: Travel time functions for roundabout links

5.4.4   Priority links

For priority links (links ending at a priority junction) the same parts as for roundabout links are necessary: estimate capacity and calculate the travel time. Again, the estimation of the capacity is based on formulas derived by Wu (2001). The capacity Q''_a (veh/h) of a minor link a, belonging to the set $\mathcal{A}^p$ (links that has to give way) conflicting with m major links (links with right of way) is defined by

$$Q''_{at} = \frac{3600 p_a}{\rho_{p3} \left(\sum\limits_{j=1}^{m} \frac{1-\varphi_{jt}}{\varphi_{jt}} \right)} \prod_{j=1}^{m} (1 - \rho_{p1} u_{jt}) \prod_{j=1}^{m} e^{-u_{jt}(\rho_{p2}-\rho_{p1})}, \quad \forall a \in \mathcal{A}^p, \, t, \tag{5.33}$$

where u_{jt} is the flow on major link j and time step t (veh/s) and ρ_{p1}, ρ_{p2} and ρ_{p3} are parameters (s), respectively the minimum gap, the zero gap and the move-up time. Again there is a relation between the parameters ρ_{p2} and ρ_{p3}, which can be expressed as $\rho_{p2} = \rho_{p4} - \rho_{p3}/2$, where ρ_{p4} again is the critical gap (s). All parameters have the same meaning as in equation 5.31. The capacity Q''_a of a minor link a, conflicting with major traffic links with flow u_c, is shown in figure 5.7.
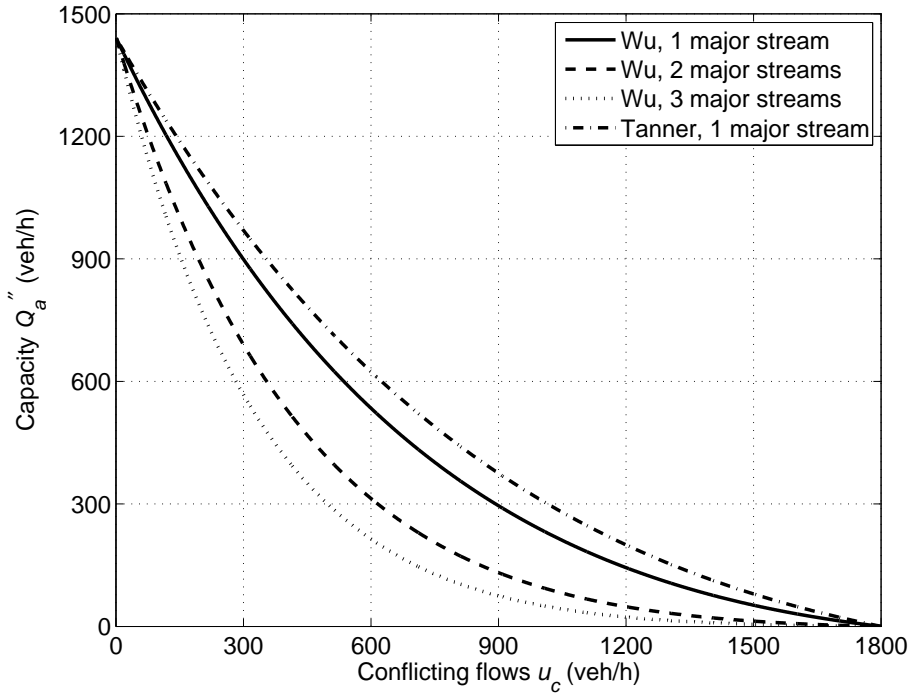
Figure 5.7: Capacity of minor stream

The figure shows the capacity calculated for the minor link if there is only one major conflicting link (Wu, 1 major stream), if there are two links which have right of way (Wu, 2 major streams) and if there are three major links (Wu, 3 major streams). In the figure also the capacity according to the well-known theory of Tanner (1962) is given. This capacity formula predicts somewhat

higher values and is given by

$$Q''_{at} = 3600 p_a (1 - \rho_{p1} u_{jt}) \frac{u_{jt} e^{-u_{jt}(\rho_{p2} - \rho_{p1})}}{1 - e^{-u_{jt}\rho_{p3}}}, \quad \forall a \in \mathcal{A}^P, t. \tag{5.34}$$

We choose the approach from Wu, because it is more general than the approach from Tanner (more conflicting streams possible). The travel time for priority links is the same as for roundabout links (equation 5.32), but the capacity used in this formula will come from equation 5.33.

5.5   Route travel time

Because the assignment is route based, the link travel times have to be transferred to route travel times (or route travel costs, if other components are added, e.g. tolling). For this, a kind of trajectory method is used. For all (or a selection of) time steps and routes, traffic is followed through the network and travel times of the visited links are summed to a route travel time. In his thesis De Romph (1994) describes several methods to do this. These methods are developed to overcome the problem of radical changes in travel time in successive periods (later departure, earlier arrival) and thus to meet the first-in-first-out (FIFO) condition. Because in our DNL model the length of the time step is small (typically 5-30 s), compared to the length of the time period in De Romph (1994) (typically 5-15 minutes), and thus link travel times do not differ much from time step to time step, we assume that the difference between these methods is small and probably irrelevant. To underpin this assumption, the difference between two of these methods was investigated to see whether it is relevant in our work.

The first method is the simplest one and does not obey the FIFO constraint. The route travel time for a departure time step t is given as

$$c_t^{rod} = \sum_{a \in \Gamma_r} \tau_{at'}, \quad \forall o, d, r \in \mathfrak{R}^{od}, t, \tag{5.35}$$

$$t' = t + \left\lfloor \frac{c_{a-t}^{rod}}{\Delta_s} \right\rfloor, \tag{5.36}$$

where Γ_r is de set of links for route r, t' is another index for the time step, $\tau_{at'}$ is the travel time on link a for that time step, c_{a-t}^{rod} is the route travel time up to, but not including link a, Δ_s is the length of a time step and $\lfloor x \rfloor$ is the floor function, which returns the largest integer less than or equal to x. The second method replaces equation 5.35 with

$$c_t^{rod} = \sum_{a \in \Gamma_r} \min_\lambda (\lambda \Delta_s + \tau_{a,t'+\lambda}), \quad \forall o, d, r \in \mathfrak{R}^{od}, t, \tag{5.37}$$

where $\lambda \geq 0$ is an integer. This equation checks if there are future time steps with lower travel times. With this method the FIFO condition is satisfied.

For the comparison of both methods, the network of Amsterdam around the A10-West motorway was used (shown in figure 6.8). For this network the travel times for 8889 routes and 10 time periods were calculated with both methods. Only 143 of the 88890 values (0.16%) showed a difference and only 57 of these differences (0.06%) was larger than 1 second. The maximum difference was 4.2 seconds on a travel time of 315 seconds. Also, most of the differences were found in the first and last period: the start and the end of the congestion. Because of the small differences in both methods, and for its simplicity, the first method was chosen. Although the second method is theoretically better, because it obeys the FIFO principle, the time to calculate the route travel times increases and the results are not so different.

To obtain the route travel time per time period k from the route travel time per time step, simple averaging is used, as specified in:

$$c_k^{rod} = \frac{1}{T} \sum_{t=(k-1)T}^{kT} c_t^{rod}, \quad \forall o,d,r \in \Re^{od}, k,t, \tag{5.38}$$

where T is the number of time steps per time period. Note that the length of the time step Δ_s should be a divisor of the length of the time period. To decrease the time needed for the calculation of route travel times, the time period, which has a length of $\Delta_s T$, is divided into a number of parts with a length $\delta\Delta_s$, where δ is an integer. Now, c_t^{rod} is not calculated for all time steps t with an interval of Δ_s, but for those t with an interval of $\delta\Delta_s$.

5.6 Critical links

So far, we have assumed that the length of the time step is smaller than the shortest free flow travel time that is needed to traverse any link: $\Delta_s < \min_a \widetilde{\tau}_a$. In realistic networks for a lot of small links this is not the case. To fulfil the requirement two options are possible: decrease the length of the time step or increase the length of these critical links (links with a free flow travel time smaller than the length of the time step). The first option is not always desirable, because the time step could become too small for an acceptable calculation time. The second option has the disadvantage that lengthening links leads to more vehicle kilometres, more congestion and more delay than expected, especially if the length of the time step increases. It was shown by Versteegt *et al.* (2003) that an increase of the time step from 10 to 30 seconds and lengthening the links, led to an increase of 13% in veh.km and 49% in total delay for the Dutch national network.

In this model the second option was chosen, with the assumption that traffic traverses critical links in one time step. To minimise the problems of this option, for every time step the critical links are determined and virtually lengthened. So, the set of critical links can be different for every time step and is defined by

$$\mathcal{A}_t^{crit} = \{a \in \mathcal{A} | \tau_{at} < \Delta_s\}. \tag{5.39}$$

Note that not the free flow travel time $\widetilde{\tau}_{at}$ is used, but the real travel time τ_{at}.

For the set of critical links some adjustments are needed in the propagation of traffic on these links. In next sections these adjustments are described and some tests are done to show the effects of using short links.

5.6.1 Adjustments for critical links

The first adjustment is related to the outflow. For critical links it is assumed that flow travels the link in one time step and without any delay. Therefore, the outflow (equation 5.4) becomes

$$v_{at} = u_{a,t-1} + \kappa_{a,t-1}, \quad \forall a \in \mathcal{A}_t^{crit}, \, t, \tag{5.40}$$

with the same constraints as before

$$v_{at} \ \leq \ Q_{ak}'' \Delta_h, \tag{5.41a}$$
$$v_{at} \ \geq \ 0 \tag{5.41b}$$

Of course, the outflow has to be adjusted in the same way as for normal links (equations 5.12 until 5.16). Also, the calculation of the storage capacity ψ (veh) of a link is different for critical links, but very important due to the effects on the onset of congestion and the blocking back mechanism. Because the real length of the link gives problems handling the queue, it is assumed that for critical links the available space is equal to a certain minimum length. It is calculated with

$$\psi_{at} = \frac{l_a^{crit} \, p_a}{\bar{l}^{veh}} - \chi_{at}, \quad \forall a \in \mathcal{A}_t^{crit}, \, t, \tag{5.42}$$

where p_a is the number of lanes and l_a^{crit} is the critical length of link a, which is defined as

$$l_a^{crit} = \max(l_a, \Delta_h \vartheta_a^f), \quad \forall a \in \mathcal{A}_t^{crit}, \tag{5.43}$$

where ϑ_a^f (km/h) is the free speed on the link a. This correction is needed to make the flow propagate downstream correctly. To make sure that congestion downstream is propagated upstream, the available space is made equal to the corrected outflow from the previous time step, so

$$\psi_{at} = v_{a,t-1}', \quad \{\forall a \in \mathcal{A}_t^{crit} | v_{a,t-1} - v_{a,t-1}' > 0\}, \forall t. \tag{5.44}$$

The condition in this equation is the trigger for congestion downstream.

A disadvantage of making a link longer is that it takes several time steps to fill the link instead of the expected one time step or less. Therefore, the queue builds up slower than for normal links. To deal with this problem, the propagation of congestion upstream is done in one time

step. Also, if a queue disappears, the queue on the link must be cleared in one step. Therefore, the formula for the queue on a critical link, if congestion occurs or disappears, is

$$\kappa_{at} = \begin{cases} \kappa_{at} \leq \frac{l_a^{crit} p_a}{\bar{l}^{veh}} - \psi_{at} & v_{at} > v'_{at} \\ 0 & v_{at} < v'_{at} \end{cases}, \quad \forall a \in \mathcal{A}_t^{crit}, t, \tag{5.45}$$

where l_a^{crit} is defined above. If there is no congestion, equation 5.17 also holds for critical links. But the constraint 5.18a is a little bit different. It becomes

$$\kappa_{at} \leq \frac{l_a^{crit} p_a}{\bar{l}^{veh}}, \quad \forall a \in \mathcal{A}_t^{crit}, t. \tag{5.46}$$

The link travel times for the critical links are calculated in the same way as for normal links. For the route travel times also the link travel times for the critical links are taken into account. Equations 5.35 and 5.36 can be used in the same way.

5.6.2   Some tests

To test the adjustments for critical links, the network shown in figure 5.8 was used. The network consists of five equal links with a length of 1000 meters, 2 lanes, a capacity of 4000 veh/h and a free speed of 100 km/h. Link 5 is the bottleneck link with only 1 lane and a capacity of 2000 veh/h. Traffic travels from left to right.

Figure 5.8: Normal network

The simulation period is 90 minutes divided into 18 periods of 5 minutes and starts at 07:00 hours. For the first quarter of an hour the demand is 1000 veh/h. Then the demand increases to 3000 veh/h for half an hour and the rest of the period the demand is 1000 veh/h again. The time step used is 10 seconds, which implies, using equation 5.43, a critical link length of ±278 meters. The results for this network and demand profile are shown in figure 5.9.

Link inflows   Link speeds
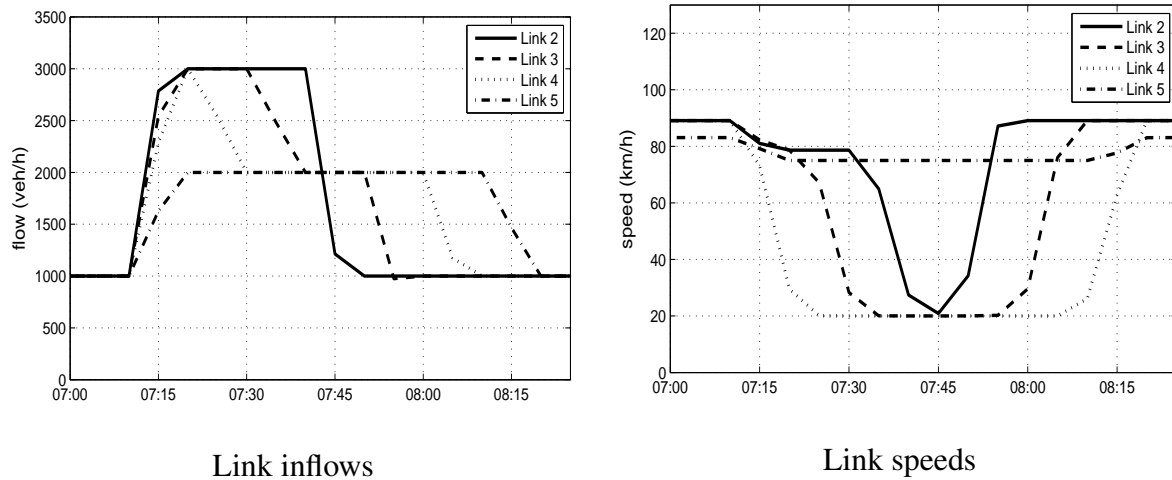
Figure 5.9: Link inflows and speeds for normal network

The figure shows the average inflows and average link speeds for the 5 minute periods as a point at the beginning of the period. From this figure it is clear that the congestion starts at link 4 and moves upstream to halfway link 2. The inflow of link 2 is not affected. The speed of link 5 for uncongested conditions is lower than for the other links, because there is only one lane and vehicles have more interaction. If we change the network and split link 2 in two parts, one part with a length of 50 meters and one with a length of 950 meters, we obtain the network shown in figure 5.10.

Figure 5.10: Adjusted network 1 with one critical link

The results in terms of link inflows and speeds for this network are shown in figure 5.11. In this figure link 2 is small part of the split up 'old' link 2. The results show that there is no difference between the link inflows. The link speed differs only for link 2, because now the congestion occupies link 21 and does not reach link 2. The lower speed for link 2 for a part of the period is due to the increase of the flow from 1000 to 3000 veh/h.
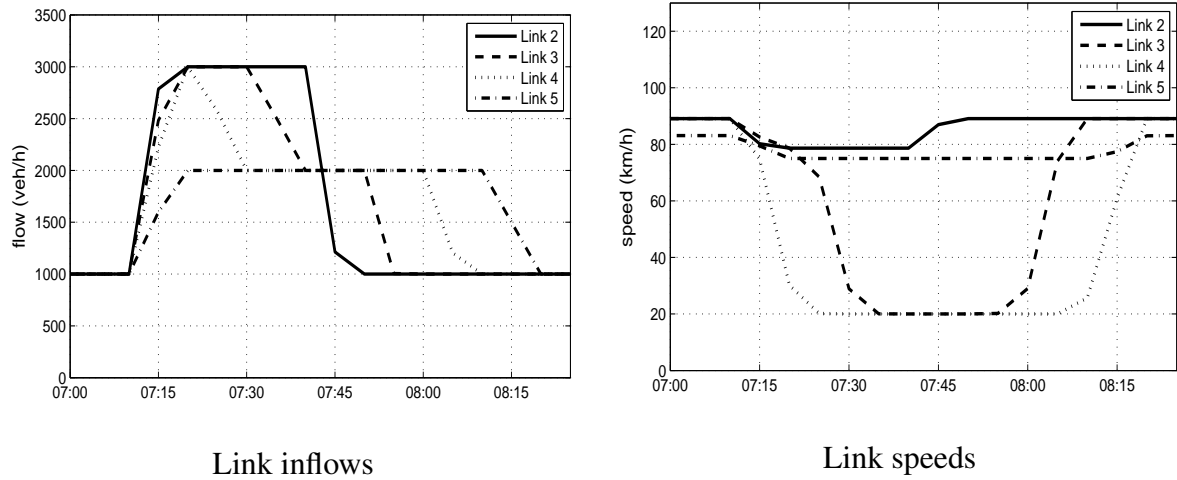
Figure 5.11: Link inflows and speeds for adjusted network 1

Again the network is changed. Link 4 is split into two parts. The resulting links 4 and 41 have a length of 30 meters and 970 meters respectively. The network is shown in 5.12 and the results for this network are shown in figure 5.13.

Figure 5.12: Adjusted network 2 with two critical links
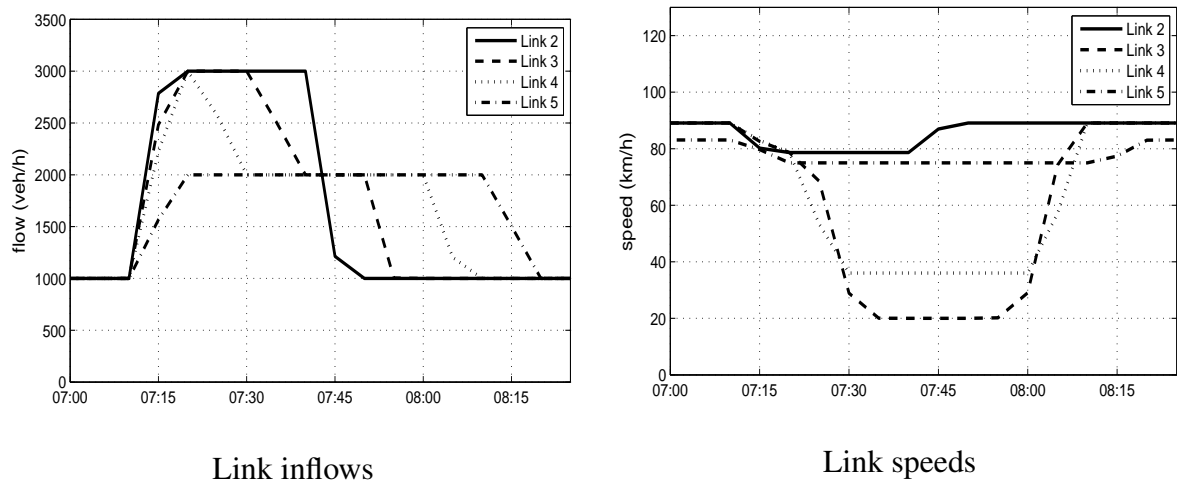
Figure 5.13: Link inflows and speeds for adjusted network 2

These results show that again there is no difference between the link inflows. For the link speeds there is now a difference for short link 4, which is caused by the length of the link. The critical

link length is too short to contain a queue long enough to result in minimum speed. If the queues in the three situations are analysed, they all have more or less the same length of 2650 metres, which is of course what we want. This also leads to remarkable corresponding travel times, which is shown in figure 5.14.
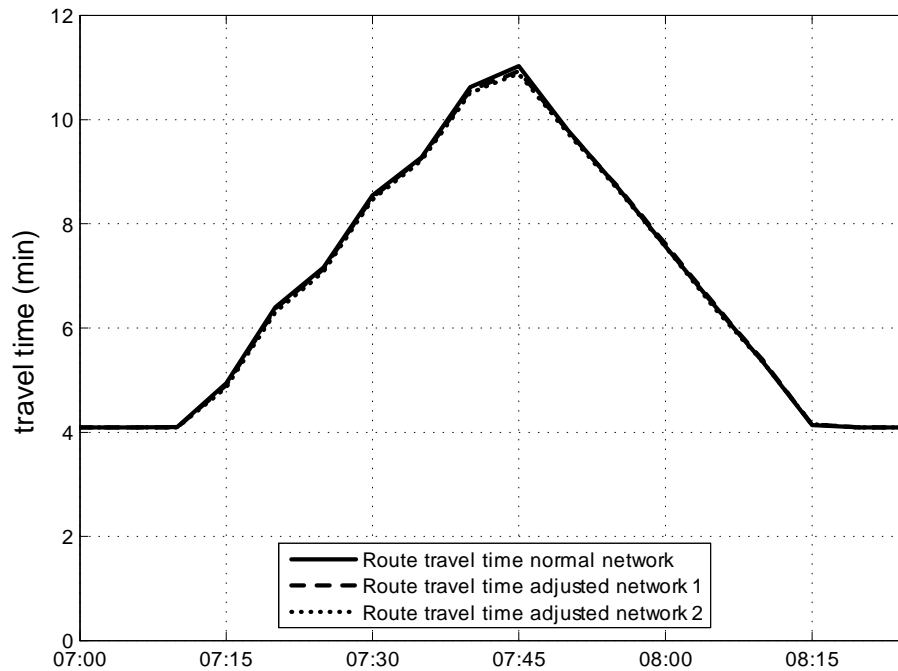
Figure 5.14: Route travel times for the test networks

5.7 Summary

In this chapter the dynamic network loading (DNL) model is described. The DNL model uses advanced travel time functions to propagate traffic through the network. New in this model is the combination of different travel time functions for different link types: normal links, signal controlled links, roundabout links and priority links. This makes it possible to simulate real-life networks with reasonable accuracy. The travel time is used to determine the outflow of links and with that the inflow of downstream links. Because of the assumption of a uniform distribution of traffic over the length of a link, differences in traffic flow at the beginning of a link are damped before they reach the end of the link, so it takes too much time before congestion starts or dissolves. To overcome this problem a new "step" mechanism is introduced, which makes sure that differences in inflow reach the end of the link in time. Horizontal queuing and blocking back is modelled by the concept of 'available space'. The available space on a link determines how much traffic can enter the link and thus how much traffic is held back on the upstream links. This blocked traffic is distributed among the upstream links according to the number of lanes.

At decision nodes traffic is distributed from the incoming to the outgoing links according to the splitting rates, which are calculated from the route flows using free flow travel times. A possible improvement would be to use real travel times, but this would imply a loop within the simulation, with the accompanying calculation time and convergence problems involved. Congestion is always caused by a capacity restriction and the resulting queue propagates upstream and horizontal, which means that blocking back is taken into account. The route travel times are calculated from the link travel times using a trajectory method. With this method the FIFO condition is not guaranteed, but that seems to give no problems, at least not in the applications we have done so far.

A special feature of the model is the treatment of critical links. This is different and new compared to other models. Where other models use the free flow travel time as the criterion to determine the critical links, in our model the travel time itself is used and the determination of critical links and the adjustments for these links are done for every time step. Tests show that this way of handling critical links gives good results.

6

Model Calibration and Validation

In the previous chapters a framework for anticipatory control has been developed. It consists of three parts. The control part was discussed in chapter 3, the assignment algorithms have been specified in chapter 4 and the dynamic network loading model was described in chapter 5. Before we use the framework for our own purpose, it is necessary to analyse whether the model gives a good representation of reality. Therefore, the model is calibrated and validated for a small and a larger real-life network. For the small network, a closer look at some model parameters is taken and for the larger network also important input for the model (the OD matrix) is the topic of discussion. For convenience, the complete model is named MARPLE, which is an abbreviation of "Model for Assignment and Regional PoLicy Evaluation ".

Every model has parameters, which have to be specified. Some of these parameters are dependent on the situation for which the model is applied, other parameters are related to generic processes of the model itself. Calibration is the process of comparing the outcomes of the model with the results of observations in reality, and changing the model parameters in such a way that the outcome of the model fits the measurements. Although, the outcome of a model will never reproduce reality perfectly, parameters can be chosen in such a way that a certain fitness criterion is optimised.

Validation is the process of comparing the outcomes of the model with another dataset, using the parameters as found in the calibration. Its purpose is to verify whether the model parameters chosen are valid for other situations and in this way increase confidence in the model.

This chapter describes this calibration and validation process for MARPLE and also discusses the related problems. First, a situation with a motorway stretch is studied. In this small network no route choice is possible, so only the DNL model is tested. The data of one day is used to calibrate some of the model parameters and datasets of two other days are used to show how

the model performs. For the second case, a large real-life network, use could be made of the input for a simulation study and of data collected for the evaluation of roadworks on a part of the A10 ring road around Amsterdam. Part of the measurements was used to calibrate model parameters and the OD matrix. The other part of the measurements was used for the validation. For this case the focus was more on the estimation of a good OD matrix, because an OD matrix has a large influence on the results.

Before the calibration and validation of MARPLE for these two cases are described, attention is paid to the goodness-of-fit measures, which are used to draw conclusions about the performance of the model. The chapter ends with a summary, containing conclusions.

6.1   Goodness-of-fit

To be able to quantify the quality of the model, goodness-of-fit measures are used. One of these measures is the root mean squared percent error (RMSPE). The RMSPE quantifies the overall error of the model, penalising large errors at a higher rate than smaller ones. It is defined as

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - y_i}{y_i}\right)^2},\tag{6.1}$$

where x_i is the simulated value and y_i the measured value. Another popular goodness-of-fit measure is the correlation coefficient (also called Pearson's correlation), which is defined as

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y},\tag{6.2}$$

where $\bar{x}$ and $\bar{y}$ are the average simulated and measured values and σ_x and σ_y are the standard deviations of the simulated and measured values. The square of r is the well-known R^2 measure.

According to Hourdakis *et al.* (2003) the RMSPE has an inherent deficiency in considering the disproportional weight of large errors while r, although being a good measure, does not provide any additional information to the modeller as to the nature of the error (difference) between real measurements and simulation. Theil (1961), in his work on economic forecasting, developed a goodness-of-fit measure called "Theil's Inequality Coefficient", which is more sensitive and accurate than the RMSPE or r and it can also be decomposed into three other metrics that provide specific information about the nature of the error. Theil's Inequality Coefficient is defined as

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}},\tag{6.3}$$

where U is bounded between 0 and 1: 0 means perfect fit and 1 means no correlation at all. The square of the numerator can be decomposed into three components:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 = (\bar{y} - \bar{x})^2 + (\sigma_y - \sigma_x)^2 + 2(1-r)\sigma_y\sigma_x. \qquad (6.4)$$

Based on these components and on equation 6.3 three separate measures of inequality can be defined: bias (U_m), variance (U_s) and covariance (U_c). The formulas for these three components are:

$$U_m = \frac{n(\bar{y} - \bar{x})^2}{\sum_{i=1}^{n}(y_i - x_i)^2}, \qquad (6.5)$$

$$U_s = \frac{n(\sigma_y - \sigma_x)^2}{\sum_{i=1}^{n}(y_i - x_i)^2}, \qquad (6.6)$$

$$U_c = \frac{2n(1-r)\sigma_y\sigma_x}{\sum_{i=1}^{n}(y_i - x_i)^2}. \qquad (6.7)$$

By definition $U_m + U_s + U_c = 1$. The bias reflects the systematic error, while the variance proportion indicates how well the simulation model replicates the variability in the observed data. Both should be as close to zero as possible, and in the opposite direction, the covariance proportion, which reflects the unsystematic error, should be close to 1.

These three goodness-of-fit measures are used to calibrate and validate the model. In the next section a motorway stretch with a bottleneck is investigated and after that a real-life network.

6.2   Motorway with bottleneck

To calibrate and validate the dynamic network loading model in MARPLE, discussed in chapter 5, a motorway stretch with a bottleneck was studied. For this stretch a part of the motorway A12 from Utrecht to The Hague was chosen. It is shown in figure 6.1.

This motorway has a clear bottleneck in the direction of The Hague (from right to left in the figure): a narrowing from four to three lanes, just downstream the junction with the rural road N204. Because no other junctions are in the neighbourhood, it is easy to build a network and OD matrix for this network. The motorway stretch has a total length of 13.5 kilometres, divided into 29 links. Each link starts at the position of a measured cross-section. With this configuration it is easy to compare the measured data with the simulation results. The measurements come from the motorway traffic management system, which provides minute data of flows and average
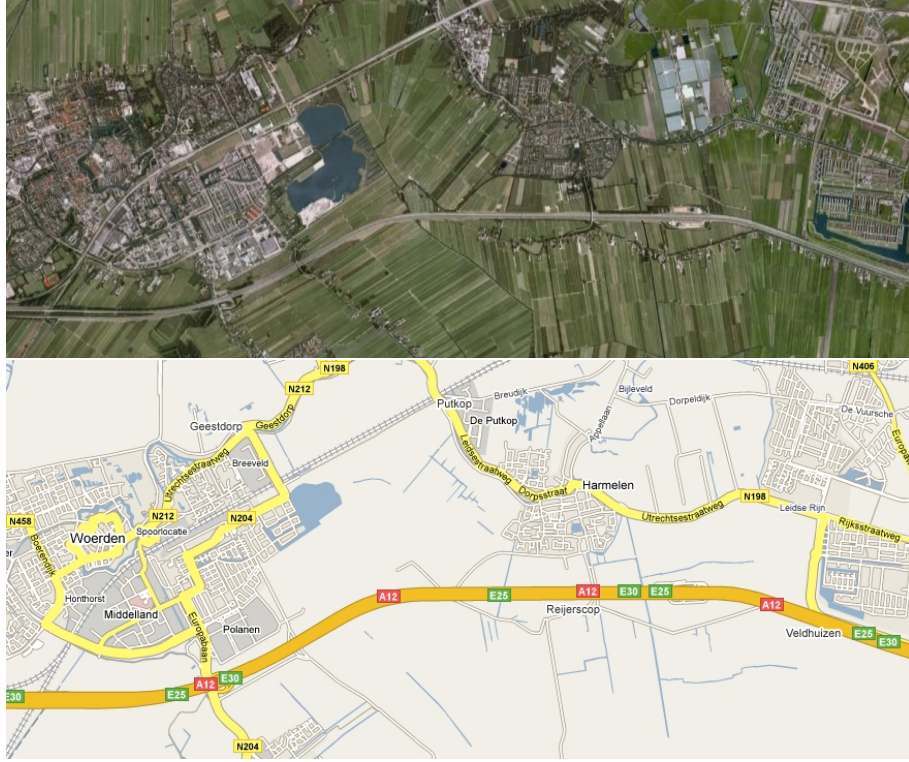
Figure 6.1: A12 motorway

speeds on cross-sections approximately every 500 meters. The data of the first cross-section and the off-ramp is used as demand for the OD matrix. The flow for the on-ramp was derived by subtracting the data of the cross-section upstream from the data of the cross-section downstream the on-ramp.

6.2.1   Calibration

For the calibration of the DNL model four parameters are important: the average length a vehicle occupies $\bar{l}^{veh}$ (including the space between vehicles), the minimum speed $\vartheta_a^{\min}$, the speed at congestion ϑ_a^c and the capacity Q' of the bottleneck. To determine the value of these parameters, the data of Wednesday April 4th, 2007, was used. The speed parameters could directly be derived from the data: $\vartheta_a^{\min} = 30$ km/h, $\vartheta_a^c = 95$ km/h. The capacity of the bottleneck Q' and the average length a vehicle occupies $\bar{l}^{veh}$ were used to calibrate the queue length and the resulting travel times. The best results for this day were obtained with $Q' = 5730$ veh/h and $\bar{l}^{veh} = 14$ metres. The capacity was determined by taking the average value of the flows just downstream the bottleneck.

Due to the nature of the DNL model (based on travel time functions and no shock waves) there is a trade-off between the representation of the travel times and the amount of congestion. In the real-life situation the speed varies in the congestion area and shock waves can be clearly recognised. In the model the speed is constant at its minimum value. This difference is clearly

displayed in the speed contour plots, shown in figure 6.2. The shock wave pattern in the measured data is not produced by the model. From the speed contour plots it can be also be seen that the congestion in the model starts earlier and takes more time to dissolve.
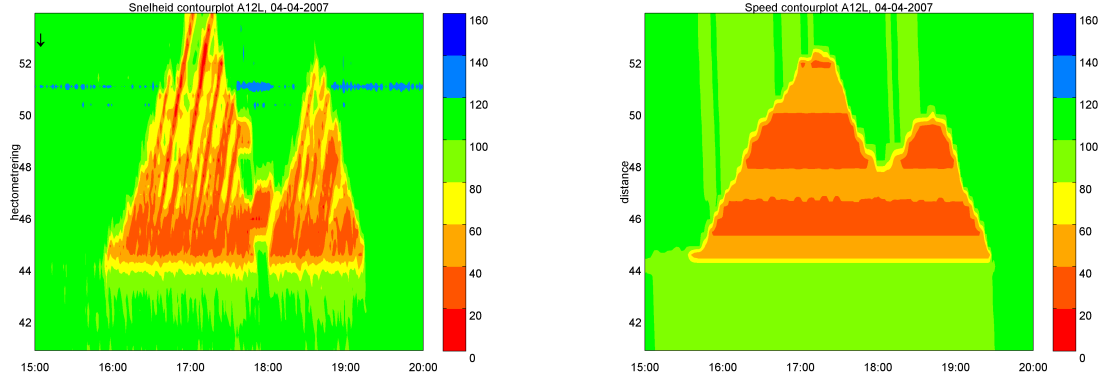
Figure 6.2: Measured and simulated speed contour plots for April 4th, 2007

From the minute data it is possible to estimate travel times with the method described by Van Lint (2004). This method estimates travel times which are close to measures ones (see Appendix D). If we compare the estimated and simulated travel times, we obtain figure 6.3. It shows us that especially at the beginning and end of the congestion period the simulated travel times are higher. This means that congestion in the model builds more gradually, as in real-life it occurs more abruptly. This is due to the shape of the travel time functions we use. If we change shape parameters for this function to get better travel times for those periods, the peak travel times get lower.
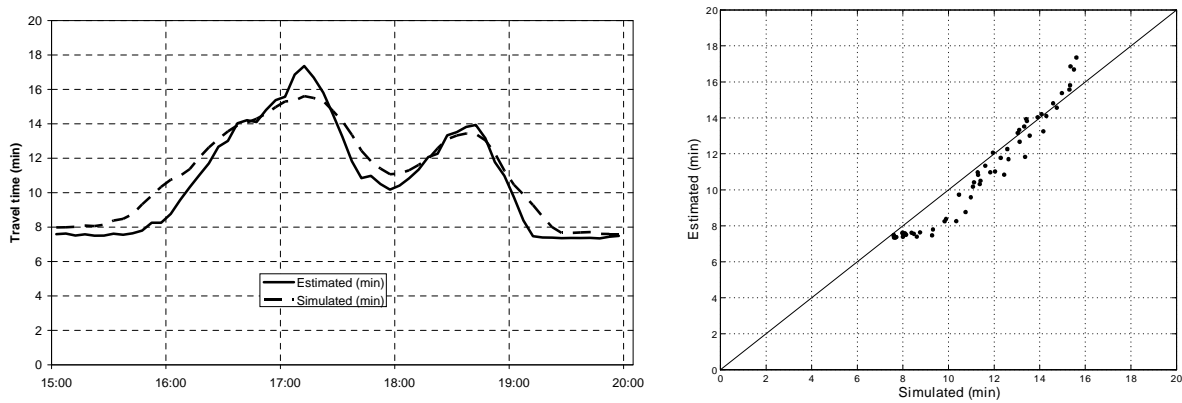
Figure 6.3: Estimated and simulated travel times for April 4th, 2007

If we calculate the goodness-of-fit measures for the flows on km. 46.206 and for the travel times, we obtain the values in table 6.1.

Table 6.1: Goodness-of-fit measures for flows and travel times on April 4th, 2007

		RMSPE	R^2	U	U_m	U_s	U_c
Flows km. 46.206	April 4th, 2007	0.102	0.759	0.046	0.011	0.071	0.935
Travel time	April 4th, 2007	0.096	0.950	0.040	0.269	0.240	0.503

Although for the travel times the RMSPE, R^2 and U are low, which means a good representation, from the U_m and the U_s values it is clear that there is a systematic error in the travel times and the travel time variance. The systematic error comes from the overestimation of the travel time in free flow conditions. The error in variance can be seen from figure 6.3: the simulated travel times show a profile which is more flat. The goodness-of-fit measures show good values for the flows on cross-section 46.206, except for the R^2 measure. This is due the absence of shock waves in the model.

Finally, we can compare some network indicators. These are shown in table 6.2. The total distance travelled is more or less equal, but the model overestimates delay with about 8%. This is due to the absence of variance in the congested area and the lower speeds in the periods without congestion.

Table 6.2: Network indicators for April 4th, 2007

	April 4th, 2007		
	measured	simulated	difference
Total distance travelled (1000 veh.km)	350.70	353.59	0.8 %
Total delay (veh.hrs)	1931.15	2078.49	7.6 %

6.2.2 Validation

Without changing the model parameters (except the capacity, because that varies from day to day) found in the previous section, two more days were simulated and analysed: March 20th and June 19th, 2007. March 20th, 2007 is not the best day to use, because of a shock wave entering the network from a downstream incident, but no "better" days were available. For those two days a capacity of the bottleneck of 5820 veh/h and 6000 veh/h was calculated respectively, also by taking the average value of the flows just downstream of the bottleneck. With these capacity values simulations were run, resulting in speed contour plots shown in the figures 6.4 and 6.5.
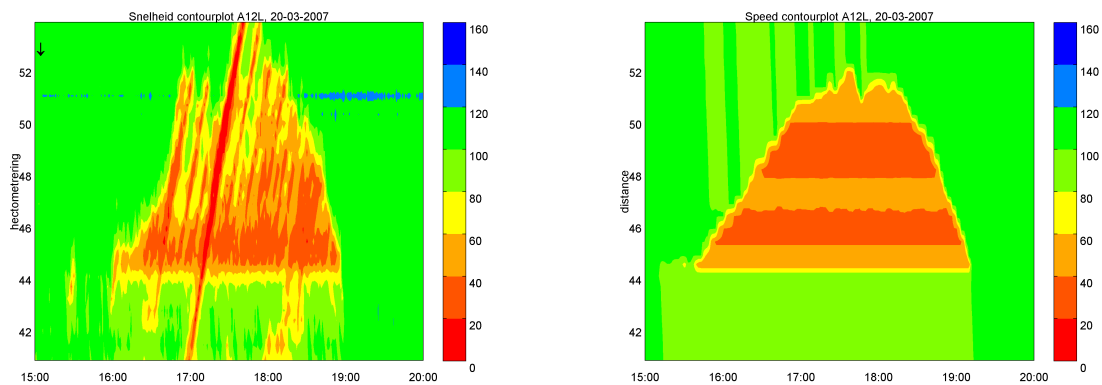
Figure 6.4: Measured and simulated speed contour plots for March 20th, 2007
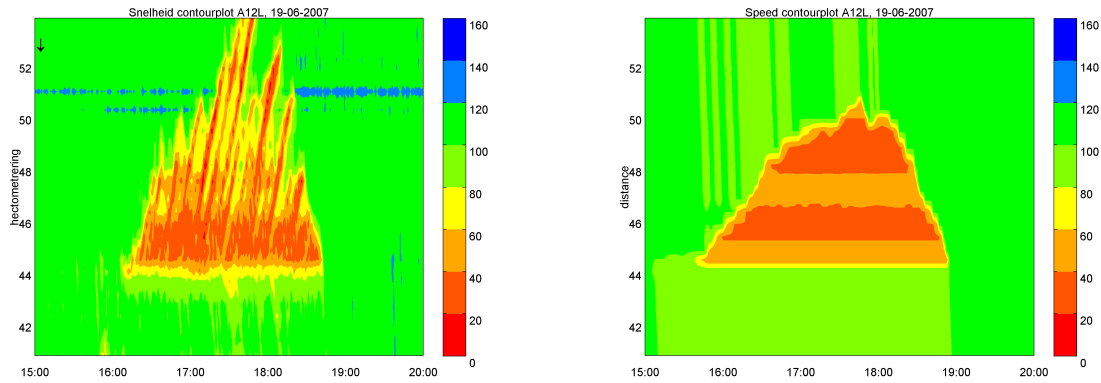
Figure 6.5: Measured and simulated speed contour plots for June 19th, 2007

From the speed contour plots it can be concluded that the model reproduces the congestion pattern reasonably well. Again it is clear that shock waves are not in the model and that the congestion starts too early and dissolves too late. If we compare the travel times, we get figures 6.6 and 6.7.
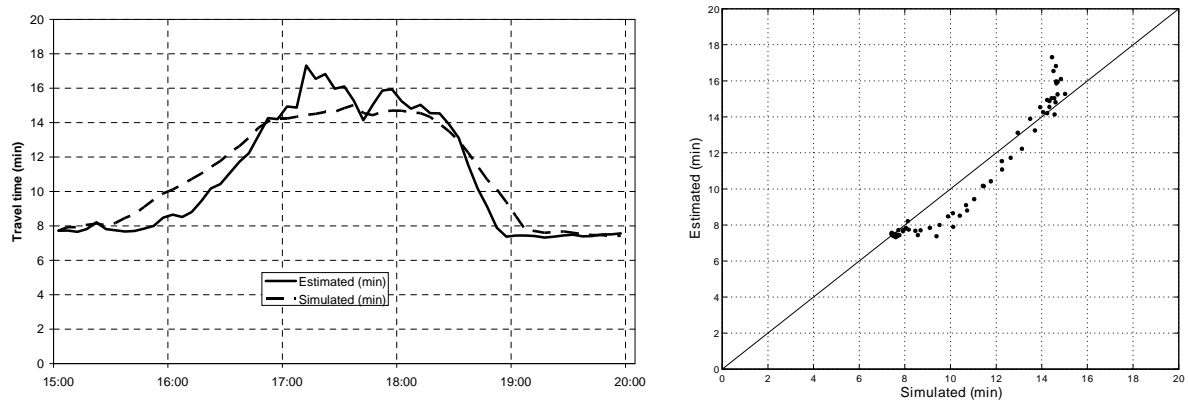
Figure 6.6: Estimated and simulated travel times for March 20th, 2007
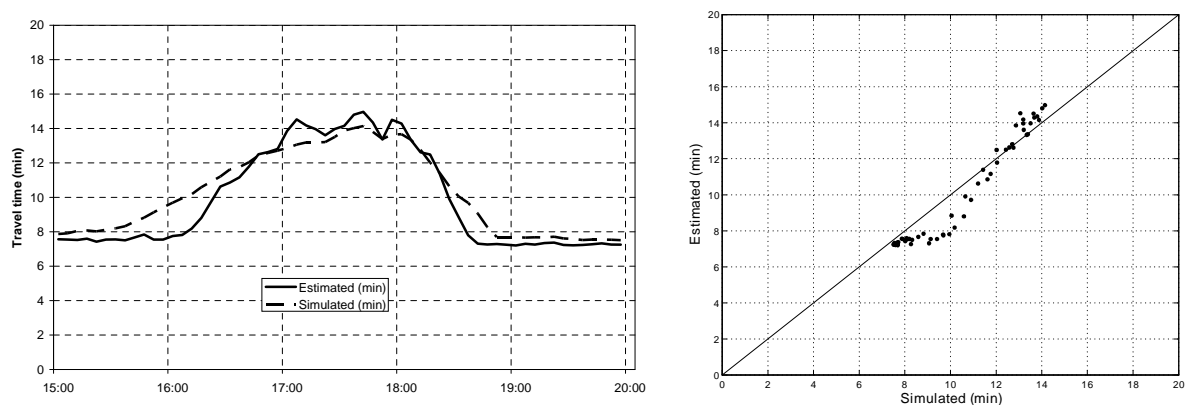
Figure 6.7: Estimated and simulated travel times for June 19th, 2007

The figures show that the modelled travel times follow the same pattern as the estimated travel times, but more smoothly. Also, the observation that in the model congestion starts earlier and

dissolves later, can be clearly seen in the travel time patterns. If we look a the goodness-of-fit errors, given in table 6.3, we see for March 20th and June 19th the same results as for April 4th: a systematic error in the travel times and the travel time variance, but the flows are looking good.

Table 6.3: Goodness-of-fit measures for flows and travel times

		RMSPE	R^2	U	U_m	U_s	U_c
Flow km. 46.206	April 4th, 2007	0.102	0.759	0.046	0.011	0.071	0.935
Travel time	April 4th, 2007	0.096	0.950	0.040	0.269	0.240	0.503
Flow km. 46.206	March 20th, 2007	0.149	0.795	0.056	0.020	0.054	0.943
Travel time	March 20th, 2007	0.105	0.928	0.047	0.047	0.324	0.645
Flow km. 46.206	June 19th, 2007	0.059	0.937	0.029	0.074	0.002	0.939
Travel time	June 19th, 2007	0.106	0.943	0.043	0.181	0.333	0.499

Finally, we again compare some network indicators. These are shown in table 6.4. Also for the two extra days the total distance travelled is simulated more or less equal to the total distance travelled which is measured. For the total delay this is different. For March 20th the total delay is underestimated in the simulation with 3%, but for June 19th the total delay is overestimated with about 7%. This is comparable with the overestimation of 8% on April 4th. Again, these differences are due to the absence of variance in the congested area and the lower speeds in the periods without congestion.

Table 6.4: Network indicators

	March 20th, 2007			June 19th, 2007		
	measured	simulated	diff.	measured	simulated	diff.
TDT (1000 veh.km)	338.95	341.73	0.8 %	340.94	345.58	1.4 %
TD (veh.hrs)	2108.63	2045.12	-3.0 %	1563.58	1665.62	6.5 %

6.3 A10-West Network

6.3.1 Introduction

In the summer of 2001 large roadworks were planned for a part of the western stretch of the ring road around Amsterdam in The Netherlands, called the A10-West. After years of intensive use, a thorough renovation was needed. To be able to carry out these roadworks, Rijkswaterstaat consulted the municipality of Amsterdam and other stakeholders and together they decided to close one carriageway during the summer holidays. From earlier experiences of this type of road works it was known that this method would cause the least disruption for the road user. Therefore, in the period from May 26th until August 26th, 2001 one carriageway with a length of 5 kilometres was closed and the carriageway in the other direction was used to handle the traffic in both directions on narrow lanes. Furthermore, all on-ramps and off-ramps on the stretch

were closed, with some exception for trucks and emergency services. This was done for both carriageways.

Beforehand, it was expected that these roadworks would have a large impact on the traffic, not only on the A10-West itself, but also on the connecting motorways and the Amsterdam road network. To prevent chaos from happening, Rijkswaterstaat, in cooperation with the municipality of Amsterdam and other parties involved, prepared all kinds of measures. The measures consisted of mobility management and traffic management measures, together with a large public relations campaign to inform the public.

Due to the large impact of the loss of capacity on the A10-West motorway, research was done before the road works started, to see whether traffic management measures would be able to mitigate the expected problems. Also, in an early stage, Rijkswaterstaat planned three studies: a dynamic modelling study to determine the effects of the traffic management measures and to adjust these measures, a survey among road users about their choice behaviour before, during and after the roadworks and a large evaluation program to monitor the traffic situation and to analyse the effects of the roadworks and measures (Taale *et al.*, 2006).

Taale *et al.* (2002) formulated the main conclusions of these three studies as follows:

- On the motorway itself no large problems occurred. In peak periods the traffic situation was reasonably good, due to the speed regime and the closure of on and off-ramps, among other things. This was different for the Amsterdam network. Especially on the north-south routes traffic was severely congested. The volume on the A10-West decreased with 38%, but a large part of this presumably local traffic chose another route to arrive at their destination, due to the closure of the on-ramps and off-ramps. So, more traffic used the neighbouring motorways and the urban network. Therefore, a part of the congestion problem moved from the motorway to the urban network.

- The public relations campaign to inform the road users, before and during the roadworks was effective. The campaign was partially responsible for the fact that large problems on the motorways did not occur. Of course, also the holiday period contributed to this. A part of the road users was on holiday or was a couple of days on leave. Also other routes, times of departure or travel modes caused less traffic on the roads.

- The measures related to mobility management (improve existing public transport and extend it with some temporary bus lines, improve the use of the bicycle, P+R facilities, teleworking, etc.) were not very effective. During the roadworks about 10% of the road users using the A10-West, chose alternatives such as public transport and the bicycle. The other mobility management measures were not used. Also the companies in the neighbourhood of the A0-West were not very interested in stimulating mobility management measures.

- The expected traffic problems did not occur. A part of the traffic disappeared or took other routes and departure times, which is in agreement with the findings of Cairns *et al.* (1998) and Goodwin *et al.* (1998). They studied nearly 100 cases covering incidents of

road closure and capacity reduction throughout the world. The report looks at the changes
in travel choice and behaviour that affect traffic conditions when road capacity is reduced
and shows how reducing road capacity can lead to traffic reduction.

In the remainder of this section attention is paid to the data which was collected during the
assessment studies and the calibration and validation of the framework for this network.

6.3.2   Data collection

Part of the evaluation program, as mentioned in the previous section, was a simulation study to
predict the effects of the traffic management measures in advance (Arcadis, 2001). The network
and OD matrix from that study were used as the basis for the calibration and validation. Figure
6.8 shows the network. It consists of 1439 links, 587 nodes and 100 zones (the grey dots in the
figure). Also, the network contains 83 controlled intersections and 4 metered on-ramps.
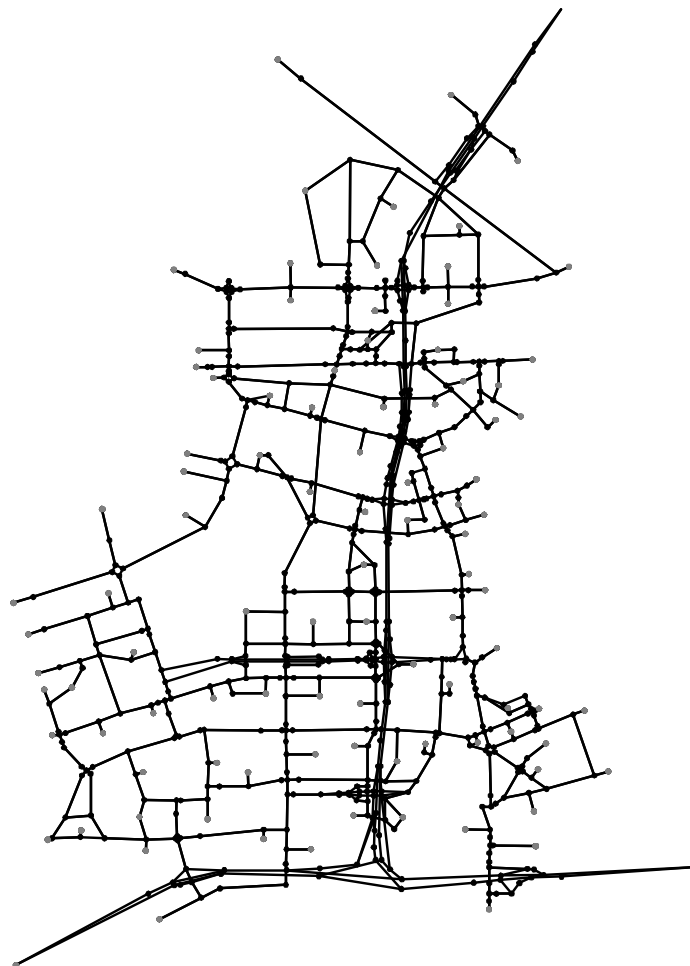
Figure 6.8: Network A10-West

The OD matrix comes from a strategic model, containing a 2-hour period (the evening peak)
from 1998. The matrix was adjusted and updated for the year 2000 and was modified into a

dynamic one. From traffic counts a departure profile was derived and this profile was used to transfer a 2-hour peak matrix into 10 separate matrices for every quarter of an hour between 15:30 and 18:00 hours.

Traffic data was gathered from the A10 (speeds and flows per minute), on-ramps and off-ramps, flows on urban roads and travel times. The before period was in the summer of 2000. The roadworks were realised between May 26th and August 26th, 2001. For the months May, June, July and August of 2000 and 2001 all data from the motorways around Amsterdam was collected. This concerned data per minute of flow and speed for the period between 06:00 and 20:00 hours. On 50 directions of the locations shown in figure 6.9 flows were measured for the on-ramps and off-ramps and the urban network. These flows were measured on a 15-minute base for 2 weeks in the summer of 2000 (June 26-30 and August 7-11) and two weeks in the summer of 2001 (June 11-15 and August 6-10).

The travel times on the routes, shown in figure 6.10, were measured for five days in the before period and seven days in the after period with roadworks. This was done with probe cars. They drove through the network in the morning and evening peak and recorded the travel time and all abnormalities that occurred, e.g. open bridges, empty fuel tanks, etc.

All these data have been used to analyse the traffic before and during the roadworks. For the validation it is important to note that it was measured that the amount of traffic decreased with 11% in the situation with roadworks. Unfortunately, it was not possible to distinguish between long distance and local trips, so a global decrease of the OD matrix was used for this situation.

6.3.3 Calibration

For the calibration of MARPLE for this network, use has been made of the data from the simulation study and the data of the measurements of the before period (without roadworks). Part of the calibration work is based on the work of Li (2005). She investigated the sensitivity of the model for certain input parameters and did a first attempt to validate the model. She found that, next to some model parameters related with route set generation, the OD matrix had the largest impact on the results.

For some of the model parameters the values suggested in Li (2005) are used. By varying the overlap threshold, she found that in the route generation process for the overlap in routes a threshold of 75% should be used. This means that if two routes have more than 75% of their links in common, they are considered as equal routes. She also found that the model is sensitive to the parameter which determines the smallest number of trips to be considered. Therefore, this parameter is set to the value 0, implying that all trips are assigned and simulated.

For the network some modifications were necessary for the use with a macroscopic model instead of a microscopic model. For example, because of the lower level of detail in the macroscopic model, it is not necessary to model all the turning lanes on an intersection. Other modifications dealt with some coding errors and missing links in the network. These errors were corrected and some links were added.
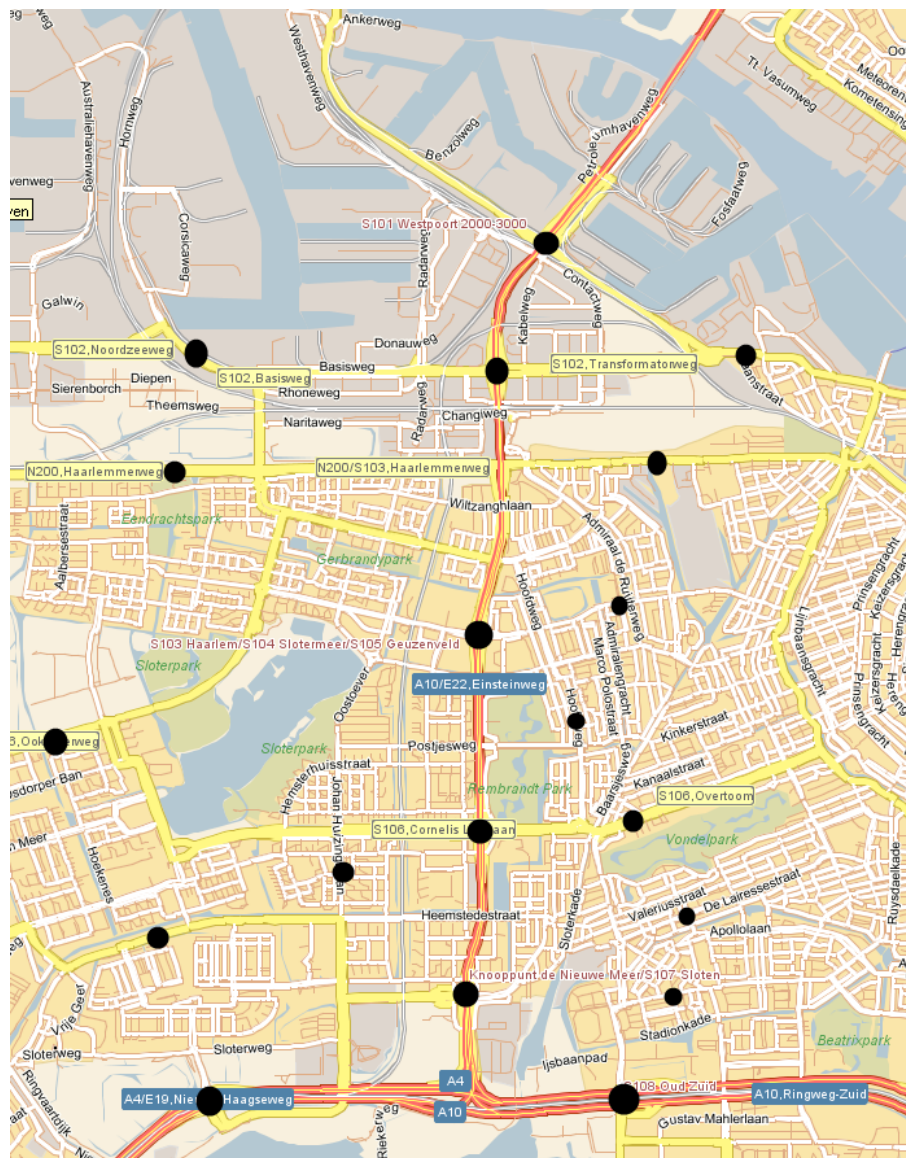
Figure 6.9: Measurement locations for flows
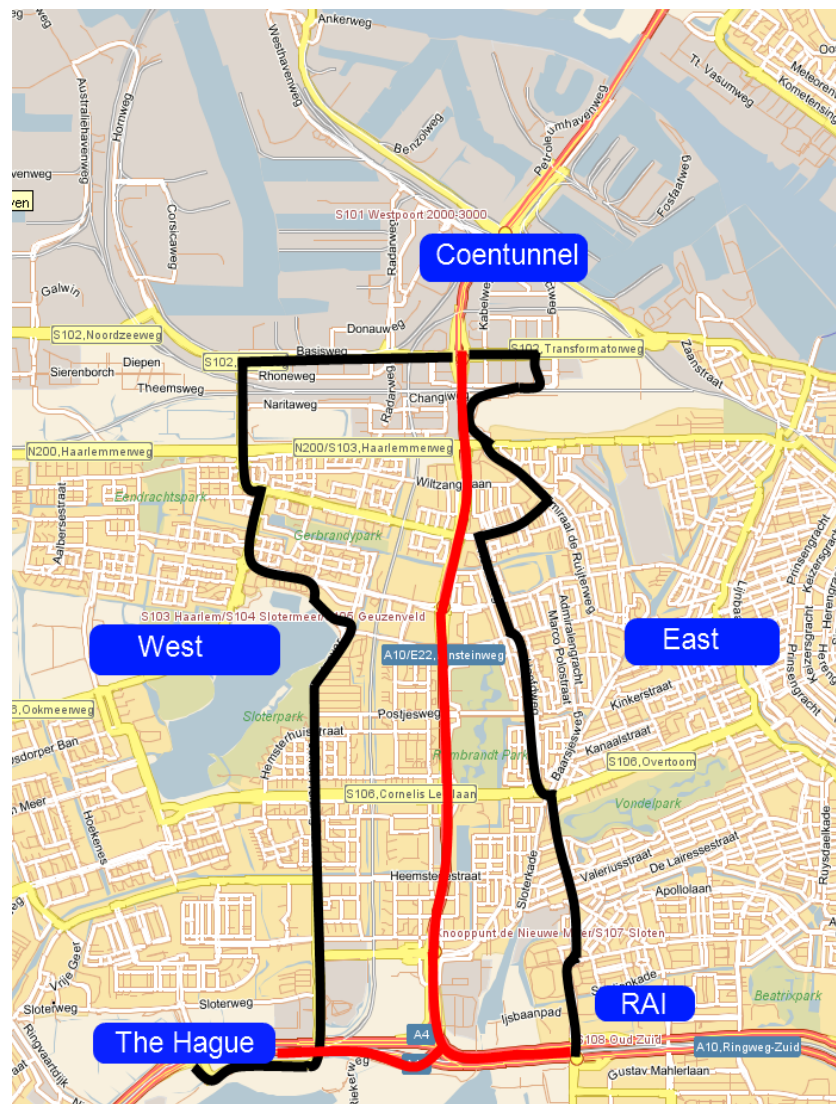
Figure 6.10: Routes for travel time

Because the OD matrix is important input for any DTA model, much attention is paid to its calibration. In the next sections first the results of using the original OD matrix as input for MARPLE is shown. Then the OD matrix is adjusted using an OD estimation method. After that the results of the assignment using this new matrix are shown.

Results of the DTA model before OD estimation

With the adjusted network and the existing OD matrix an equilibrium assignment with MARPLE was done for the before period without road works (situation 2000). For this situation the flows and travel times were compared with the measured ones. The flows are compared for 64 measurement locations on a quarter per hour bases for two and an half hours (15:30 - 18:00 hours). For the flow measurements the summer week in June 2000 was chosen (26-30 June). The values for the 5 days of measurements were averaged to obtain the measured flow values. For the travel time, data from the same week was used, but not per quarter of an hour, but the average for the whole simulation period. This was done, because the number of measurements is too small to distinguish between the separate quarter of an hour periods. The results for the flows (per quarter of an hour in veh/h) are shown in figure 6.11 and for the travel times in figure 6.12.
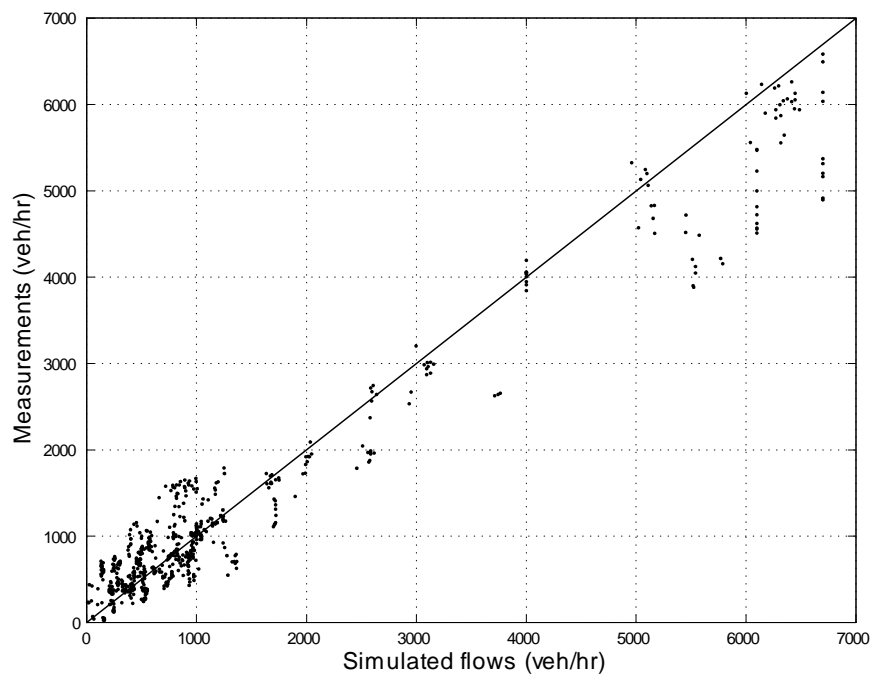
Figure 6.11: Flows before calibration of the OD matrix

For the flows it can be seen that for the motorway flows, the simulated flows are too high and the variation in smaller flows is large. The simulated travel times show a good resemblance with the measured ones. For the flows, separate goodness-of-fit measures are calculated for the motorways, on-ramps and off-ramps, the urban network and the total network. For the travel times only 8 values are available and therefore only the RMSPE and R^2 measures are given. These measures are shown in table 6.5. From this table it can be concluded that for the
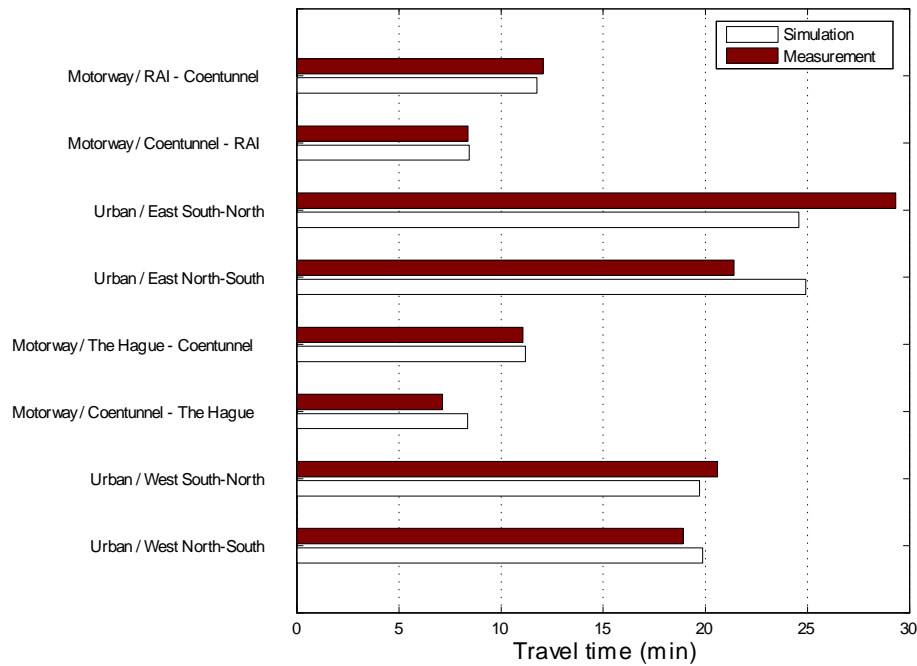
Figure 6.12: Travel times before calibration of the OD matrix

motorway flows the correlation is high, but there is a strong bias, which can also be seen in figure 6.11 (the large flows). The explanation of the variance is better, which is also true for the on-ramp and off-ramp flows. For the urban flows the bias is small, but the variance is not explained that well. The prediction of the travel times is very good, which can be seen in figure 6.12 and the small value for the RMSPE and the large value for R^2.

Table 6.5: Goodness-of-fit for flows and travel times before calibration of the OD matrix

		RMSPE	R^2	U	U_m	U_s	U_c
Flows	Motorways	0.2057	0.8586	0.0823	0.4883	0.0239	0.4930
	On and off-ramps	0.6742	0.4440	0.1973	0.0906	0.0215	0.8906
	Urban network	0.4006	0.7662	0.1456	0.0815	0.1822	0.7406
	Total network	0.5419	0.9489	0.1023	0.0007	0.2966	0.7043
Travel time		0.1051	0.8914				

OD matrix calibration method

To improve the estimation of the flows, the OD matrix was re-estimated using the method described by Van Zuylen and Willumsen (1981) and adjusted by Van Zuylen (1981). This method uses a base OD matrix and adjusts OD flows to match the measured flows as close as possible. The original, static method is described in detail in appendix C. This algorithm was adjusted a little bit and incorporated into a dynamic environment and an iteration loop with the stochastic dynamic assignment (algorithm 4.3). The adjustment of the algorithm concerns

equation C.7. To make calculations simpler, this equation is changed into

$$V_a = \sum_{o,d} p_a^{od} X_o^n q_0^{od} \left(\prod_{a'} (X_{a'}^n)^{p_{a'}^{od}} \right) Y_a, \tag{6.8}$$

where V_a is the measured flow on link a, p_a^{od} is the proportion of trips from o to d using link a, X_o^n is a scaling factor for iteration n, q_0^{od} is the initial OD matrix, X_a^n are the matrix adjustment factors and Y_a is a scaling parameter for link a, which is now general, instead of dependent on the OD pair.

With this adjustment the algorithm was used for dynamic matrices by applying it for every time period separately. The dynamics are dealt with by taking travel times into account for the calculation of the OD link proportions. Equation 6.8 now becomes

$$V_{a,k} = \sum_{o,d} \left(\sum_{t=1}^{k} p_{a,t}^{od} \right) X_o^n q_0^{od} \left(\prod_{a'} (X_{a'}^n)^{\sum_{t=1}^{k} p_{a',t}^{od}} \right) Y_a. \tag{6.9}$$

The complete algorithm is described in algorithm 6.1. In step 5, the convergence criterion used, is the total RMSPE, as defined by equation 6.1.

Algorithm 6.1 *Algorithm for calibration of OD matrix*

Step 1: *Calculate the link inflows u_a and travel times τ_a using the existing route flows and the dynamic network loading model (chapter 5).*
Set $j := 1$

Step 2: *For each time period k:*
For every link a determine OD proportions $p_{a,k}^{od(j)}$ using $u_{a,k}^{(j)}$ and $\tau_{a,k}^{(j)}$.
Calculate $q_k^{od(j)}$ with one iteration of Van Zuylen's algorithm, described in appendix C and using equation 6.9.

Step 3: *Calculate new route flows $\mathbf{f}_k^{(j)} \in \Omega$ with equations 4.15 and 4.16.*

Step 4: *Calculate the link inflows $u_{a,k}^{(j)}$ and travel times $\tau_{a,k}^{(j)}$ using the route flows $\mathbf{f}_k^{(j)}$ and the dynamic network loading model*

Step 5: *If convergence criterion is met, then stop.*
Otherwise, set $j := j+1$ and return to Step 2.

Results after OD estimation

The OD matrix calibration method described in the previous chapter was used to construct an adjusted OD matrix. For the first runs it turned out that the results for the motorway were

not that good, caused by some measurement locations on road sections with congestion. The measurement locations were removed and the process was repeated to obtain a better OD matrix. This matrix was used in a new equilibrium assignment and the results for flows and travel times were again compared with the measured values. The results are shown in the figures 6.13 and 6.14.
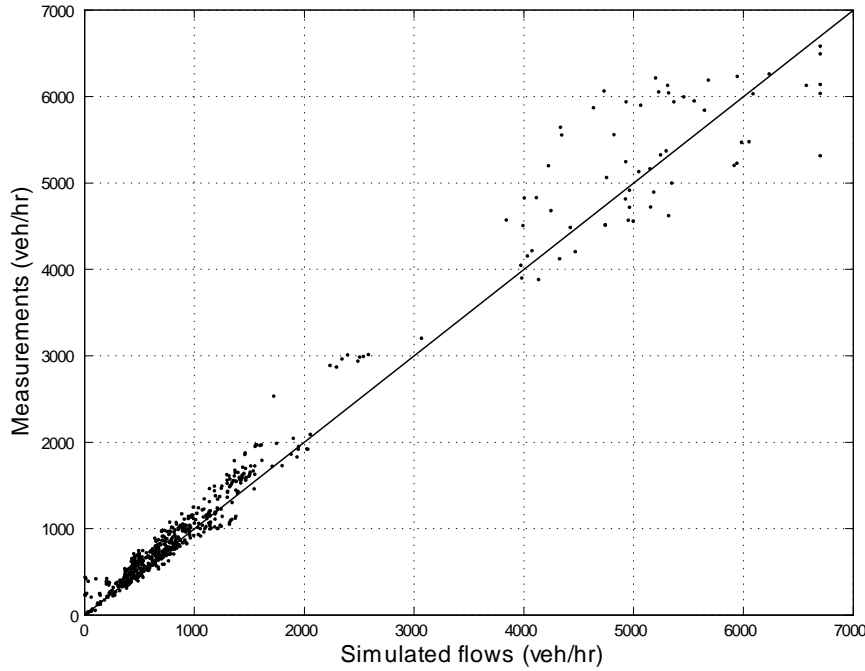
Figure 6.13: Flows after calibration

From these figures it is clear that the simulated flows show a much better correspondence with the measured flows, due to the adjusted OD matrix. The correspondence for the travel times is a little bit worse, which is caused by the fact that in the simulation there is less congestion on the A10-West (less demand in the new OD matrix). This gives shorter travel times for the motorway routes The Hague-Coentunnel and RAI-Coentunnel. But still the travel times in the assignment are close to the measured ones, which is also shown in table 6.6, that gives the goodness-of-fit measures. It is clear that the OD matrix calibration improved the results a lot. For all the network parts RMSPE and U are smaller, which is good. The only problem is that U_m for the on and off-ramps is much larger than before. This means that there is a systematic error for these flows, which is possibly due to the OD estimation algorithms, which contains only one iteration of the assignment and is not included in the complete assignment loop.

Table 6.6: Goodness-of-fit measures for flows and travel times after calibration

		RMSPE	R^2	U	U_m	U_s	U_c
Flows	Motorways	0.1361	0.8802	0.0616	0.1469	0.0547	0.8092
	On and off-ramps	0.1426	0.9584	0.0715	0.4109	0.1756	0.4153
	Urban network	0.2451	0.9317	0.0749	0.0656	0.2356	0.7033
	Total network	0.1832	0.9751	0.0638	0.1193	0.0184	0.8637
Travel time		0.1403	0.9034				
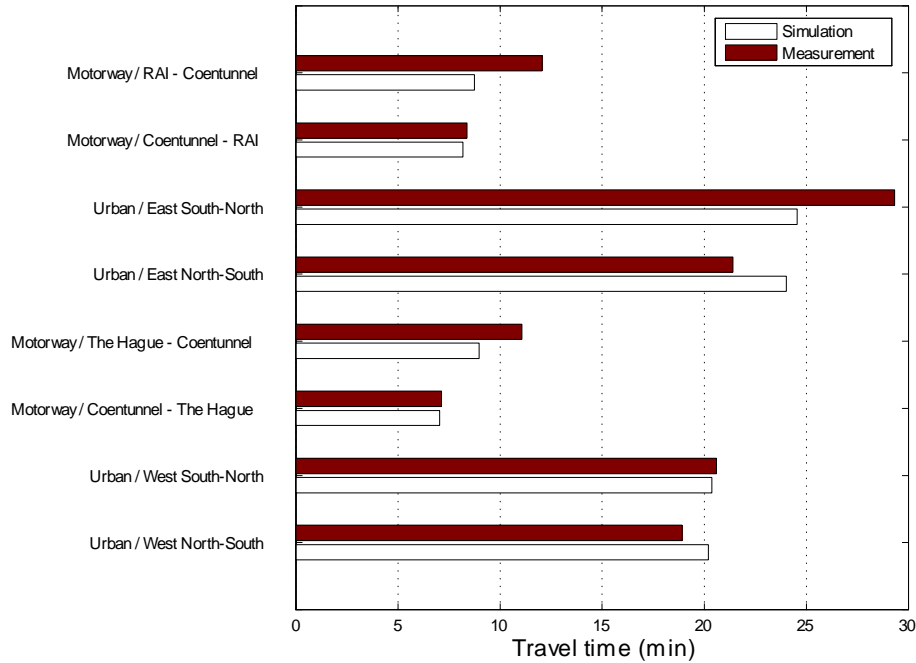
Figure 6.14: Travel times after calibration

6.3.4   Validation

For the validation the situation with roadworks was used. For this situation the network was adjusted, because of the closure of on and off-ramps and the reduced capacity and speed limit for the motorway section where the roadworks were done. From the measurement data during the roadworks it could be concluded that the demand was decreased with 11%, although the decrease was not measured for every OD relation separately. Therefore, for the situation with roadworks a global scale factor of 0.89 was used for all OD pairs. Of course, also new routes had to be generated because of the differences in the network. An equilibrium assignment with the new network and the new OD matrix gives the results for the flows and travel times shown in the figures 6.15 and 6.16 and the results for the goodness-of-fit measures shown in table 6.7.

Table 6.7: Goodness-of-fit measures for flows and travel times for the validation

		RMSPE	R^2	U	U_m	U_s	U_c
Flows	Motorways	0.1634	0.9350	0.0667	0.3353	0.0483	0.6248
	On and off-ramps	0.3933	0.4990	0.1872	0.2655	0.0028	0.7357
	Urban network	0.3844	0.5756	0.1533	0.0352	0.0000	0.9688
	Total network	0.3624	0.9438	0.0978	0.1603	0.0670	0.7744
Travel time		0.1425	0.9807				

From figure 6.15 it is clear that the predicted flows for the motorway are a little bit lower than the measured ones, which also results in a larger bias U_m, but still the error is acceptable. For the on and off-ramp the systematic error U_m is also larger, but the total error remains within acceptable limits. This also is true for the urban network, where U is larger than before. Taking
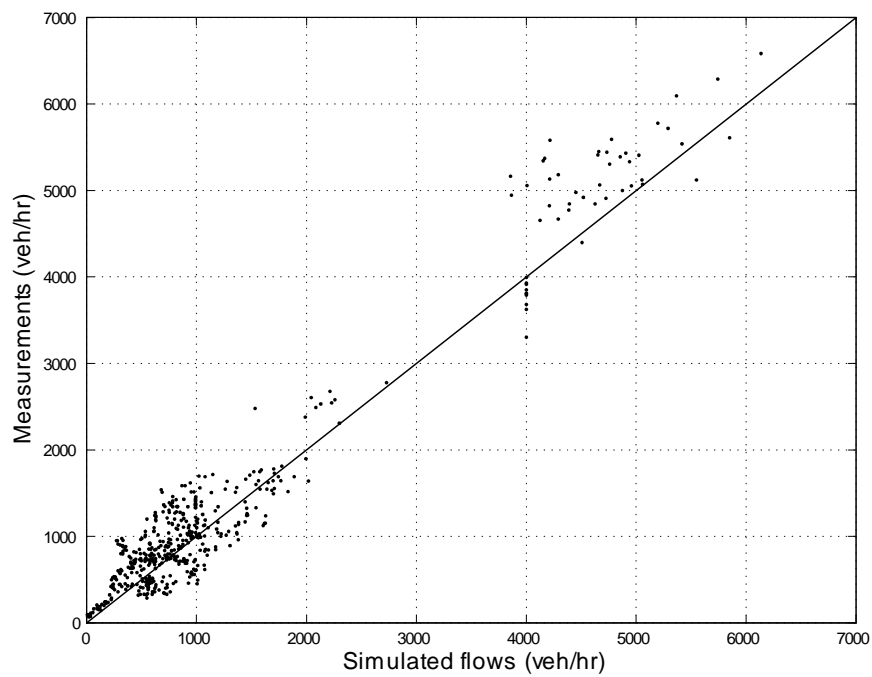
Figure 6.15: Flows for situation with road works
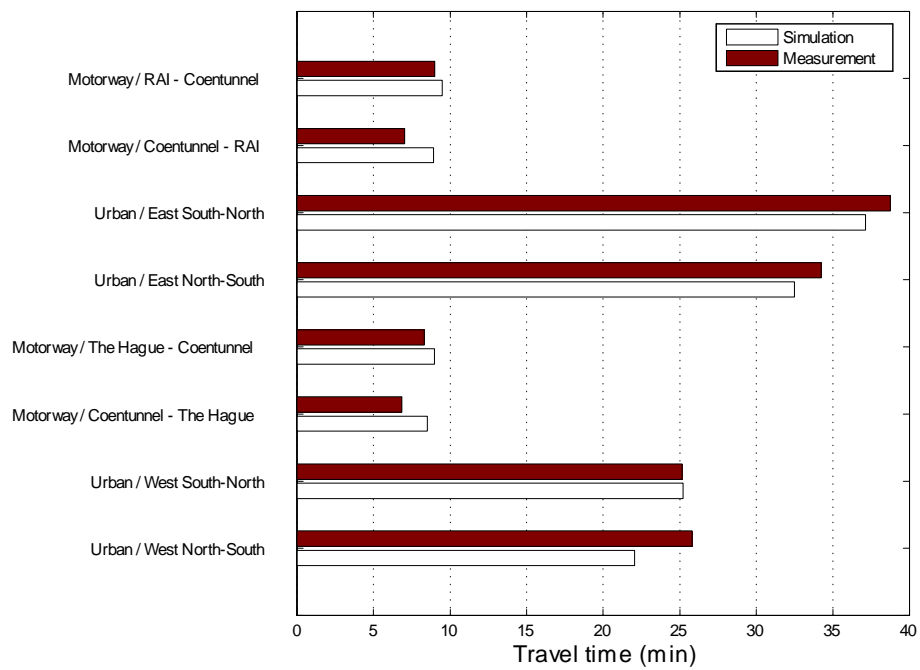
Figure 6.16: Travel times for situation with road works

into account that the OD matrix used was scaled with a global factor and not per relation, the results for the flows are good. Also the simulated travel times reproduce the measured travel times very well.

6.4 Summary

In this chapter the assignment model was calibrated and validated for two cases: a motorway with bottleneck and for a real-life network. For the motorway case data from one day was used to calibrate a number of parameters. Data from two more days were used to validate the model, especially the travel times.

The real-life network was simulated, using data from an assessment of traffic management measures during roadworks on the A10-West near Amsterdam. First, an assignment was done using the existing network and OD matrix. Comparing the simulated flows and travel times with the measured ones showed large deviations. From a previous study done by Li (2005) it was clear that the OD matrix plays an important role. Therefore the OD matrix was calibrated using an adjusted version of the OD estimation method developed by Van Zuylen and Willumsen (1981) and Van Zuylen (1981). The results after the OD matrix calibration are good. The model predicts flows and travel times in good agreement with the measured values. For the validation the situation with roadworks was used. The network and OD matrix were adjusted and the assignment was run. The results for the flows are less good than before, but, taking into account that the OD matrix used was scaled with a global factor and not per relation, the results are satisfactory. The results for the travel times are very good.

The conclusion of this chapter is that the dynamic network loading model is capable of simulating bottlenecks fairly accurate. Also, the dynamic traffic assignment model MARPLE is capable of simulating medium-sized networks with good results. Another conclusion is that calibrating and validating a DTA model is a difficult task and that preparing the input, especially the OD matrix, needs a lot attention, which is confirmed by Balakrishna *et al.* (2005).

7

Case Studies

The framework for anticipatory control, described in chapter 3 and elaborated in the chapters 4 and 5, was used to study a few theoretical cases. These cases represent small networks to find out whether the concept works. Also the network used for the calibration and validation was studied to see whether or not the framework can be used for medium-sized urban networks. For all cases the stochastic assignment (see section 4.2) was used to reach equilibrium in the main loop. In each section the case is described and the results for all control strategies are shown, sometimes for a few alternatives in network layout or demand. Total delay (veh.hours) is the main indicator for the comparison of the control alternatives, sometimes supplemented with green times, flows or travel times. Because of the stochastic nature of the optimisation methods for the Local GA, anticipatory control and system optimum control strategies, for those strategies 10 runs were done to obtain an average value, which for all cases was sufficient to obtain reliable results (see also section 3.4.5). Except for those cases where it is explicitly mentioned, a stochastic assignment is used in the optimisation loop also. Parameter values for the assignment and control strategies are given in section 7.1 and are the same for the other cases, unless stated otherwise.

7.1   Simple case

We start with the most simple network for this research: a network with one OD pair (A-B), having two routes: one route with a controlled intersection and a bypass route (see figure 7.1). The other OD pair in the network (C-D) is added to get a controlled intersection and to get more realistic results for the total network. The black dot represents the signal controlled intersection, the blank dots represent the nodes. Characteristics of the links are given in table
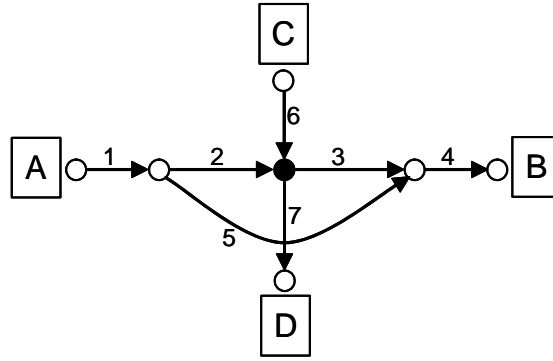
Figure 7.1: Network simple case

Table 7.1: Link characteristics for the simple case

Link	Length	Number of lanes	Capacity	Maximum speed
1	999	2	4000	50
2	1570	1	2000	50
3	571	1	2000	50
4	499	2	4000	50
5	4492	2	4000	80
6	941	1	2000	50
7	553	1	2000	50

7.1. The number of time periods for this case and the demand for the two OD pairs is given in table 7.2. The time periods have a length of 15 minutes.

For this network the control strategies described in sections 3.2, 3.4 and 3.5 are used. For the fixed-time control strategy the green times for all cases are based on the busiest time period of the initial assignment. For the optimisation strategies the green times are allowed to vary between 7 and 40 seconds. The cycle time is not fixed, but is based on the sum of the green times and the intersection lost time, which is 10 seconds. Following Chen (1998), the parameters for the stochastic assignment are chosen to be $\beta = 1.0$, $\gamma = 1.0$ and $\theta = 1.0$ or 3.0 (see section 4.2). For the local GA strategy the number of generations was set to 50, with a population size of 20. For anticipatory control the minimum number of function evaluations was set to 100, with a maximum of five times the number of variables to optimise. The values for μ and λ are set automatically. Finally, for the system optimum control the maximum number of function evaluations was set to 30 times the number of variables to optimise. This should lead to a near optimal solution, considering green times and route flows, as concluded by Hansen and Ostermeier (2001).

Table 7.2: Demand for the simple case (veh/h)

OD pair	Time periods				
	1	2	3	4	5
AB	1200	3000	1800	800	800
CD	400	1500	800	300	300

To show the difference between the two types of assignment, for this case a deterministic and stochastic equilibrium was calculated, using the framework described in chapter 3. For the stochastic equilibrium two different values of the θ parameter were used: θ = 1.0 and θ = 3.0. This was done to determine the influence of this important parameter on the results. θ is a parameter that reflects the degree of uncertainty in the travel time knowledge of the road users. The larger θ is, the less uncertain road users are about their travel time or it can also be stated that the larger θ is, the more information is available. The framework was run until convergence was reached and for convergence the maximum difference in flows between two iterations was used (see section 4.2.5). The threshold was set to 1% for the stochastic assignment and 0.1% for the deterministic assignment. The results for the control strategies are shown in figure 7.2.
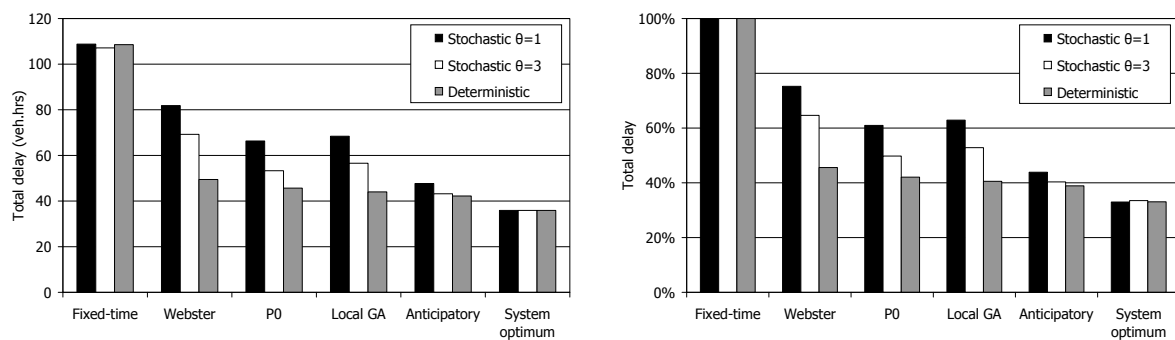
Figure 7.2: Absolute and relative results for different assignment types

From this figure it is clear that anticipatory control can improve the traffic situation in terms of total delay. The improvement compared with the local control strategies, is dependent on the level of knowledge the road users have. If they are completely informed (deterministic assignment) the improvement is less (4% - 15%) than when they do not have perfect knowledge of the situation in the network (20% - 42%). It seems that the more uncertain road users are, the larger the improvements can be. This seems logical because more road users will change route if the uncertainty is higher.

To see what the effects of anticipatory control are, some traffic indicators (for the stochastic assignment with θ = 1.0) are shown in table 7.3. For the two busiest time periods (time period 2 and 3, see table 7.2) the green times of the signal controlled intersection, the route flow for the route using link 2 and the route travel times for the routes using link 2 or 5 are shown. These two periods are chosen, because in the second period congestion starts, while in the third period the queue from the second period is still there. The table shows clearly that it is advantageous to put as much traffic as possible on the longer route. The advantage is less delay, the disadvantage is more vehicle kilometres travelled (about 5% - 10% for anticipatory control and 18% - 25% for the system optimum), due to the shift of traffic to the longer route. It can be seen that the local strategies already give less green to link 2 than the optimised fixed-time strategy. But anticipatory control gives link 2 even less green time, 'forcing' route users to choose the bypass, which is better for the network as a whole.

The robustness of the results in relation with the demand was tested by taking a 20% lower and a 20% higher demand. The results for those simulations are shown in figure 7.3. Again, it is

Table 7.3: Some results for the simple case for time periods 2 and 3

| | Green time | | | | Flow link 2 | | Travel time | | | |
| | link 2 (s) | | link 6 (s) | | (veh/h) | | link 2 (min) | | link 5 (min) | |
Period	2	3	2	3	2	3	2	3	2	3
Control strategy										
Fixed-time	40.0	40.0	35.0	35.0	1161	585	6.71	6.79	6.25	6.07
Webster	30.5	11.5	40.0	25.9	970	308	7.05	9.83	6.27	6.10
P_0	21.6	18.7	40.0	40.0	824	302	7.37	8.79	6.28	6.11
Local GA	23.4	19.0	40.0	39.1	863	319	7.30	8.85	6.28	6.10
Anticipatory	8.2	7.4	39.7	38.6	311	167	8.55	8.16	6.35	6.12
System Optimum	7.0	7.2	40.0	40.0	117	0	5.81	5.52	6.38	6.14

clear that anticipatory control is better than local control strategies also for the lower and higher demand scenarios, although the improvements for the higher demand are somewhat lower.
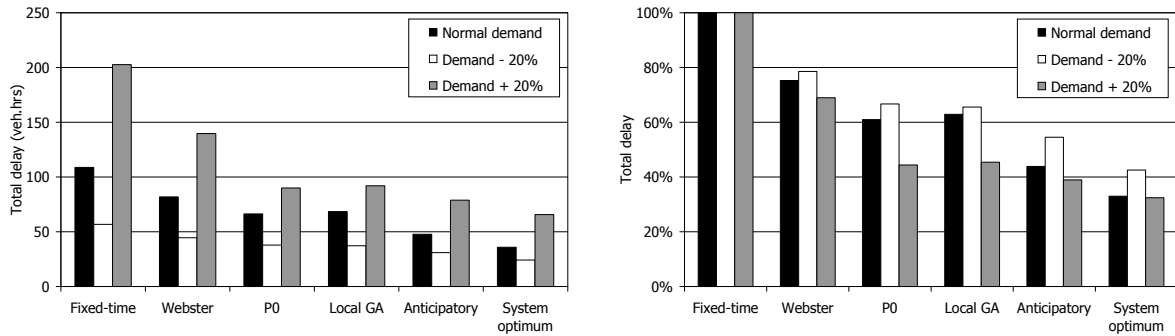
Figure 7.3: Absolute and relative results for different demands

7.2   Other small cases

7.2.1   Case 2

Apart from the simple case illustrated in the previous section, also other small cases were tested. The first one looks a lot like the example from section 7.1, but with a difference: the bypass route also crosses the conflicting traffic stream (C-D). The network is shown in figure 7.4 and the links have the same characteristics as in table 7.1. Link 5 is split into two links with equal characteristics and link 7 is longer. An extra link is added from the intersection to zone D with the same link characteristics as link 7. The demand is the same as for the first case and is given in table 7.2. For the routes there are two possibilities: OD pair AB has 2 routes or OD pair AB has 3 routes, where the third route uses part of the route for OD pair CD. The results for this case (with two possibilities for the number of routes) are shown in figure 7.5. Again, anticipatory control is better than local control. The improvements are as high as 4% - 11% in terms of total delay, at the expense of an insignificant small number of extra vehicle kilometres. For the situation with 3 routes, the system optimum is considerable lower than for the situation
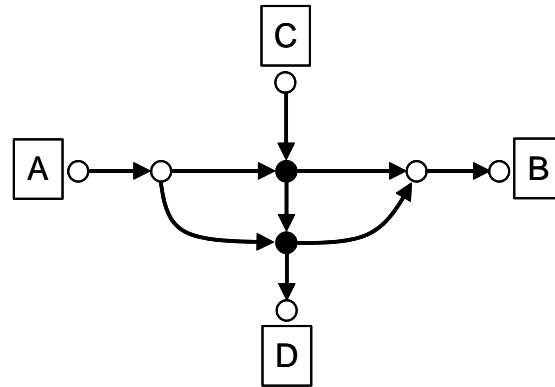
Figure 7.4: Network second case

with 2 routes. The extra alternative route could lead to a better distribution of the traffic in the network, but the anticipatory control method for the situation with 3 routes does not converge to this situation, although it is better than the local control strategies.
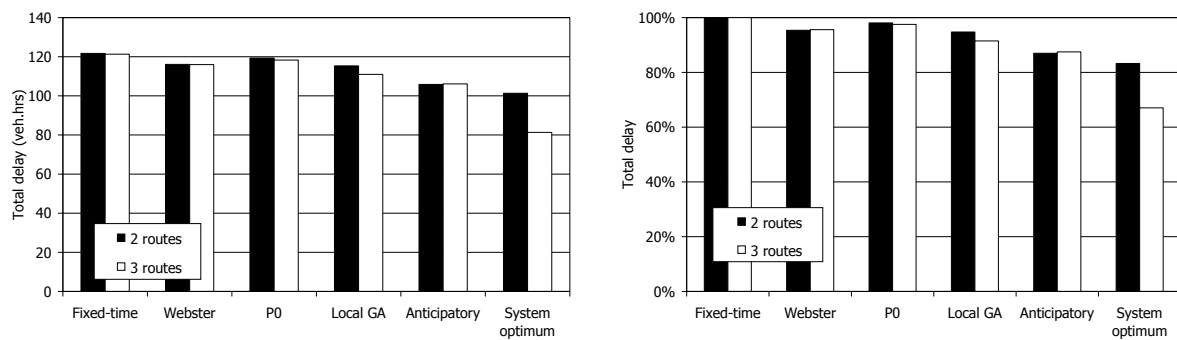
Figure 7.5: Absolute and relative values for the total delay for case 2

Also for case 2 (with 2 routes from A to B: upper and lower route) the results in terms of flows and travel times for the periods 2 and 3 are shown in table 7.4.

Table 7.4: Some results for the simple case for time periods 2 and 3

	Flow				Travel time			
	up (veh/h)		low (veh/h)		up (min)		low (min)	
Period	2	3	2	3	2	3	2	3
Control strategy								
Fixed-time	1214	540	1786	1260	6.93	7.19	6.54	6.33
Webster	1094	449	1906	1351	7.36	10.07	6.71	6.98
P_0	981	428	2019	1372	7.66	10.02	6.80	8.38
Local GA	1045	481	1955	1319	7.51	9.52	6.58	7.92
Anticipatory	1126	367	1874	1433	7.13	8.13	6.61	6.59
System Optimum	926	615	2074	1185	6.18	5.99	7.02	6.95

Multiple equilibria

A similar case was investigated by Van Zuylen and Taale (2000). Using an analytical approach with a function minimisation for the route choice problem and Webster's 2-term delay formula, they found that different equilibria could exist, dependent on the demand. To test if this kind of behaviour also occurs with MARPLE, the network of case 2 was modified slightly: both routes from A to B were made equal in length. The demand of OD pair AB was assigned to the two routes according to a pre-defined distribution. In the first simulation run all traffic was assigned to the upper route and in successive runs 10% of the traffic was taken from the upper route and assigned to the lower route. The results of the simulation runs were checked to see if there could be more than one equilibrium. Different demand patterns were tried, but with none of them multiple equilibria occurred. It seems that the analytical results from Van Zuylen and Taale (2000) are due to the travel time function and minimisation function used.

7.2.2   Case 3

The third case is more symmetric than the previous two cases. Again, there are two OD pairs, but now each with two possible routes. The network is drawn in figure 7.6.
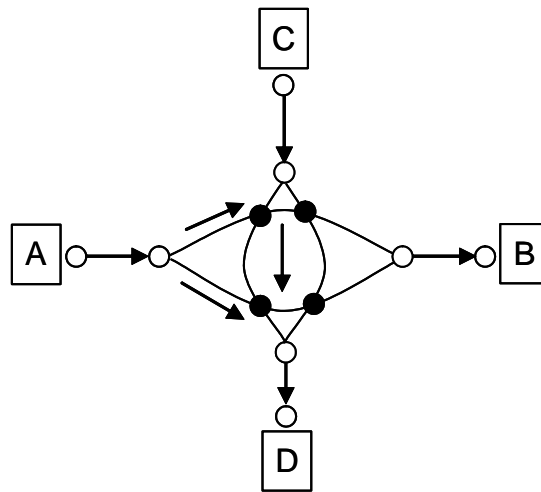
Figure 7.6: Network for case 3

All the links have one lane and a capacity of 2000 veh/h, except for the connector links (between origins and destinations and subsequent nodes), which have two lanes and a capacity of 3000 veh/h. The maximum speed for the network links is 50 km/h. The demand for the OD pairs AB and CD is again the same as given in table 7.2. Both OD pairs have two possible routes. For this case we distinguish between two alternatives: one were all routes have equal length (4639 metres) and an alternative were one route from A to B has a length of 4639 metres and the other a length of 6639 metres and were the two routes for OD pair CD both have a length of 5331 metres. For these two alternatives the results are shown in figure 7.7.

For the first alternative anticipatory control is better than Webster control (9%) and P_0 (4%), but not better than Local GA (local optimisation with genetic algorithms). If we compare the
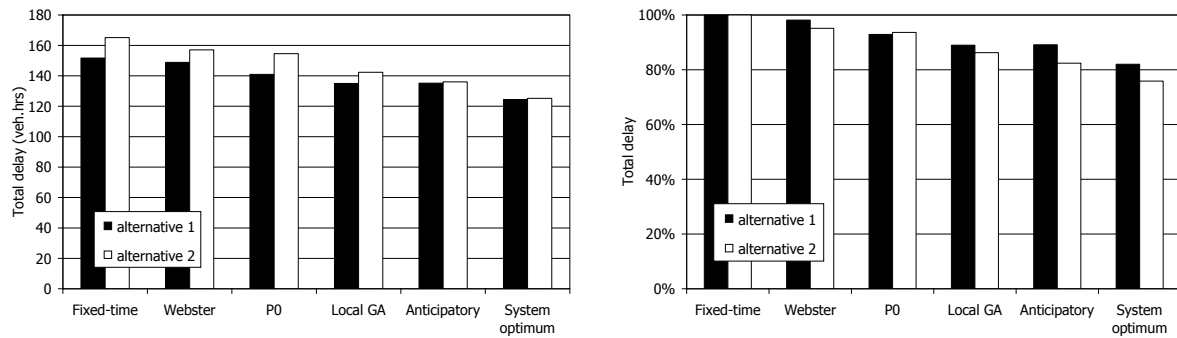
Figure 7.7: Absolute and relative values for the total delay for case 3

flow pattern in equilibrium of Local GA and anticipatory control there are some differences, but these differences are due to the different green times and not due to a different equilibrium.

It seems that for pure symmetric networks anticipatory control does not improve much. If the network is not symmetric, like in the second alternative, anticipatory control does improve the situation compared to local control strategies. The improvement in total delay varies from 5% to 14%. The difference with the system optimum is 9% for both alternatives. For now this statement is based on a few examples. Further research should reveal if this is also true for other networks.

7.2.3   Case 4

The network for the fourth case is sketched in figure 7.8. This network has two signal controlled intersections. All the links have a maximum speed of 50 km/h. Other characteristics are shown in table 7.5. OD pair AB has two routes with length 4000 metres and OD pair AC has three routes with length 5000 metres. Also, for this case two alternatives were studied. These alternatives have different demand profiles, which are given in table 7.6. The first demand profile has no traffic for OD pair AB, thus zone B cannot be reached. In the second demand profile, link 10 is opened for traffic, attracting traffic from zone C to Zone B. The results are shown in figure 7.9. The benefits of anticipatory control can be clearly seen, especially for the first demand profile. The control strategy puts as much traffic as possible on the route using links 1, 5 and 6, which is better for the network as a whole, while the local strategies try to handle the traffic at the local level. Due to the different saturation flows, P_0 is the best local strategy, because it tries to use the available capacity in the network.
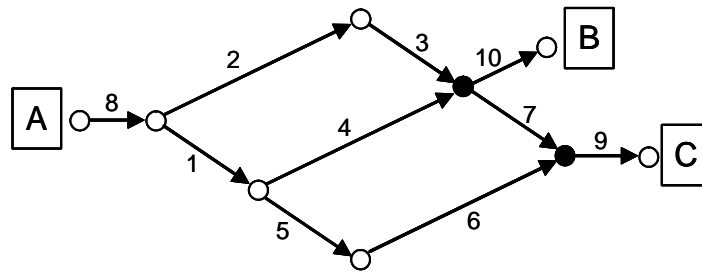
Figure 7.8: Network for case 4

Table 7.5: Link characteristics for case 4

Link	Length	Number of lanes	Capacity	Maximum speed
1	1000	2	4000	50
2	2000	1	2000	50
3	1000	1	2000	50
4	2000	1	2000	50
5	1000	2	4000	50
6	2000	2	4000	50
7	1000	1	2000	50
8	500	3	6000	50
9	500	3	6000	50
10	500	3	6000	50

Table 7.6: Demand for case 4 (veh/h)

		Time periods				
	OD pair	1	2	3	4	5
Demand 1	AB	0	0	0	0	0
	AC	2000	3700	2000	1500	1000
Demand 2	AB	500	1000	500	300	300
	AC	1500	2700	1500	1200	700
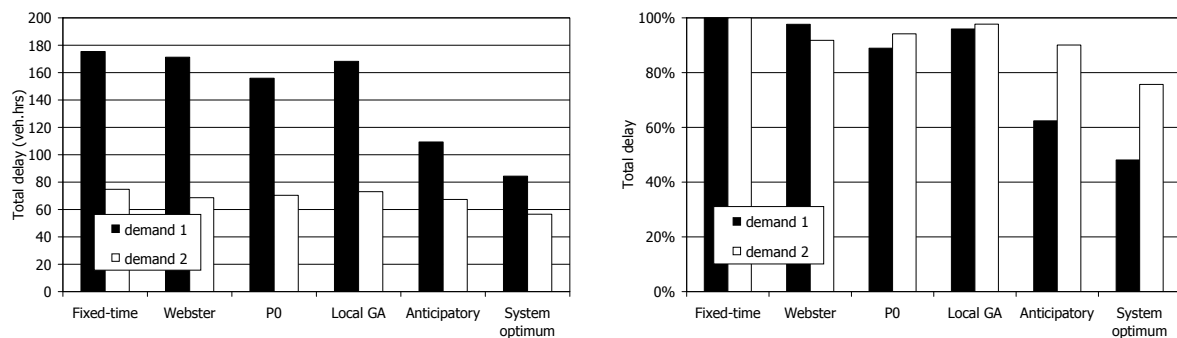
Figure 7.9: Absolute and relative values for the total delay for case 4

7.3 Cases with signal control and ramp metering

Traffic management is not restricted to traffic signal control, but involves also other measures such as ramp metering. For local control strategies, ramp metering is included in the model with a simple rule: the amount of traffic allowed to enter the motorway from the on-ramp is equal to the capacity downstream minus the traffic flow upstream the motorway. This is translated to a green time varying between 2 and 12 seconds, where the cycle time is always 12 seconds. In this way the green time can be optimised, just like normal signal control. So if the saturation flow of the on-ramp is 1800 veh/h and the green time is 2 seconds, only 300 veh/h are allowed to enter. If the green time is 6 seconds, 900 veh/h are allowed to enter. For anticipatory and system optimum control the metering rates are included in the optimisation.

Integrated control was studies for two cases: a network with one controlled intersection and two metered on-ramps and a network with two controlled intersections, also with two metered on-ramps. The networks are drawn in figures 7.10 and 7.11. The grey nodes represent the metered on-ramps and the black nodes the intersections with signal control.
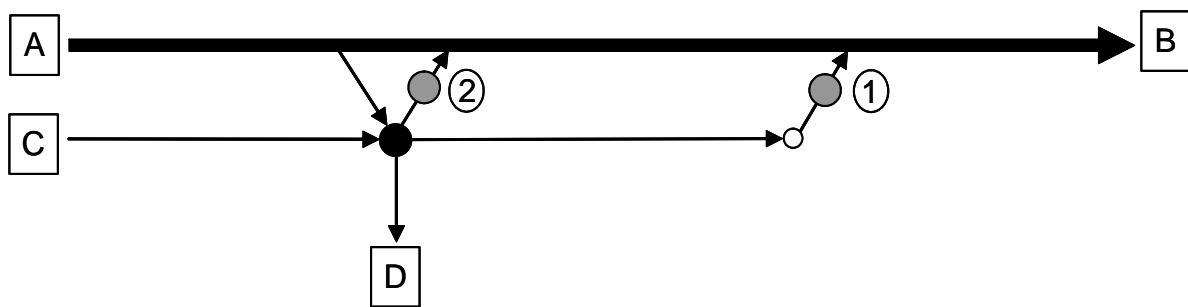
Figure 7.10: Network for case 5a
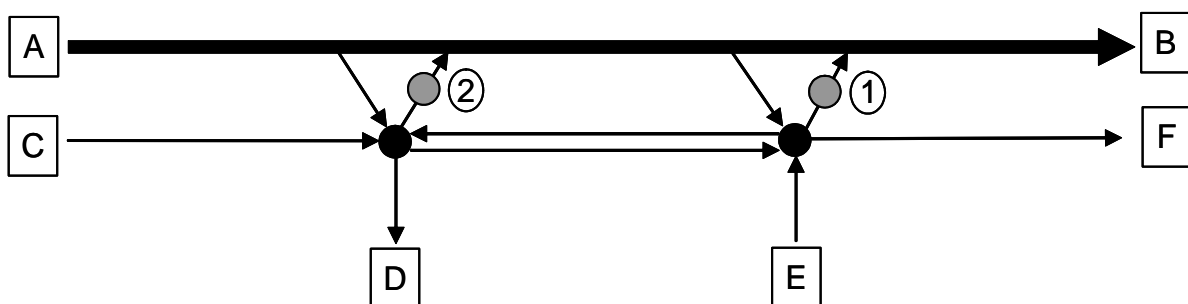
Figure 7.11: Network for case 5b

7.3.1 Case 5a

The network for case 5a is the simplest: a motorway with 2 lanes with a capacity of 4200 veh/h runs from A to B. Downstream there is a bottleneck with a capacity of 4000 veh/h. The maximum speed on the motorway is 100 km/h. Traffic from the urban network, coming from

origin C, can choose between two on-ramps to enter the motorway. For this traffic the through and left-turn movements are controlled with separate signals. Furthermore, there is a conflicting movement with traffic leaving the motorway going to destination D. To make the case no too complicated, for this traffic no left turn is possible. So, we have three OD pairs (AB, AD and CB) and only OD pair CB has two routes. The motorway is 6474 metres long and the distance between the on-ramps is 2045 metres. The routes for OD pairs CB have a length of 4599 metres (on-ramp 1) and 5705 metres (on-ramp 2). The maximum speed on the urban network is 70 km/h and on the on-ramps and off-ramps 80 km/h.

The demand for this network is shown in table 7.7. The simulation period of one hour and a half is split into six 15-minutes periods, with a peaked demand profile. The demand exceeds the capacity of the bottleneck for the first 3 periods. In these periods congestion builds up and in the last three periods congestion decreases.

Table 7.7: Demand for case 5 (veh/h)

	OD pair	Time periods					
		1	2	3	4	5	6
Case 5a	AB	3100	3800	3500	3000	2000	2000
	AD	150	200	200	100	100	100
	CB	1000	1500	1200	1000	600	300
Case 5b	AB	3100	3800	3500	3000	2000	2000
	AD	150	200	200	100	100	100
	AF	150	200	200	100	100	100
	CB	500	750	600	500	300	150
	CF	100	100	100	100	100	100
	EB	500	750	600	500	300	150
	ED	100	100	100	100	100	100

7.3.2   Case 5b

The network of case 5b has more OD pairs and therefore more route choice possibilities. The motorway is the same as in case 5a: maximum speed of 100 km/h, 2 lanes, a capacity of 4200 veh/h and downstream a bottleneck of 4000 veh/h. The off-ramps and on-ramps have a maximum speed of 80 km/h and the urban links a maximum speed of 50 km/h. The two signal controlled intersections have both three conflicting movements. All movements are allowed, except taking the off-ramp and directly the on-ramp. All OD pairs have two routes to choose from, except OD pair ED, which has only one route. The demand for this case is also given in table 7.7.

7.3.3   Results

The results for both cases are given in figure 7.12. Again the figure shows the absolute values for the total delay in the network at equilibrium and the relative percentages of the total delay

compared with fixed-time control.
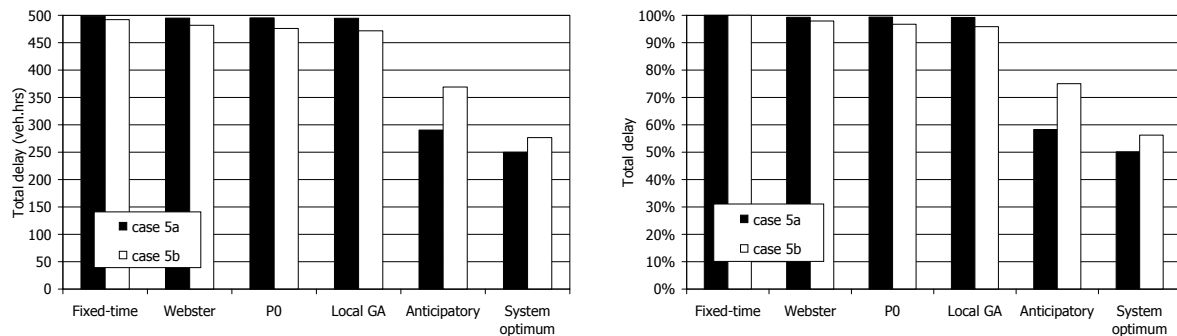
Figure 7.12: Absolute and relative values for the total delay for case 5

The results show that local control strategies cannot improve the situation much. Only a few percent (1% - 4%) less delay compared with fixed-time control (and metering for the on-ramps). In both situations anticipatory control improves the situation a lot (42% for case 5a and 25% for case 5b) and lets the network operate closer to the system optimum. For case 5a this result is obtained by shifting traffic from the route using on-ramp 2 to the route using on-ramp 1. The signal controls are set in such a way that metering on the on-ramps is not so strict, but the intersection control is used to meter the traffic.

For case 5b anticipatory control keeps the traffic for OD pair AB on the motorway. So, the detour for this OD pair via the urban network is not attractive enough. For OD pair CB most of the traffic is directed via on-ramp 2 and for OD pair EB via on-ramp 1. In this way a better use of the network is obtained with less delay (about 23%) than with local control strategies.

7.4 Cases with three players

For case 5 in the previous section we have assumed that signal control and ramp metering are operated by the same road authority. In The Netherlands that is often not the case. Ramp metering is done by the motorway authority Rijkswaterstaat and signal control by the province or the neighbouring municipality. To study what happens if two road authorities are involved, two cases were studied. The first case (case 6) is shown in figure 7.13. The second case (case 7) has the network of case 5b (see figure 7.11), but a different demand. It also has more control possibilities.

7.4.1 Description of the cases

The two road authorities are two players, the road users are considered the third one (see also section 3.6). In the network of case 6 all three players have one thing to control: the motorway authority controls the ramp metering, the municipality controls the intersection and the road users control their route choice. To keep it simple, only traffic going from C to F has a choice of
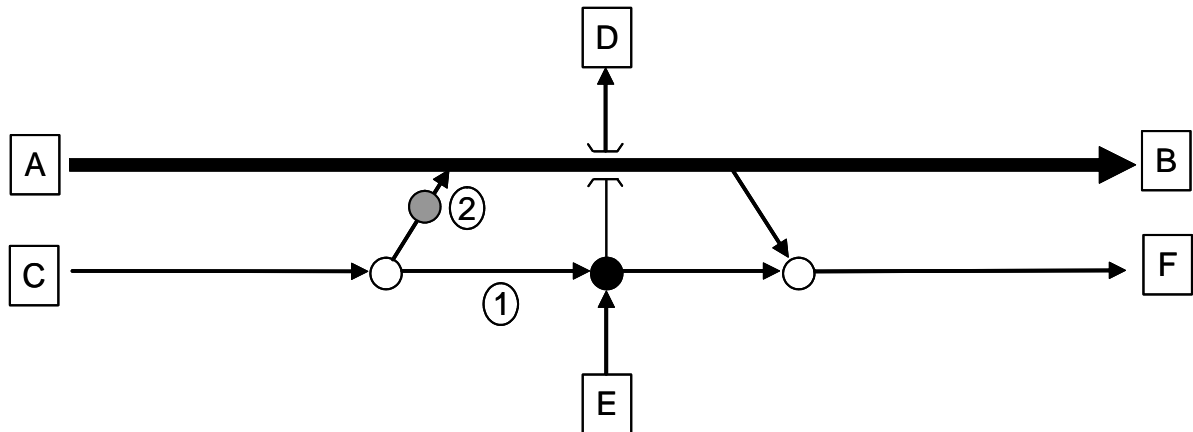
Figure 7.13: Network of three player case

route: route 1 via the controlled intersection (black dot), with a length of 5200 metres, or route 2 via the metered on ramp (grey dot), with a length of 5800 metres. The intersection also controls a second flow: from E to D. Traffic from A to B can use the motorway: 2 lanes with a total capacity of 4000 veh/h with a length of 5000 metres. The maximum speed on the motorway is 100 km/h, on the urban streets 50 km/h and on the on and off ramp 80 km/h. The demand for this network is shown in table 7.8. The simulation period of one hour and a quarter of an hour is split into five 15-minutes periods, with a peaked demand profile for all OD pairs.

Table 7.8: Demand for case 6 (veh/h)

	Time periods				
OD pair	1	2	3	4	5
AB	3255	3570	3570	3150	2100
CF	1050	1575	1365	1050	630
ED	630	1155	945	525	420

The network of case 7 is the same as for case 5b. It has more route choice options and more control possibilities for both road operators. The demand is similar to the demand shown in table 7.7, but higher. For this case the demand in table 7.9 is used.

Table 7.9: Demand for case 7 (veh/h)

	Time periods					
OD pair	1	2	3	4	5	6
AB	3100	3800	3500	3000	2000	2000
AD	350	400	400	300	300	300
AF	350	400	400	300	300	300
CB	700	950	800	700	500	350
CF	300	300	300	300	300	300
EB	700	950	800	700	500	350
ED	300	300	300	300	300	300

7.4.2   Possible control situations

Because of the higher speeds on the motorway, traffic to destinations B and F has a preference for routes using the motorway, causing congestion on the motorway, but favouring traffic on the urban roads. This can be considered the base situation. If the motorway operator introduces ramp metering, travel times on motorway routes become larger and more traffic will choose urban routes, inflicting also more delay for other urban traffic. Therefore, the urban authority may want to change the traffic control for the intersection. We consider this the reference situation, which leads to a Nash equilibrium between the three players, because the players react on each other's moves. A different equilibrium arises if we assume that one player anticipates the reaction of the two other players. There are two possibilities: a Stackelberg equilibrium with the motorway operator as leader or a Stackelberg equilibrium with the operator of the urban roads as leader. The leader knows the strategies of the road users (route choice) and the other road operator (traffic control) and anticipates their reaction. This anticipation can also be done by both road operators separately or coordinated. In the first case they both anticipate the route choice of the road users independently, in the second case their sub-networks are considered to be one network and they work together and have an integrated strategy which anticipates the route choice. This also leads to a Stackelberg equilibrium with the cooperating road operators as the leader. As a benchmark we have the system optimum situation in which control and route choice is chosen such that total delay in the whole network is minimised.

7.4.3   Results for case 6

The results for all situations are shown in table 7.10. The results are given in terms of the total delay of the equilibrium solution. The total delay is split into two parts: delay for the motorway and delay for the urban roads. From the table it is clear that the introduction of ramp metering is beneficial for the motorway. It gives some extra delay on the urban network, but the total delay is much less (-60%) than in the situation without ramp metering.

Table 7.10: Total delay for three player game of case 6

Situation	Total delay (veh.hrs)			Total delay (%)		
	motorway	urban	total	motorway	urban	total
Base	334.75	27.98	362.74			
Reference	95.56	46.27	141.83	100.0	100.0	100.0
Stackelberg motorway	88.23	56.78	145.01	92.3	122.7	102.2
Stackelberg urban	95.67	41.79	136.46	100.1	90.3	96.9
Stackelberg separately	95.65	41.60	137.25	100.1	89.9	96.8
Stackelberg integrated	94.27	45.09	139.36	98.7	97.4	98.3
System Optimum	85.36	44.92	130.28	89.3	97.1	91.9

The Stackelberg equilibrium where the motorway is the leader (Stackelberg motorway) improves the situation on the motorway (-8%), but this has negative consequences for the urban

network (+23%) and also the total delay is higher than for the reference situation (+2%). The opposite is true for the Stackelberg equilibrium if the urban authority is the leader (Stackelberg urban): the urban delay is smaller (-10%), the motorway delay is more or less the same and the total delay is also smaller (-3%). If both authorities anticipate the reaction of the drivers separately (Stackelberg separately), only the situation on the urban roads improves (-10%), giving an overall improvement of 3%. Integrated anticipatory control (Stackelberg integrated) improves the situation on both sub-networks a little bit, resulting in a total improvement of 2%. The system optimum distributes traffic and optimises control in such a way that the improvement for the motorway is 10% and for the urban roads 3%, giving a total improvement of 8%.

For this case it can be concluded that anticipatory control improves the total delay in most situations, except for the situation in which the motorway authority is the leader. If the motorway authority is the leader, the situation on the motorway is the best, but due to the strict control strategy it has large negative effects for the urban network. For the other situations the urban network profits the most from anticipatory control.

7.4.4   Results for case 7

The results for case 7 are obtained for the same situations and control strategies as case 6. They are shown in table 7.11.

Table 7.11: Total delay for three player game of case 7

	Total delay (veh.hrs)			Total delay (%)		
Situation	motorway	urban	total	motorway	urban	total
Base	629.15	250.71	879.86			
Reference	446.90	254.61	701.51	100.0	100.0	100.0
Stackelberg motorway	178.79	349.81	528.60	40.0	137.4	75.4
Stackelberg urban	449.69	198.21	647.90	100.6	77.8	92.4
Stackelberg separately	447.66	194.71	642.37	100.2	76.5	91.6
Stackelberg integrated	221.95	238.91	460.86	49.7	93.8	65.7
System Optimum	187.09	262.39	449.48	41.9	103.1	64.1

From the results it can be concluded that for this case integrated anticipatory control gives the best results: a decrease of 34% in total delay compared with the reference situation. The results are also close to the system optimum solution. It is striking that integrated anticipatory control (Stackelberg integrated) is much better than separate anticipatory control (Stackelberg separately). This is different from the result of case 6, but an explanation could be that in case 7 there are more control possibilities than in case 6. The second best strategy is anticipatory control where the motorway operator is the leader (Stackelberg motorway). This is because for this case the biggest profit can be earned on the motorway: 60% less delay.

7.5 Applicability of the framework

The applicability of the framework for real-life networks is for a large part determined by its calculation time. Generally, we can state that the larger the network, the larger the calculation time will be, because the time needed to run the DNL model is linearly dependent on the size of the network. And what is more, the anticipatory control strategy, as it is developed in this thesis, is also time consuming, due to the heuristic optimisation method and the use of the DNL model in the method itself. In this section the calculation times needed to run the framework for the anticipatory control strategy are investigated.

The cases discussed in the previous sections are all relatively small cases. In table 7.12 the network characteristics of all cases discussed are summarised. The characteristics are the number of time steps in the DNL model (# steps), the number of links (# links), the number of nodes (#nodes), the number of OD pairs (#OD) and the number of routes (# routes). These characteristics determine the calculation time for one run of the DNL model.

The table also shows some characteristics of the evolutionary algorithm, used for the optimisation in the anticipatory control strategy, which are important for the total calculation time. These characteristics are the number of parameters to optimise (#param), the number of iterations in the main assignment loop (# iter) and the number of total control plan evaluations (#planeval). Note that for anticipatory control one control plan evaluation contains two runs of the DNL model and one assignment.

In the last two columns, the table shows the CPU time to run the anticipatory control strategy (AC) and, for comparison reasons, the CPU time to run the Webster local control strategy (LC). The CPU time is measured on a laptop with an Intel® Core^{TM}2 Duo T7100 processor at 1.80 GHz and 2 GB internal memory.

Table 7.12: Calculation characteristics for all cases

Cases	Network characteristics					Optimisation characteristics			CPU time (min)	
	# steps	# links	# nodes	# OD	# routes	# param	# iter	# planeval	AC	LC
Case 1	225	7	7	2	3	10	14	1428	2.25	0.06
Case 2	225	9	8	2	3	20	12	1320	2.18	0.06
Case 3	150	16	12	2	4	40	15	3180	4.28	0.05
Case 4	150	9	8	1	3	20	15	1650	1.85	0.05
Case 5a	270	14	12	3	4	30	15	2340	4.95	0.06
Case 5b	270	17	14	7	13	40	14	3388	8.14	0.06
Case 6	225	12	12	3	4	15	8	880	1.59	0.05
Case 7	270	17	14	7	13	48	16	3872	9.28	0.06

From the table we conclude that the CPU time increases rapidly if the number of parameters to optimise increases. This number determines how many evaluations are needed and the larger the number of evaluations, the larger the CPU time. Generally, the number of parameters to optimise is dependent on the size of the network, the number of control measures and the number of time periods in the simulation. The size of the network determines also the time needed to run the DNL model, which is done twice in the anticipatory control strategy.

So, we have two effects of the size of the network on the calculation time, which reinforce each other. Therefore, we can conclude that the framework can be applied to larger network, but it will be time consuming. And, in this case, with 'time' we refer to CPU time.

7.6   Summary

In this chapter the framework for anticipatory control was used to study a number of cases. The first case is the most simple one: only route choice between two routes for one OD pair and only one controlled intersection. This case was used to show that the concept of anticipatory control works and to test its robustness for a number of important variables, such as assignment type and demand. For these variables different values were specified to see if the results of the default scenario still hold. And the result is that in all scenarios anticipatory control performs better (in terms of total delay) than the other control strategies used, except for the system optimum, which is more a kind of benchmark.

Then three other small cases were simulated to test if the improvements found for the first case are also valid for somewhat more complex cases. This is indeed the case, although it can be stated that the results do not improve much for the symmetric networks investigated and the improvements are the largest if control and/or route choice possibilities increase.

If we add ramp metering to the existing traffic control we get cases with integrated traffic management. Also for these cases anticipatory control gives the best results in terms of total delay. If traffic management is not integrated, but distributed among different road operators, we get cases with more than two players. Two cases with three players were simulated and also for these cases integrated anticipatory control was the best, especially if the number of control options increased.

The conclusion of this chapter is that the framework for anticipatory control can be used to test all kinds of control strategies for all kinds of cases. For the cases studied, anticipatory control gives the best results in almost all circumstances and these results are also robust with respect to a route choice parameter and demand. However, calculation time remains an issue to study further. For larger networks it increases rapidly.

8

Conclusions and Further Research

This final chapter summarises the work of this thesis and lists the conclusions, which can be derived from the results. Furthermore, it shows the contributions of this work to modelling practise and finally it gives directions for further research.

8.1 Conclusions

In this thesis a framework for anticipatory control has been developed. Within this framework, a dynamic network loading (DNL) model and a dynamic traffic assignment model are important parts to test the performance of the anticipatory control strategy. Before the framework was applied to several cases, the models were validated. Therefore, the conclusions are listed into the categories: framework, assignment models, DNL model, validation and application. Overall it can be concluded that anticipatory control is a promising strategy to deal with network control. For all the cases studied, anticipatory control showed decreases in total delay and this decrease was robust for demand and some control parameters.

8.1.1 Framework

The first conclusion is that game theory is useful to describe different control strategies. It is shown that traditional traffic control, in which control parameters are adapted to changing traffic conditions, is related to the Nash game or Cournot game, in which each player reacts on the moves of other players. Anticipatory control is related to the Stackelberg game, in which one or more players can anticipate the moves of other players if they have perfect knowledge

about how players react. System optimum control can be seen as a Monopoly game, in which everyone chooses the moves that make the system profit the most.

Secondly, from the tested networks, it can be concluded, that for anticipatory control it is enough to predict route choice one day ahead. In the framework, control is optimised by predicting ahead a few iterations in the route choice process. An iteration in our route choice process can be considered to be a day of route choice experience. It was found that making a prediction for the next iteration (so one day ahead) gives comparable results with making a prediction for the equilibrium situation. Whether this conclusion is valid for other networks should be investigated.

For anticipatory control, the resulting bi-level optimisation problem is heuristically solved by an evolutionary algorithm. The conclusion is that the CMA-ES algorithm performed better compared with another genetic algorithm: it gave better final solutions (less total delay) and less stochastic variation in less computation time. Although the computation time can be problematic for larger networks.

Another conclusion related to the framework is that it is suitable not only to study the interaction between two players, but also the interaction between three or even more players. In this thesis the example of two road authorities and the road users is given, but the framework allows extensions to more players.

Finally, we can conclude that the framework can be applied not only for intersection control, but also for other types of control, such as ramp metering. In the framework all control types can be integrated and the effects of these integrated traffic control measures can be determined.

8.1.2   Assignment models

For the deterministic dynamic user optimal assignment the variational inequality approach leads to a quadratic programming optimisation problem. One parameter in this problem is the contraction factor ρ. The conclusion is that a variable ρ, dependent on the route flows and costs, leads to faster convergence than a fixed ρ.

In the stochastic dynamic user optimal assignment flows of an iteration are smoothed with flows from the previous iteration. The parameters for this smoothing process are important for convergence. It can be concluded that the adjusted smoothing factors proposed in this thesis, have a faster convergence than the traditional factors.

Another conclusion related to convergence is the convergence criterion. To determine convergence, several criteria can be used. From the results in this thesis, we conclude that a convergence criterion, based on the maximum change in flows from one iteration to the other, leads to faster convergence than the duality gap.

Solving the system optimum assignment problem analytically is difficult, especially when it is combined with optimal traffic control. It can be concluded that using an evolutionary algorithm is a suitable method to solve this problem, provided that the problem is not too large, otherwise computation time becomes a problem.

8.1.3 DNL model

Due to the assumption of an uniform distribution of traffic along a link, problems with congestion and blocking back can occur. This has to do with the propagation of changes in inflow and outflow (due to congestion downstream) along the link. We conclude that the step mechanism developed in this thesis is a workable method to solve this problem. The step mechanism makes sure that differences in inflow and outflow reach the end of the link in time.

In DNL models a problem with short links occurs if the time step is large. Normally, traffic is not allowed to traverse a link within one time step. We conclude that some adjustments in the DNL model for this type of links is enough to relax this condition. It is shown that these adjustments do not affect propagation of traffic much.

8.1.4 Validation

The DNL model was calibrated and validated for a motorway bottleneck with real-life data. The results showed that the model was able to reproduce the congested situation fairly accurate: the flows, travel times and network indicators show small goodness-of-fit errors. Therefore, the conclusion is that the DNL model gives a simple, but adequate representation of real-life traffic conditions.

Before the complete framework with local control was tested for a medium-sized real network, a method to estimate a dynamic OD matrix was developed. The method is based on the theory described by Van Zuylen (1981). It can be concluded that this method is suitable for the estimation of dynamic OD matrices. It is shown that it reduces the difference between modelled and measured flows and travel times.

For the validation of a larger real-life network, a situation with roadworks was used. The network and OD matrix were adjusted and the assignment was run. The conclusion is that the results for the flows are less good than for the calibration data, but, taking into account the assumptions made, the results are satisfactory, especially for the travel times.

8.1.5 Application

From the case studies it can be concluded that anticipatory control decreases the total delay in the network. The decrease compared with local control strategies varies from $\pm$ 4% to $\pm$ 40%. This positive result is robust in relation with demand and assignment parameters. From the case studies it can also be concluded that anticipatory control works best if there are enough control and route choice possibilities.

The conclusions about the results of anticipatory control we have drawn so far, are also valid if different traffic control measures are integrated in the framework for anticipatory control. The results show that also in these cases anticipatory control has the smallest total delay in the network.

The framework described in this thesis is suitable to study the interaction between different road operators and road users. We did that for two cases with two road operators. The conclusion is that in both cases anticipatory control was better, again especially when the number of control options and route choice increased. Cooperation between the road operators was only beneficial in one case, the more complex one.

Concerning the calculation time, the size of the network has the largest influence on the calculation time. For anticipatory control this influence is strengthened, because of the use of an heuristic method and the use of the model in this method. Therefore, we conclude that the framework can be applied to larger networks, but it will be time (CPU time) consuming.

8.2   Main contributions

The main contribution of this thesis is a framework for developing, testing and evaluating all kinds of network control strategies. These control strategies do not limit themselves to traffic signal control only, but can include other control measures, such as ramp metering or tolling, as well. The combination with a dynamic traffic assignment model gives the opportunity to study the impact of traffic management on network conditions and especially route choice.

The second main contribution is the development of a network control strategy that anticipates route choice behaviour of road users and includes multiple network types (and accompanying road authorities) and multiple control measures. This control strategy could form the basis of integrated, network-wide traffic management.

8.3   Further research

The framework described in this thesis is far from complete. A number of aspects are open for further research. In this section we discuss these aspects, divided into the three main components of our work: the framework, the assignment models and the dynamic network loading models.

8.3.1   Research aspects for the framework

- *Optimisation for multiple criteria.* In the optimisation problem for the anticipatory control strategy, formulated in section 3.4.2, travel costs are used as an optimisation criterion. In our work travel costs are restricted to travel time, but it is possible to extend travel costs with safety costs, environmental costs, comfort costs and tolling (see next subsection). This could be done for the optimisation in the anticipatory control strategy only, or also for the assignment in the main loop.

- *Simulation in the loop*. The framework uses the same DNL model for the optimisation of the traffic control plan as for the evaluation of this control plan in the overall assignment loop. To avoid that the assignment in the optimisation loop profits from the assignment in the main loop, a different model should be used to evaluate the control strategy. In Taale and Van Zuylen (2003b) this is done using the microscopic simulation model FLEXSYT-II- (Taale and Middelham, 1995, 1997). In this article it is concluded that the results if a microscopic simulation model is used for the evaluation, are consistent with the results if the DNL model is used. Further research should investigate this conclusion for other networks and multiple players.

- *Intersection control*. The networks in our examples all use a simple control strategy: an arm of the intersection is controlled by one phase. Normally, intersection control is more complex, at least in The Netherlands. Every movement is controlled separately with a signal and besides that, pedestrians, bicycles and public transport make intersection control even more complex. Further research should try to make the connection between anticipatory control and real-life, signal-based traffic control. A research direction could be that anticipatory control gives the constraints for the local intersection control. The local control strategy then takes these constrains into account while dealing with the local traffic situation. Also, the comparison of anticipatory control with network control systems such as SCATS and SCOOT, is a challenging topic.

- *Faster optimisation*. The applicability of the framework depends for a large part on the calculation time needed. These calculation times can be long if the network size increases. This is partially due to the optimisation method used, which is a heuristic one. Therefore, it will be interesting to investigate if it is possible to use faster optimisation algorithms to solve the anticipatory control problem or the solve the optimisation problem using parallel computing.

8.3.2   Research aspects for the assignment models

- *Assignment properties*. In chapter 4 a number of properties for the deterministic and stochastic assignment, related to convergence, were investigated. The results for an example network showed that the assignment process can convergence faster if adjustments to step size factors and convergence criteria are made. The question is raised if these improvements also hold for other, small and large, networks.

- *Departure time choice*. Anticipatory control anticipates the reaction of travellers on changes in travel characteristics. In the work of this thesis changes in travel time are used to anticipate route choice. This could be extended to departure time choice, both within the main assignment loop, as within the optimisation process for anticipatory control. For that, a change in the definition of an equilibrium is needed, to allow for choice of the departure period. Also, the specification of travel costs should be extended by penalties on departure and/or arrival. A useful formulation for departure time choice has been given by Bliemer (2001).

- *Tolling*. Another extension in the definition of travel costs is possible with tolling. Tolling can be applied in several ways, e.g. a tolling station or a fee per kilometre driven. For all tolling options, the fee can be converted into time using a value-of-time parameter and added to the normal travel time. In the assignment this generalised 'travel time' can be used for the route flow distribution. Tolling can be considered a control measure and it would be interesting to investigate the control strategies in this thesis, in combination with departure time choice and/or elastic demand. The effects of different control strategies on demand (induced or suppressed demand) can be neglected. This is different if a combination with tolling is made. Smith (2006) would be a good start for this investigation.

- *Multiple user classes*. In this thesis only one user class has been considered. In the literature different classifications of user classes can be found. The most popular ones are (Bliemer, 2001):

  - level of information;
  - vehicle type;
  - driver properties;
  - network access restrictions;
  - travel purpose and/or destination.

In this thesis only the distinction in level of information is made. The level of information is varied by the parameter θ in the stochastic assignment. In the model several user classes with a different value of θ can be specified. In section 7.1 a small network is simulated using one user class with different values of θ. In Taale and Van Zuylen (2003b) a mix of user classes is used. Further research should focus on the distinction between vehicle types, because the interaction between vehicle types is important for the traffic situation on roads (again see Bliemer (2001).

8.3.3   Research aspects for the network loading model

- *Travel time functions*. In the dynamic network loading (DNL) model traffic is moved through the network using link travel time functions. Different functions for different link types are used. Recent research by Viti (2006) has shown that for signal controlled intersections improvements in the estimation of travel times can be obtained. This improved travel time function can be implemented in the DNL model of this thesis and the consequences in terms of travel times, but also calculation time, can be investigated. Also, the results of the different control strategies can be simulated.

- *Reliability*. In his PhD thesis Tu (2008) states that travel time reliability has a significant effect on route choice. He specified an extra term, to be added to the travel cost function, representing the costs of unreliability. It is easy to add this term to the travel cost function used in the DNL model in this thesis and to investigate what the effects are on route choice and on the control strategies. The question to answer will be how anticipatory control will react if reliability is included.

Bibliography

Abdelfatah, A.S. and H.S. Mahmassani (1998). System Optimal Time-Dependent Path Assignment and Signal Timing in a Traffic Network. *Proceedings of the 77th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Abdelfatah, A.S. and H.S. Mahmassani (2000). Time-Dependent Signal Control and System Optimal Assignment in Congested Networks: Algorithm Implementation and Numerical Results. *Proceedings of the 79th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Abdelghany, K.F., D.M. Valdes, A.S. Abdelfatah and H.S. Mahmassani (1999). Real-Time Dynamic Traffic Assignment and Path-Based Signal Coordination: Application to Network Traffic Management. *Proceedings of the 78th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Akçelik, R. (1989). *Traffic Signals: Capacity and Timing Analysis*. Technical Report ARR No. 123, Australian Road Research Board.

Akçelik, R. (1991). Travel Time Functions for Transport Planning Purposes: Davidson's Function, its Time-Dependent Form and an Alternative Travel Time Function. *Australian Road Research*, Vol. 21, No. 3, pp. 49–59. Minor revisions: December 2000.

Akçelik, R. (2003). *Speed-Flow Models for Uninterrupted Traffic Facilities*. Technical Report, Akcelik and Associates Pty Ltd, Melbourne, Australia.

Akçelik, R. and M.J. Maher (1977). Route Control of Traffic in Urban Road Networks. *Transportation Research*, Vol. 11, No. 1, pp. 15–24.

Allsop, R.E. (1974). Some Possibilities for Using Traffic Control to Influence Trip Distribution and Route Choice. *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*. Sydney, Australia.

Allsop, R.E. and J.A. Charlesworth (1977). Traffic in a Signal-Controlled Road Network: An Example of Different Signal Timings Inducing Different Routeings. *Traffic Engineering + Control*, Vol. 18, No. 5, pp. 262–264.

Arcadis (2001). *Results Modeling Study A10-West. From Current Situation to Scenarios*. Report, AVV Transport Research Centre. (in Dutch).

Atkins (2007). SATURN. Website: www.saturnsoftware.co.uk/7.html, accessed on October 1st, 2007.

Bäck, Th. (1996). *Evolutionary Algorithms in Theory and Practise: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, New York, U.S.A.

Balakrishna, R. (2002). *Calibration of the Demand Simulator in a Dynamic Traffic Assignment System*. Master's thesis, Massachusetts Institute of Technology, Boston, U.S.A.

Balakrishna, R., H.N. Koutsopoulos and M. Ben-Akiva (2005). Calibration and Validation of Dynamic Traffic Assignment Systems. In: H.S. Mahmassani (ed.), *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, pp. 407–426. Maryland, U.S.A.

Başar, T. and G. J. Olsder (1999). *Dynamic Noncooperative Game Theory*. Classics in Applied Mathematics 23. Society for Industrial and Applied Mathematics, Philadelphia, U.S.A.

Bazaraa, M.S., H.D. Sherali and C.M. Shetty (1993). *Nonlinear Programming: Theory and Algorithms*. Second Edition. John Wiley & Sons Inc., Singapore.

Bell, M.G.H., C.M. Shield, F. Busch and G. Kruse (1997). Stochastic User Equilibrium Path Flow Estimator. *Transportation Research Part C*, Vol. 5, No. 3-4, pp. 197–210.

Bindel, D. (2004). Implementation of Dijkstra's Method Using a Matlab Sparse Matrix as an Adjacency Matrix. C-source file, found at www.cs.berkeley.edu/~dbindel, accessed on December 20th, 2004.

Bliek, C., P. Spellucci, L.N. Vicente, A. Neumaier, L. Granvilliers, E. Huens, P. Van Hentenryck, D. Sam-Haroud and B. Faltings (2001). *Algorithms for Solving Nonlinear Constrained and Optimization Problems: The State of The Art*. COCONUT Deliverable D1, The Coconut Project.

Bliemer, M.C.J. (2001). *Analytical Dynamic Traffic Assignment with Interacting User-Classes*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Bliemer, M. and H. Taale (2006). Route Generation and Dynamic Traffic Assignment for Large Networks. *Proceedings of the First International Symposium on Dynamic Traffic Assignment*. Leeds, U.K.

Bogers, E.A.I., M. Bierlaire and S.P. Hoogendoorn (2006a). Modeling Learning in Route Choice. *Proceedings of the 86th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Bogers, E.A.I., F. Viti, S.P. Hoogendoorn and H.J. Van Zuylen (2006b). Valuation of Different Types of Travel Time Reliability in Route Choice : Large-Scale Laboratory Experiment. *Transportation Research Record*, No. 1985, pp. 612–170.

Bossom, R., V. Avontuur, J.-F. Gaillet, G. Franco and P. Jesty (2000). *European ITS Framework Architecture Overview*. Deliverable D3.6, KAREN project, part of the EC Framework IV - Telematics for Transport programme.

Cairns, S., C. Hass-Klau and P. Goodwin (1998). *Traffic Impact of Highway Capacity Reductions: Assessment of the Evidence*. Landor Publishing, London, United Kingdom.

Cantarella, G.E., G. Improta and A. Sforza (1991). Iterative Procedure for Equilibrium Network Traffic Signal Setting. *Transportation Research Part A*, Vol. 25, No. 5, pp. 241–249.

Cascetta, E., M. Gallo and B. Montella (1998a). An Asymmetric SUE Model for the Combined Assignment-Control Problem. *Proceedings of the 8th World Conference on Transport Research*. Antwerp, Belgium.

Cascetta, E., M. Gallo and B. Montella (1998b). Optimal Signal Setting on Traffic Networks with Stochastic Equilibrium Assignment. *Proceedings of TRISTAN III*. Puerto Rico.

Cascetta, E., A. Nuzzolo, F. Russo and A. Vitetta (1996). A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and some Calibration Results for Interurban Networks. *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pp. 697–711. Lyon, France.

Černý, V. (1985). A Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm. *Journal of Optimization Theory and Applications*, Vol. 45, No. 1, pp. 41–51.

Ceylan, H. and M.G.H. Bell (2004). Traffic Signal Timing Optimisation Based on Genetic Algorithm Approach, Including Drivers' Routing. *Transportation Research Part B*, Vol. 38, No. 4, pp. 329–342.

Ceylan, H. and M.G.H. Bell (2005). Genetic Algorithm Solution for the Stochastic Equilibrium Transportation Networks under Congestion. *Transportation Research Part B*, Vol. 39, No. 2, pp. 169–185.

Charlesworth, J.A. (1977). The Calculation of Mutually Consistent Signal Settings and Traffic Assignment for a Signal-Controlled Road Network. *Proceedings of the 7th International Symposium on Transportation and Traffic Theory*. Tokyo, Japan.

Chen, H.-K. (1999). *Dynamic Travel Choice Models: A Variational Inequality Approach*. Springer, Heidelberg, Germany.

Chen, H.-K. and C.-F. Hsueh (1997). Combining Signal Timing Plan and Dynamic Traffic Assignment. *Proceedings of the 76th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Chen, O.J. (1998). *Integration of Dynamic Traffic Control and Assignment*. Ph.D. thesis, Massachusetts Institute of Technology, Boston, U.S.A.

Chiou, S.-W. (1997). Optimisation of Area Traffic Control Subject to User Equilibrium Traffic Assignment. *Proceedings of the 25th European Transport Forum*. Brunel University, United Kingdom.

Chiou, S.-W. (1998). Optimisation of Traffic Signal Timings for User Equilibrium Flow. *Proceedings of the Third International Symposium on Highway Capacity*. Copenhagen, Denmark.

Chiou, S.-W. (1999). Optimization of Area Traffic Control for Equilibrium Network Flows. *Transportation Science*, Vol. 33, No. 3, pp. 279–289.

Chiou, S.-W. (2004). Solution Algorithms for Equilibrium Network Design with Signal Settings. *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Cipriani, E. and G. Fusco (2004). Combined Signal Setting Design and Traffic Assignment Problem. *European Journal of Operational Research*, Vol. 155, No. 3, pp. 569–583.

Clegg, J., A. Clune, M. Smith and Y. Xiang (1999). A Decision Support System Based on the Cone Projection Method - Theory and First Results. *Proceedings of the 11th Mini-EURO Conference on AI in Transportation Systems and Science*. Helsinki, Finland.

Clegg, J., M. Smith, Y. Xiang and R. Yarrow (2001). Bilevel Programming Applied to Optimising Urban Transportation. *Transportation Research Part B*, Vol. 35, No. 1, pp. 41–70.

Clement, S. and J. Anderson (1997). Traffic Signal Timing Determination: The Cabal Model. *Proceedings of the Second International Conference on Genetic Algorithms in Engineering Systems*, pp. 63–68. Glasgow, United Kingdom.

Davidson, K.B. (1966). A FlowTravel Time Relationship for Use in Transportation Planning. *Proceedings of the 3rd ARRB Conference*, Part 1, pp. 183–194.

Davidson, K.B. (1978). The theoretical basis of a flow-travel time relationship for use in transportation planning. *Australian Road Research*, Vol. 8, No. 1, pp. 32–35.

De Romph, E. (1994). *A Dynamic Traffic Assignment Model. Theory and Applications*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Di Taranto, C. and V. Mauro (1989). UTOPIA. *Proceedings of the 6th IFAC/IFIP/IFORS Symposium on Control, Computers, Communications in Transportation*, pp. 575–597. Paris, France.

Dickson, T.J. (1981). A Note on Traffic Assignment and Signal Timings in a Signal-Controlled Road Network. *Transportation Research Part B*, Vol. 15, No. 4, pp. 267–271.

Eiben, A.E. (2002). Evolutionary Computing: The Most Powerful Problem Solver in the Universe? *Dutch Mathematical Archive*, Vol. 5/3, No. 2, pp. 126–131.

Eiben, A.E. and J.E. Smith (2003). *Introduction to Evolutionary Computing*. Springer, Berlin.

Eppstein, D. (1998). Finding the K Shortest Paths. *SIAM Journal of Computing*, Vol. 28, No. 2, pp. 652–673.

Fisk, C.S. (1984). Game Theory and Transportation Systems Modelling. *Transportation Research Part B*, Vol. 18, No. 4/5, pp. 301–313.

Florian, M. and H. Spiess (1982). The Convergence of Diagonalization Algorithms for Asymmetric Network Equilibrium Problems. *Transportation Research Part B*, Vol. 16, No. 6, pp. 477–483.

Foy, M.D., R.F. Benekohal and D.E. Goldberg (1992). Signal Timing Determination Using Genetic Algorithms. *Transportation Research Record*, No. 1365, pp. 108–115.

FRAME Project, Part of the EC Framework V Information Society Technologies Programme (2007). European ITS Architecture. Website: http://www.frame-online.net/home.htm, accessed on July 30th, 2007.

Fudenberg, D. and J. Tirole (1993). *Game Theory*. MIT Press, Cambridge, Massachussetts, U.S.A.

Gartner, N.H. (1977). Influence of Control Measures on Traffic Equilibrium in an Urban Highway Network. *Transportation Research Record*, No. 644, pp. 125–129.

Gartner, N.H. (1983). OPAC: A Demand-Responsive Strategy for Traffic Signal Control. *Transportation Research Record*, Vol. 906, pp. 75–81.

Gartner, N.H. and M. Al-Malik (1996). Combined Model for Signal Control and Route Choice in Urban Traffic Networks. *Transportation Research Record*, No. 1554, pp. 27–35.

Gartner, N.H. and M. Al-Malik (1997). Combined Control and Route Assignment in Traffic Signal Networks. *Proceedings of the 8th IFAC Symposium on Transportation Systems*. Chania, Crete, Greece.

Gartner, N.H., S.B. Gershwin, J.D.C. Little and P. Ross (1980). Pilot Study of Computer-Based Urban Traffic Management. *Transportation Research Part B*, Vol. 14, pp. 203–217.

Gartner, N.H., C. Stamatiadis and P.J. Tarnoff (1994). Development of Real Time, Traffic-Adaptive Control Strategies for IVHS. *Proceedings of the First World Congress on ATT and IVHS*, Volume 2, pp. 423–430. Paris, France.

Gershwin, S.B., P. Ross, N.H. Gartner and J.D.C. Little (1978). Optimization of Large Traffic Systems. *Transportation Research Record*, No. 682, pp. 8–15.

Ghali, M.O. and M.J. Smith (1993). Traffic Assignment, Traffic Control and Road Pricing. *Proceedings of the 12th International Symposium on Transportation and Traffic Theory*. Berkeley, California, U.S.A.

Giezen, J. and V. Avontuur (2002). *Inter-Urban Traffic Management in AVB and KAREN - AVB-KAREN Compliance Check and an AVB Example System in KAREN Style*. TNO report RPT-020046, TNO TPD, for AVV Transport Research Centre.

Goodwin, P., C. Hass-Klau and S. Cairns (1998). Evidence on the Effects of Road Capacity Reduction on Traffic Levels. *Traffic Engineering + Control*, Vol. 39, No. 6, pp. 348–354.

Hadi, M.A. and C.E. Wallace (194). Optimization of Signal Phasing and Timing Using Cauchy Simulated Annealing. *Transportation Research Record*, No. 1456, pp. 64–71.

Hadi, M.A. and C.E. Wallace (1993). Hybrid Genetic Algorithm to Optimize Signal Phasing and Timing. *Transportation Research Record*, No. 1421, pp. 104–112.

Hansen, N. (2006). The CMA Evolution Strategy: A Comparing Review. In: J.A. Lozano *et al.* (eds.), *Towards a New Evolutionary Computation. Advances in Estimation of Distribution Algorithms*, pp. 75–102. Springer-Verlag, Berlin.

Hansen, N., S.D. Müller and P. Koumoutsakos (2003). Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, Vol. 11, No. 1, pp. 1–18.

Hansen, N. and A. Ostermeier (1996). Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation. *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation*, pp. 312–317. Aachen, Germany.

Hansen, N. and A. Ostermeier (2001). Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, Vol. 9, No. 2, pp. 159–195.

Hawas, Y.E. (2000). Integrated Traffic Assignment and Signal Control for On-Line Operation. *Proceedings of the 7th World Congress on Intelligent Transport Systems*. Turin, Italy.

Head, K.L., P.B. Mirchandani and D. Sheppard (1992). Hierarchical Framework for Real-Time Traffic Control. *Transportation Research Record*, Vol. 1360, pp. 82–88.

Henry, J.J. and J.L. Farges (1989). PRODYN. *Proceedings of the 6th IFAC/IFIP/IFORS Symposium on Control, Computers, Communications in Transportation*, pp. 505–507. Paris, France.

Heydecker, B.G. (1983). Some Consequences of Detailed Junction Modelling in Road Traffic Assignment. *Transportation Science*, Vol. 17, No. 3, pp. 263–281.

Hoogendoorn-Lanser, S., R. Van Nes and P. Bovy (2005). Path-Size Modeling in Multi-Modal Route Choice Analysis. *Proceedings of the 84th Annual Meeting of the Transportation Research Board*. Washington, D.C., U.S.A.

Houck, C.R., J.A. Joines and M.G. Kay (1995). *A Genetic Algorithm for Function Optimization: A Matlab Implementation*. NCSU-IE Technical Report 95-09, North Carolina State University.

Hourdakis, J., P.G. Michalopoulos and J. Kottommannil (2003). Practical Procedure for Calibrating Microscopic Traffic Simulation Models. *Transportation Research Record*, No. 1852, pp. 130–139.

Hunt, P.B., D.I. Robertson, R.D. Bretherton and R.I. Winton (1981). *SCOOT - a Traffic Responsive Method of Co-Ordinating Signals*. TRRL Report 1014, TRRL.

Institute of Transportation Engineers (1992). *Results of the 1991 North American Traffic Signal Inventory*. Report, Institute of Transportation Engineers, Washington, D.C.

Jayakrishnan, R., H.S. Mahmassani and T.-Y. Hu (1994). An Evaluation Tool for Advance Traffic Information and Management Systems in Urban Networks. *Transportation Research Part C*, Vol. 2, No. 3, pp. 129–147.

Joksimovic, D. (2007). *Dynamic Bi-Level Optimal Toll Design Approach for Dynamic Traffic Networks*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Kelman, Les (2004). A History of Traffic Control In Toronto. Website: www.signalsystems.org.vt.edu/documents.html, accessed on April 29th, 2006.

Kirkpatrick, S., C.D. Gelatt Jr. and M.P. Vecchi (1983). Optimization by Simulated Annealing. *Science*, Vol. 220, No. 4598, pp. 671–680.

Kock, R.W. and E. Van Den Hoogen (2002). Frames of Reference - Quantified Policy Objectives for Effective Network Management. *Proceedings of the 11th International Conference on Road Transport Information and Control*, pp. 126–130. Institute of Electrical Engineers, London, United Kingdom.

Lee, C. and R.B. Machemehl (1998). Combining Route Choice and Signal Control: Algorithms, Numerical Results and its Implication to Advanced Traffic Management Systems. *Proceedings of the 5th World Congress on Intelligent Transport Systems*. Seoel, Korea.

Lee, C. and R.B. Machemehl (1999). Local & Iterative Searches for the Combined Signal Control and Assignment Problem: Implementation and Numerical Results. *Proceedings of the 78th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Lee, S. and M. Hazelton (1996). Stochastic Optimisation of Combined Traffic and Signal Control Junction Modelling. *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Lyon, France.

Leonard, D.R., P. Gower and N.B. Taylor (1989). *CONTRAM: Structure of the Model*. TRL Report RR 178, Transport Research Laboratory, Crowthorne, U.K.

Li, H. (2005). *Calibration and Validation of MARPLE*. Technical Report, AVV Transport Research Centre.

Li, H., P.H.L. Bovy and G. Hooghiemstra (2007). *Probabilistic Properties of Stochastic Choice Set Generation*. TRAIL Studies in Transportation Science No. S2007/1, TRAIL.

Li, Z. and S. Shi (2003). A Differential Game Modeling Approach to Dynamic Traffic Assignment and Signal Control. *European Journal of Transport and Infrastructure Research*, Vol. 3, No. 3, pp. 281–297.

Lowrie, P.R. (1982). SCATS - Principles, Methodology and Algorithms. *Proceedings of the International Conference on Road Traffic Signalling*. London, United Kingdom.

Maher, M.J. and R. Akçelik (1975). The Re-Distributional Effects of an Area Traffic Control Policy. *Traffic Engineering + Control*, Vol. 16, No. 9, pp. 383–385.

Maher, M.J. and R. Akçelik (1977). Route Control - Simulation Experiments. *Transportation Research*, Vol. 11, No. 1, pp. 25–31.

Maher, M.J., X. Zhang and D. Van Vliet (2001). A Bi-Level Programming Approach for Trip Matrix Estimation and Traffic Control Problems with Stochastic User Equilibrium Link Flows. *Transportation Research Part B*, Vol. 35, No. 1, pp. 23–40.

Marcotte, P. (1983). Network Optimization with Continuous Control Parameters. *Transportation Science*, Vol. 17, No. 2, pp. 181–197.

Meneguzzer, C. (1995). An Equilibrium Route Choice Model with Explicit Treatment of the Effect of Intersections. *Transportation Research Part B*, Vol. 29, No. 5, pp. 329–356.

MinVenW (2005). *Mobility Policy Document – Towards reliable and predictable accessibility*. Report, Ministry of Transport, Public Works and Water Management.

MinVenW (2008). *Policy Framework for Utilisation*. Report, Ministry of Transport, Public Works and Water Management.

Montana, D.J. and S. Czerwinski (1996). Evolving Control Laws for a Network of Traffic Signals. *Genetic Programming 1996*, pp. 333–338. The MIT Press, Cambridge, Massachusetts, U.S.A.

Mordridge, M.J.H. (1997). The Self-Defeating Nature of Urban Road Capacity Policy; a Review of Theories, Disputes and Available Evidence. *Transport Policy*, Vol. 4, No. 1, pp. 5–24.

Muller, Theo H.J. and Martijn de Leeuw (2006). A New Method to Design Traffic Control Programs. *Proceedings of the 85th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Nagurney, A. (1993). *Network Economics: A Variational Inequality Approach*. Kluwer Academic Publishers, Boston, MA, U.S.A.

Neumaier, A. (2004). Complete Search in Continuous Global Optimization and Constraint Satisfaction. *Acta Numerica*, Vol. 13, pp. 271–369.

Nihan, N., M. Hamed and G. Davis (1995). Interactions Between Driver Information, Route Choice, and Optimal Signal Timing on a Simple Network. *Journal of Advanced Transportation*, Vol. 29, No. 2, pp. 163–182.

Peeta, Srinivas (1994). *System Optimal Dynamic Traffic Assignment in Congested Networks with Advanced Information Systems*. Ph.D. thesis, University of Texas, Austin, Texas, U.S.A.

Ramming, M.S. (2002). *Network Knowledge and Route Choice*. Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA, U.S.A.

Rijkswaterstaat (2001). *Traffic Management Architecture - Traffic Control Architecture*. AVV Transport Research Centre, Arcadis and TNO Inro, Rotterdam, The Netherlands. (in Dutch).

Rijkwaterstaat (2003). *Handbook Sustainable Traffic Management*. AVV Transport Research Centre, Rotterdam, The Netherlands.

Roberston, D.I. (1969). *TRANSYT: A Traffic Network Study Tool*. Road Research Laboratory Report LR 253, Road Research Laboratory, Crowthorne, United Kingdom.

Rouphail, N., A. Tarko and J. Li (2002). Traffic Flow at Signalized Intersections. *Traffic Flow Theory, A State of the Art Report*, chapter 9. Federal Highway Administration. Website: www.tfhrc.gov/its/tft/tft.htm, accessed on January 17th, 2005.

Schuurman, H. (2003). BOSS Launched: Predictions for the Motorway. *Verkeerskunde*, Vol. 54, No. 4, pp. 20–25. (in Dutch).

Schwarz, S. (1999). GM and the Red Cars. Website: http://1134.org/stan/ul/GM-et-al.html, accessed on July 31st, 2007.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, N.J., U.S.A.

Sheffi, Y., A.M. Asce and W.B. Powell (1983). Optimal Signal Settings over Transportation Networks. *Journal of Transportation Engineering*, Vol. 109, No. 6, pp. 824–839.

Smith, M.J. (1979a). The Existence, Uniqueness and Stability of Traffic Equilibria. *Transportation Research Part B*, Vol. 13, No. 4, pp. 295–304.

Smith, M.J. (1979b). Traffic Control and Route-Choice: A Simple Example. *Transportation Research Part B*, Vol. 13, No. 4, pp. 289–294.

Smith, M.J. (1980). A Local Traffic Control Policy Which Automatically Maximises the Overall Travel Capacity of an Urban Road Network. *Traffic Engineering + Control*, Vol. 21, No. 6, pp. 298–302.

Smith, M.J. (1981a). The Existence of an Equilibrium Solution to the Traffic Assignment Problem When There are Junction Interactions. *Transportation Research Part B*, Vol. 15, No. 6, pp. 443–451.

Smith, M.J. (1981b). Properties of a Traffic Control Policy Which Ensures the Existence of a Traffic Equilibrium Consistent with the Policy. *Transportation Research Part B*, Vol. 15, No. 6, pp. 453–462.

Smith, M.J. (1981c). A Theoretical Study of Traffic Assignment and Traffic Control. *Proceedings of the 8th International Symposium on Transportation and Traffic Theory*. Toronto, Canada.

Smith, M.J. (1982). Junction Interactions and Monotonicity in Traffic Assignment. *Transportation Research Part B*, Vol. 16, No. 1, pp. 1–3.

Smith, M.J. (1985). Traffic Signals in Assignment. *Transportation Research Part B*, Vol. 19, No. 2, pp. 155–160.

Smith, M.J. (1987). Traffic Control and Traffic Assignment in a Signal-Controlled Network with Queueing. *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*. Cambridge, Massachusetts, U.S.A.

Smith, M.J. (2006). Bilevel optimisation of prices and signals in transportation models. In: S. Lawphongpanich *et al.* (eds.), *Mathematical and computational models for congestion charging*, pp. 159–200. Springer.

Smith, M.J. and M.O. Ghali (1990a). Dynamic Traffic Assignment and Dynamic Traffic Control. *Proceedings of the 11th International Symposium on Transportation and Traffic Theory*. Yokohama, Japan.

Smith, M.J. and M.O. Ghali (1990b). The Dynamics of Traffic Assignment and Traffic Control: A Theoretical Study. *Transportation Research Part B*, Vol. 24, No. 6, pp. 409–422.

Smith, M.J. and T. Van Vuren (1993). Traffic Equilibrium with Responsive Traffic Control. *Transportation Science*, Vol. 27, No. 2, pp. 118–132.

Smith, M.J., T. Van Vuren, B.G. Heydecker and D. Van Vliet (1987). The Interactions Between Signal Control Policies and Route Choice. *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*. Cambridge, Massachusetts, USA.

Smith, M.J., Y. Xiang and J. Clegg (1998). Bilevel Optimization of Road Traffic Networks. *Proceedings of the 8th World Conference on Transport Research*. Antwerp, Belgium.

Stankova, K., M.C.J. Bliemer and G.J. Olsder (2006). Inverse Stackelberg Games and Their Application to Bilevel Optimal Toll Design Problem. *Proceedings of the 12th International Symposium on Dynamic Games and Applications*. Sophia Antipolis, France.

Taale, H. (1995). An Algorithm for Calculating the Cycletime and Greentimes for a Signalized Intersection. *Studies of the Faculty of Management Science, University of Transport and Communications, Žilina, Slovakia*, Vol. 4, pp. 87–94.

Taale, H. (2000). Optimising Traffic Signal Control with Evolutionary Algorithms. *Proceedings of the 7th World Congress on Intelligent Transport Systems*. Turin, Italy.

Taale, H. (2002). Comparing Methods to Optimise Vehicle Actuated Signal Control. *Proceedings of the 11th International Conference on Road Transport Information and Control*, pp. 114–119. Institute of Electrical Engineers, London, United Kingdom.

Taale, H. (2006). Regional Traffic Management - Method and Tool. *Preprints of the 11th IFAC Symposium on Control in Transportation Systems*, pp. 457–462. Delft, The Netherlands.

Taale, H., Th. Bäck, M. Preuß, A.E. Eiben, J.M. de Graaf and C.A. Schippers (1998). Optimizing Traffic Light Controllers by Means of Evolutionary Algorithms. *Proceedings of the EUFIT'98 Conference*, pp. 1730–1734. Aachen, Germany.

Taale, H. and F. Middelham (1995). Simulating Traffic and Traffic Control with FLEXSYT-II-. In: M. Snorek *et al.* (eds.), *Modelling and Simulation*, pp. 335–339. Society for Computer Simulation International, Istanbul.

Taale, H. and F. Middelham (1997). FLEXSYT-II-, A Validated Microscopic Simulation Tool. In: M. Papageorgiou and A. Pouliezos (eds.), *Proceedings of the 8th IFAC/IFIP/IFORS Symposium*, pp. 923–928. Chania, Crete, Greece.

Taale, H., H. Schuurman and G. Bootsma (2002). *Evaluation Large Road Works on the A10-West*. Report, AVV Transport Research Centre. (in Dutch).

Taale, H., H. Schuurman and G. Bootsma (2006). Evaluation of Traffic and Mobility Management During Road Works on the A10-West Motorway. *Proceedings of the 13th World Congress on Intelligent Transport Systems*. London, United Kingdom.

Taale, H. and H.J. van Zuylen (2000). Traffic Control Types and Route Choice: Some Simple Examples. *Proceedings of the 7th World Congress on Intelligent Transport Systems*. Turin, Italy.

Taale, H. and H.J. van Zuylen (2001). Testing the HCM 1997 Delay Function for Dutch Signal Controlled Intersections. *Proceedings of the 80th Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Taale, H. and H.J. van Zuylen (2002). Anticipatory Control: Effects of Uncertainty in Travel Times. *Proceedings of the 13th Mini-EURO Conference "Handling Uncertainty in the Analysis of Traffic and Transportation Systems"*, pp. 126–130. Polytechnical University of Bari, Bari, Italy.

Taale, H. and H.J. van Zuylen (2003a). The Effects of Anticipatory Control for Several Small Networks. *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington D.C., U.S.A.

Taale, H. and H.J. van Zuylen (2003b). Effects of Anticipatory Control with Multiple User Classes. *European Journal of Transport and Infrastructure Research*, Vol. 3, No. 1, pp. 61–76.

Taale, H. and M. Westerman (2005). The Application of Sustainable Traffic Management in The Netherlands. *Proceedings of the European Transport Conference 2005*. Association for European Transport, Strasbourg, France.

Taale, H., M. Westerman, H. Stoelhorst and D. Van Amelsfort (2004). Regional and Sustainable Traffic Management in The Netherlands: Methodology and Applications. *Proceedings of the European Transport Conference 2004*. Association for European Transport, Strasbourg, France.

Tanner, J.C. (1962). A Theoretical Analysis of Delays at an Uncontrolled Intersection. *Biometrica*, Vol. 49, pp. 163–170.

Taylor, N.B. (2003). The CONTRAM Dynamic Traffic Assignment Model. *Networks and Spatial Economics*, Vol. 3, No. 3, pp. 297–322.

Theil, H. (1961). *Economic Forecasts and Policy*. North-Holland Publishing Company, Amsterdam, The Netherlands.

Transportation Research Board (1997). *Highway Capacity Manual*. Special Report 209, 3rd Edition, 3rd Revision, National Research Council, Washington D.C., U.S.A.

Transportation Research Board (2000). *Highway Capacity Manual*. Report, National Research Council, Washington D.C., U.S.A.

Troutbeck, R.J. and W. Brilon (2002). Unsignalized Intersection Theory. *Traffic Flow Theory, A State of the Art Report*, chapter 8. Federal Highway Administration. Website: www.tfhrc.gov/its/tft/tft.htm, accessed on January 17th, 2005.

Tu, H. (2008). *Monitoring Travel Time Reliability on Freeways*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Van Berkum, E.C. and P.H.J. Van der Mede (1993). *The Impact of Traffic Information: Dynamics in Route and Departure Time Choice*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Van Katwijk, R.T. (2008). *Multi-Agent Look-Ahead Traffic-Adaptive Control*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Van Lint, J.W.C. (2004). *Reliable Travel Time Prediction for Freeways*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Van Vliet, D. (1982). SATURN - A Modern Assignment Model. *Traffic Engineering and Control*, Vol. 23, No. 12, pp. 578–581.

Van Vuren, T. and D. Van Vliet (1992). *Route Choice and Signal Control: The Potential for Integrated Route Guidance*. Institute for Transport Studies 5, Avebury, United Kingdom.

Van Zuylen, H.J. (1980). *A Delay Minimizing Traffic Assignment for (Nearly) Congested Road Networks*. Intenal Report, Verkeersakademie Tilburg.

Van Zuylen, H.J. (1981). Some Improvements in the Estimation of an OD Matrix from Traffic Counts. *Proceedings of the 8th International Symposium on Transportation and Traffic Theory*. Toronto, Canada.

Van Zuylen, H.J. (2001). The Assessment of Economic Benefits of Dynamic Traffic Management. *Proceedings of the 9th World Conference on Transport Research*. Seoel, Korea.

Van Zuylen, H.J. and H. Taale (2000). Anticipatory Optimization of Traffic Control. *Transportation Research Record*, No. 1725, pp. 109–115.

Van Zuylen, H.J. and H. Taale (2004). Urban Networks with Ring Roads: A Two-Level, Three Player Game. *Transportation Research Record*, No. 1894, pp. 180–187.

Van Zuylen, H.J. and L.G. Willumsen (1981). The Most Likely Trip Matrix Estimated from Traffic Counts. *Transportation Research Part B*, Vol. 14, No. 3, pp. 281–293.

Varia, H.R. and S.L. Dhingra (2004). Dynamic User Equilibrium Traffic Assignment on Congested Multidestination Networks. *Journal of Transportation Engineering*, Vol. 130, No. 2, pp. 211–221.

Versteegt, H.H., K.M. Malone, A.I. Martins Botto de Barros, R.J. Castenmiller and M.C.J. Bliemer (2003). *Follow-Up Study on Testing INDY for Use as the Dynamic Traffic Assignment Model in the Dutch National Model*. Report 2003-53, TNO Inro.

Viti, F. (2006). *The Dynamics and the Uncertainty of Delays at Signals*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

Wardrop, J.G. (1952). Some Theoretical Aspects of Road Traffic Research. *Proceedings of the Institute of Civil Engineers, Part II*, 1, pp. 325–378.

Warren, Mark (1990). Traffic Control in NYC - A Brief History. Website: www.nyc.gov/html/nypd/html/transportation/newpage5.html, accessed on April 28th, 2006.

Watling, D.P. (1997). Modelling Responsive Signal Control and Route Choice. In: D.A. Henster *et al.* (eds.), *Selected Proceedings of the 7th World Conference on Transport Research*, pp. 123–138. Pergamon.

Webster, F.V. (1958). *Traffic Signal Settings*. Road Research Technical Paper No. 39, Road Research Laboratory, United Kingdom.

Wikipedia (2002). Traffic Light. Website: http://en.wikipedia.org/wiki/Traffic_light, accessed on April 28th, 2006.

Wilson, Ad (1999). *Verkeersregelinstallaties in Nederland - Enquête-Onderzoek Onder Wegbeheerders*. Report, Nationale Hogeschool voor Toerisme en Verkeer and DTV Consultants, also by order of ITS Netherlands. (in Dutch).

Wong, S.C., J.S.W. Chung and C.O. Tong (1999). Equilibrium Networks: Practical Schemes for Signal Optimisation. *Traffic Engineering + Control*, Vol. 40, No. 12, pp. 586–589.

Wong, S.C. and C. Yang (1999). An Iterative Group-Based Signal Optimization Scheme for Traffic Equilibrium Networks. *Journal of Advanced Transportation*, Vol. 33, No. 3, pp. 210–217.

Wong, S.C., C. Yang and H.K. Lo (2001). A Path-Based Traffic Assignment Algorithm Based on the TRANSYT Traffic Model. *Transportation Research Part B*, Vol. 35, No. 2, pp. 163–181.

Wu, N. (2001). A Universal Procedure for Capacity Determination at Unsignalized (Priority-Controlled) Intersections. *Transportation Research Part B*, Vol. 35, No. 6, pp. 593–623.

Yang, H. and S. Yagar (1994). Traffic Assignment and Traffic Control in General Freeway-Arterial Corridor Systems. *Transportation Research Part B*, Vol. 28, No. 6, pp. 463–486.

Yang, H. and S. Yagar (1995). Traffic Assignment and Signal Control in Saturated Road Networks. *Transportation Research Part A*, Vol. 29, No. 2, pp. 125–139.

A

Traffic Management Architecture

A.1 Traffic management architecture

"The purpose of traffic management is to inform, induce and, if necessary, direct road users towards a safer and more efficient use of the existing infrastructure while safeguarding the quality of the environment of those living and working in the vicinity of the road network." The introduction of the Handbook Sustainable Traffic Management (STM) of Rijkwaterstaat (2003) opens with this definition of traffic management. The Handbook is part of the Dutch National Traffic Management Architecture.

The Traffic Management Architecture (TMA) is a structured description of the complex system of traffic and traffic management measures. It can be used to develop and implement a consistent and accepted set of traffic management measures and the necessary technical and information infrastructure.

The Dutch Traffic Management Architecture is consistent with the European ITS Framework Architecture, as was shown by Giezen and Avontuur (2002). The European ITS Framework Architecture is described by Bossom *et al.* (2000) and must accommodate national plans and support the various efforts in research, standardisation, deployment and investment in ITS. It must also provide a migration plan which incorporates and builds upon existing 'legacy' systems. A common framework provides specifications that enables:

- Compatibility of information delivered to end-users through different media;

- Compatibility of equipment with infrastructures, thus enabling seamless travel across Europe;

Figure A.1: Handbook Sustainable Traffic Management

- A basis for regional, national and European authorities to produce master plans and recommendations to facilitate ITS deployment;

- An open market for services and equipment where compatible sub-systems are offered (no more ad hoc solutions);

- Economies of scale in equipment manufacture permitting competitive prices and cheaper investments when compatibility is guaranteed;

- A known market place into which producers can supply products with reduced financial risk.

The European ITS Framework Architecture has much in common with the US National ITS Architecture and other architecture initiatives around the world. Aside from any differences in the scope of the user requirements, the main difference between the European and American architecture is that the US National ITS Architecture describes its functionality in a much greater level of detail (Bossom *et al.*, 2000). More information on the European ITS Framework Architecture can be found at the website of the FRAME Project (2007).

The TMA consists of five sub-architectures, each describing one aspect of traffic management. For defining and using a consistent set of traffic management measures, the Traffic Control Architecture is used. For the integration of the hardware and software an Application Architecture is defined. The Architecture of the Technical Infrastructure describes the general ICT services in traffic management systems. The Information Architecture should harmonise the exchange and use of information and finally the Organisation Architecture gives a picture of the organisation required to facilitate traffic management. Of these five sub-architectures, the Traffic Control Architecture is the most developed one and plays a leading role in the design, implementation and operational use of traffic management (Rijkswaterstaat, 2001).

A.2 Traffic control architecture

The Traffic Control Architecture (TCA) describes the process (typical Dutch!) to go from policy objectives to operational traffic control. In this process, cooperation between all parties involved is a key issue. Parties involved can be national, regional and local road authorities or road operators, public transport companies, chambers of commerce, major companies, road user associations, police departments, etc. With these different parties policy objectives are discussed and defined. These objectives are made concrete by linking them to a specific route or location and translating them into specific traffic indicators, such as average speed on a certain road or travel time for an origin to a certain destination. Comparing the actual situation with the target situation, results into the locations with problems and the bottlenecks. The next step will be to agree on traffic management measures to solve or mitigate the problems.

Also the operational side of traffic control, such as defining, testing and implementing control scenarios for roadworks, incidents, football matches, etc., is part of the TCA. As described by Schuurman (2003) in the operational part the use of models to assess different control scenarios becomes very important. To structure the process to come to a widely accepted traffic control architecture, the Handbook Sustainable Traffic Management was developed.

A.3 Handbook

The handbook describes the nine steps, which are defined in the process. The steps are hierarchical in the sense that they lead the user from a high-level strategic goal to low-level operational traffic management measures. The first step is to initiate and organise the project and bringing together the relevant parties. The next step is to formulate the common policy objectives in a clear and unambiguous way, such that the target situation is clear. After that, a control strategy has to be developed, which defines the way to handle the traffic in a region in case of problems and sets priorities to areas and routes (step 3). So, a control strategy defines important areas and roads and less important areas and roads. Attached to the control strategy, a frame of reference is needed, which quantifies the control strategy (step 4). A frame of reference makes the control strategy measurable and makes the step from a qualitative judgement (this road is important and has high priority) to a quantitative measure (the speed on this road must be higher than 60 km/h at all times). So, the frame of reference includes measurable criteria and thresholds to indicate the acceptable situation. An example of a frame of reference for a situation with roadworks is given by Kock and Van Den Hoogen (2002).

In the fifth step the target situation is compared with the actual or future situation. From this comparison traffic problems and bottlenecks are derived in step 6. To solve or decrease the traffic problems for these bottlenecks services (desired effect on a location in the network) can be defined (step 7). Step 8 consists of linking the services to one of more traffic management measures. The final step completes the project by integrating the products from the previous steps into one document and formulating an implementation plan and policy document.

Of course, the steps described in the handbook are very important to implement successful traffic management, but the next steps are even more important, namely to realise the goals and implement the measures defined and to come to operational traffic control at a network level.

A.4   Regional Traffic Management Explorer

The Regional Traffic Management Explorer (RTME) follows the exact same steps as the STM method, but from a quantitative perspective and in a dynamic modelling environment. It provides tools to formulate policy objectives and a control strategy. The objectives are quantified in a so called frame of reference and a user can input reference values for different (flexible) criteria, such as average speed on links, or (parts of) routes, travel times between origins and destinations, etc. When the policy objectives are confronted with the actual situation, bottlenecks arise for the different criteria. The objective in the remainder of the STM process will then be to eliminate as much of these bottlenecks as possible. The Handbook STM and RTME facilitate this in two steps. In the first step, called services, resolving the bottlenecks is thought of in general terms, such as reduce inflow, restrict speed and increase capacity. In the second step these services are translated into actual measures, such as ramp metering, tidal flow lanes, speed limits, etc. The effects of services and measures can be calculated using the dynamic equilibrium assignment model MARPLE, which is part of the RTME. More information about the method and the tool can be found in Taale *et al.* (2004), Taale and Westerman (2005) and Taale (2006).
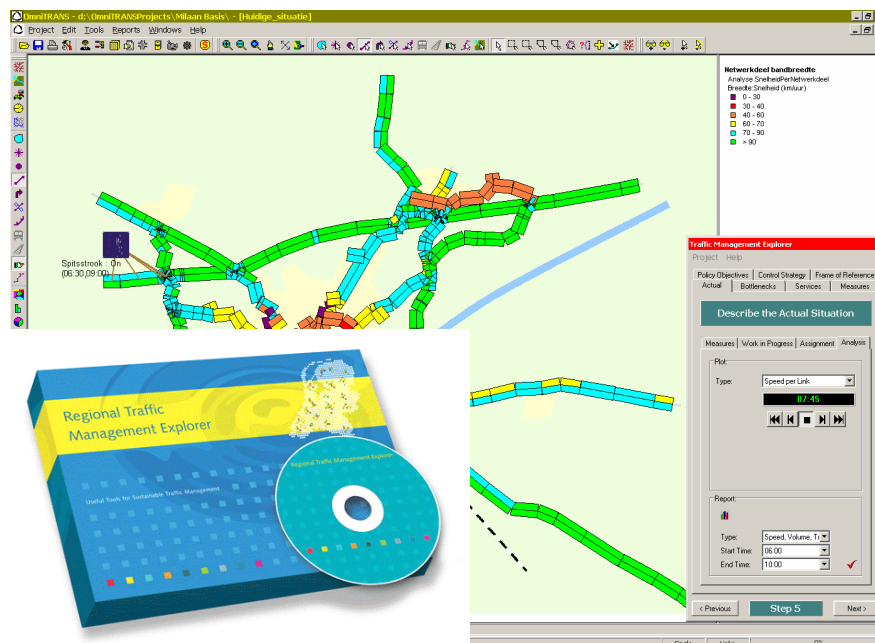
Figure A.2: Regional Traffic Management Explorer

B

MSA Step Size

For a stochastic assignment to converge, flows have to be smoothed. If we assume that for iteration j we have a route flow vector $\mathbf{f}^{(j)}$, then we smooth the flows with

$$\mathbf{f}^{(j)} = \mathbf{f}^{(j-1)} + \zeta^{(j)}(\mathbf{f}^{(j)} - \mathbf{f}^{(j-1)}). \tag{B.1}$$

Normally, the method of successive averages (MSA) is used for this. This method uses the iteration number for the step size: $\zeta^{(j)} = 1/j$. The convergence of the MSA is slow, because the step size ζ quickly becomes small and slowly decreases. To overcome this problem, a somewhat adjusted MSA is proposed in section 4.2.4. The step size for smoothing the flows is defined as

$$\zeta^{(j)} = a_1 e^{-a_2 j} + \frac{a_3}{j}, \tag{B.2}$$

where j is the iteration number and a_1, a_2 and a_3 are parameters. Sheffi (1985) states that for convergence the step size has to satisfy certain conditions. These condition are

$$\sum_{j=1}^{\infty} \zeta^{(j)} = \infty \text{ and } \sum_{j=1}^{\infty} \left(\zeta^{(j)} \right)^2 < \infty. \tag{B.3}$$

It is easy to proof that $\zeta^{(j)} = 1/j$ meets these conditions, because from the mathematical literature it is known that

$$\sum_{j=1}^{\infty} \frac{1}{j} = \infty, \tag{B.4}$$

and

$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6} < \infty. \tag{B.5}$$

It remains to proof that the step size defined in equation B.2 meets the conditions. So, we want to proof that

$$\sum_{j=1}^{\infty} \left(a_1 e^{-a_2 j} + \frac{a_3}{j} \right) = \infty, \tag{B.6}$$

and

$$\sum_{j=1}^{\infty} \left(a_1 e^{-a_2 j} + \frac{a_3}{j} \right)^2 < \infty. \tag{B.7}$$

Equation B.6 follows directly from equation B.4, but only if $a_3 > 0$. Proving equation B.7 is slightly harder. We have

$$\left(a_1 e^{-a_2 j} + \frac{a_3}{j} \right)^2 = a_1^2 e^{-2a_2 j} + \frac{a_1 a_3}{j} e^{-a_2 j} + a_3^2 \frac{1}{j^2}. \tag{B.8}$$

We will prove equation B.7 by proving that all terms in equation B.8 are less than ∞. For the *third* term that is easy to see, because equation B.5 holds. For the *first* term we have

$$\sum_{j=1}^{\infty} a_1^2 e^{-2a_2 j} = \sum_{j=1}^{\infty} a_1^2 \left(e^{-2a_2} \right)^j. \tag{B.9}$$

Using the standard geometric series

$$\sum_{j=1}^{\infty} a x^j = \frac{ax}{1-x}, \tag{B.10}$$

which is valid for $|x| < 1$, we obtain

$$\sum_{j=1}^{\infty} a_1^2 \left(e^{-2a_2} \right)^j = \frac{a_1^2 e^{-2a_2}}{1 - e^{-2a_2}} < \infty. \tag{B.11}$$

To meet the condition for the geometric series, a_2 should be larger than zero. For the *second* term we know that

$$\frac{a_1 a_3}{j} e^{-a_2 j} < a_1 a_3 e^{-a_2 j} \Rightarrow \sum_{j=1}^{\infty} \left(\frac{a_1 a_3}{j} e^{-a_2 j} \right) < \sum_{j=1}^{\infty} \left(a_1 a_3 e^{-a_2 j} \right), \tag{B.12}$$

and, using the geometric series again, it is true that

$$\sum_{j=1}^{\infty} a_1 a_3 e^{-a_2 j} = \frac{a_1 a_3 e^{-a_2}}{1 - e^{-a_2}} < \infty. \tag{B.13}$$

So also the second term

$$\sum_{j=1}^{\infty} \frac{a_1 a_3}{j} e^{-a_2 j} < \infty, \tag{B.14}$$

which completes the proof.

C

Method for Calibrating OD Matrix

In this study the calibration of OD matrices was done with the method described by Van Zuylen and Willumsen (1981) and adjusted by Van Zuylen (1981). This method uses a base OD matrix and adjusts OD flows to match the measured flows as close as possible. Originally, the method was meant for static matrices and one time period. In that case, assuming that the matrix contains minimum information, the equation to solve is

$$q^{od} = q_0^{od} \prod_a X_a^{p_a^{od}},$$ (C.1)

where q^{od} is the estimation for OD pair od, q_0^{od} the original, a priori demand for OD pair od, X_a are factors to adjust the matrix and p_a^{od} is the fraction of trips for relation od using link a (the outcome of the assignment). If it is assumed that the total number of trips is not fixed, we get the following equation

$$q^{od} = q_0^{od} X_o \prod_a X_a^{p_a^{od}},$$ (C.2)

where X_o is a scaling factor, which is defined by

$$X_o = \frac{\sum\limits_{o,d} q^{od}}{\sum\limits_{o,d} q_0^{od}}.$$ (C.3)

The factors X_o and X_a are to be solved with

$$V_a = \sum_{o,d} p_a^{od} q_0^{od} X_o \prod_a X_a^{p_a^{od}},$$ (C.4)

and

$$\sum_{o,d} q_0^{od} = \sum_{o,d} q_0^{od} \prod_a X_a^{p_a^{od}}, \tag{C.5}$$

where V_a is the flow measured on link a. The algorithm to solve this starts with initialising the iteration number $n = 0$ and setting the factors $X_a^0 = 1$ for all links a and

$$X_o^0 = \frac{\sum\limits_a V_a}{\sum\limits_a \sum\limits_{o,d} p_a^{od} q_0^{od}}. \tag{C.6}$$

For each iteration n and for each link a calculate X_a^{n+1} by solving

$$V_a = \sum_{o,d} p_a^{od} X_o^n q_0^{od} \left(\prod_{a'} (X_{a'}^n)^{p_{a'}^{od}} \right) Y_a^{p_a^{od}}, \tag{C.7}$$

for Y_a and setting $X_a^{n+1} = X_a^n Y_a$. Calculate the new OD adjustment factors with

$$\widehat{X}_o^{n+1} = \frac{X_o^n \sum\limits_{o,d} q_0^{od} \prod\limits_a \left(X_a^{n+1} \right)^{p_a^{od}}}{\sum\limits_{o,d} q_0^{od}}, \tag{C.8}$$

and smooth them with $X_o^{n+1} = 1/2(X_o^n + \widehat{X}_o^{n+1})$. Now the estimated OD matrix q_{n+1}^{od} is given by

$$q_{n+1}^{od} = q_0^{od} X_o^{n+1} \prod_a \left(X_a^{n+1} \right)^{p_a^{od}}. \tag{C.9}$$

If the flows, resulting from the assignment with q_{n+1}^{od}, are sufficient close to the measurements, then stop, otherwise, set $n := n + 1$ and return to equation C.7.

D

Estimating Travel Times from Loop Data

Van Lint (2004) describes a trajectory method to estimate travel times from loop data. The loop data contain flows and time-mean speeds per minute. To correct for the bias using time-measured speeds, the speed variance is used. This speed variance is estimated using a time series approach. The method was implemented and compared with travel time measurements. These measurements stem from a camera system with numberplate recognition software, implemented on the A13 motorway between The Hague and Rotterdam. To show the validity of the estimation method, 6 days of data were compared. In these 6 days different situations occur, such as a small amount of congestion, a normal amount of congestion, a large amount of congestion and an incident situation (March 30th, 2006). For every day the speed contour plot and the travel times are shown in the figures below. From the figures it is clear that the method gives a good estimate of the travel times, even in situations with incidents.
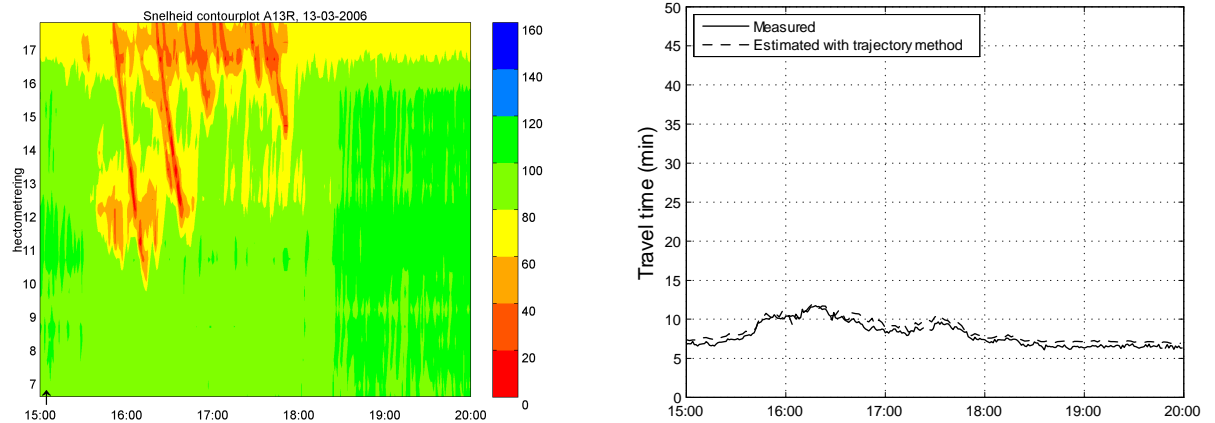
Figure D.1: Speed contour plot and travel times for March 13th, 2006
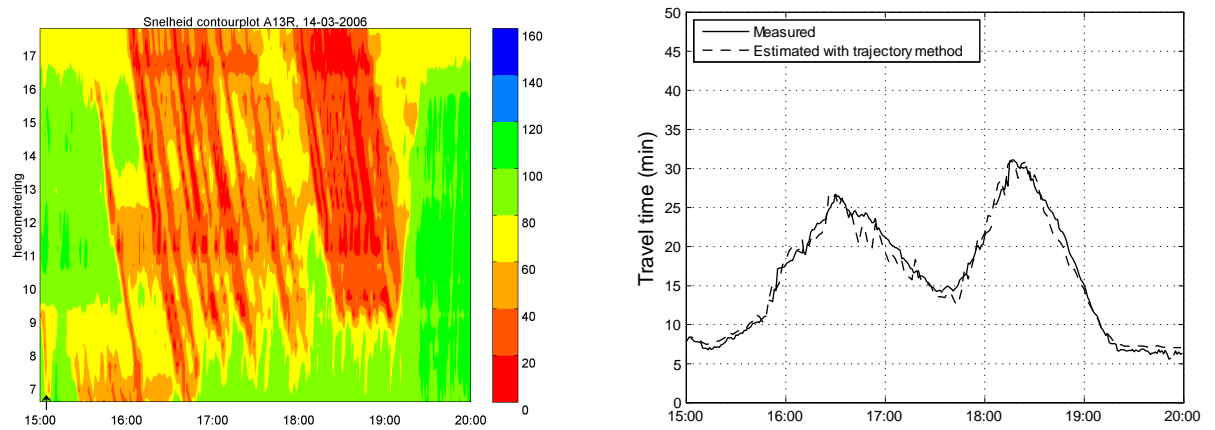
Figure D.2: Speed contour plot and travel times for March 14th, 2006
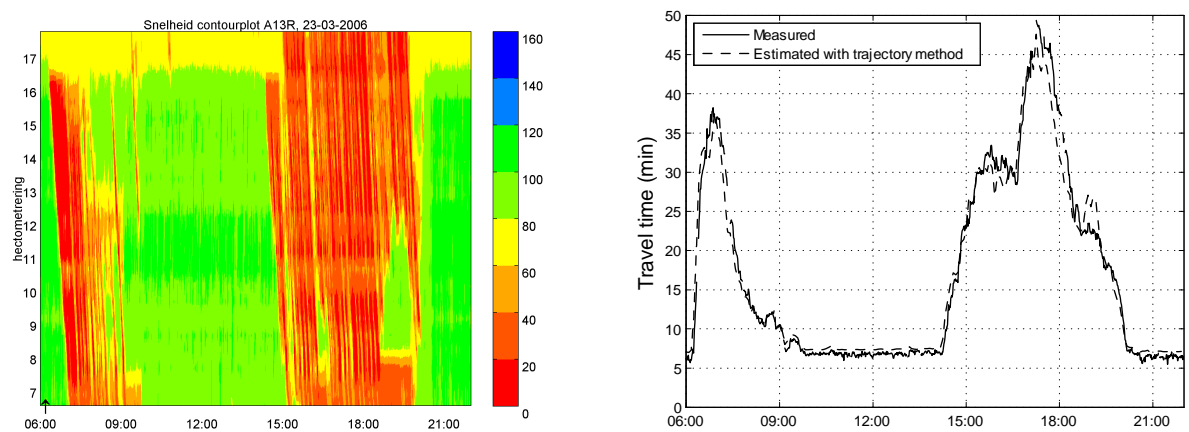
Figure D.3: Speed contour plot and travel times for March 23rd, 2006

Figure D.4: Speed contour plot and travel times for March 29th, 2006

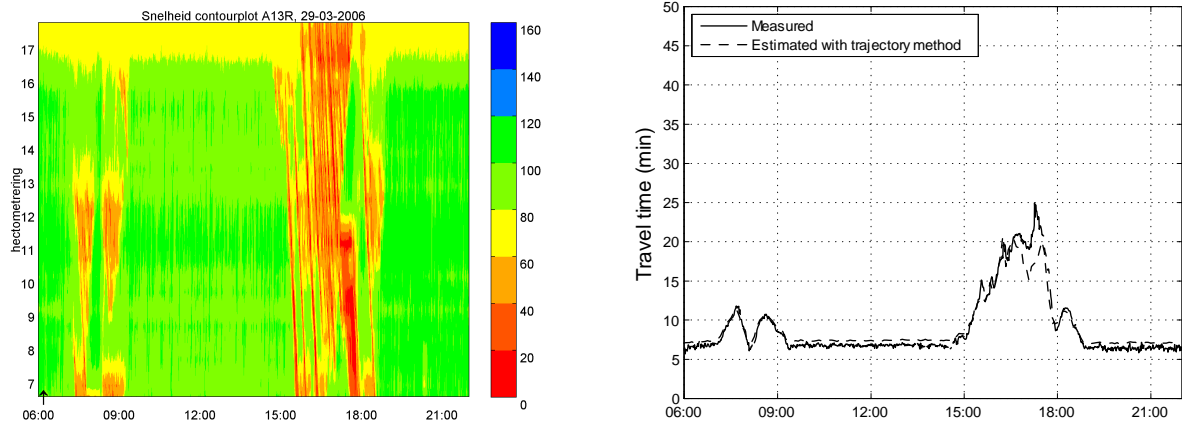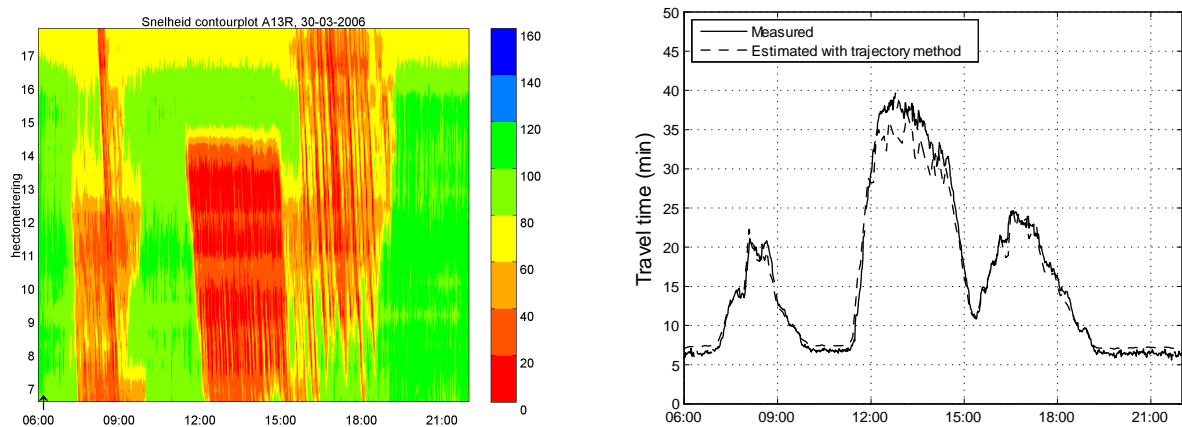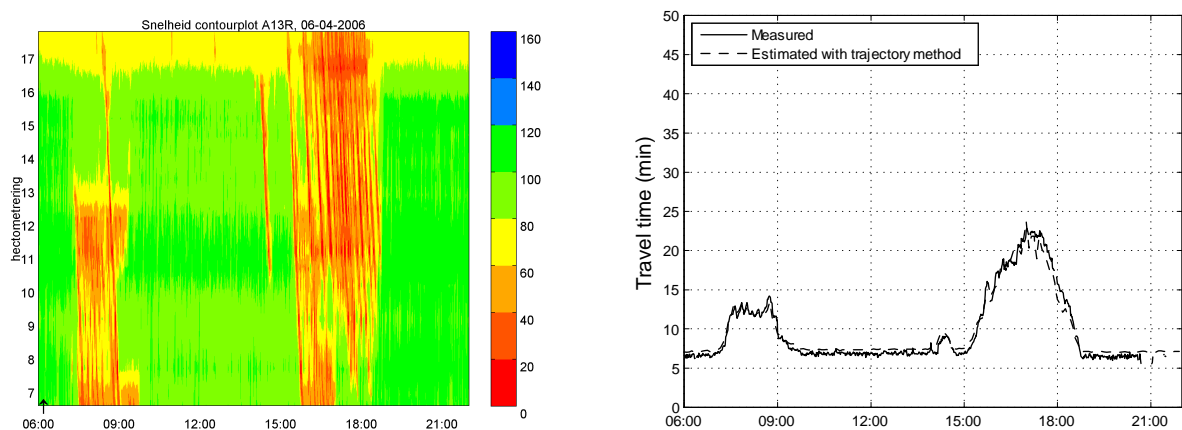**Figure D.5:** Speed contour plot and travel times for March 30th, 2006

Figure D.6: Speed contour plot and travel times for April 6th, 2006

Summary

Integrated and Anticipatory Control of Road Networks
A game-theoretical approach

Henk Taale

In The Netherlands transport and traffic policy heavily relies on traffic management. Building new roads is either too expensive or takes too much time due to procedures related to spatial and environmental conditions. Technically and politically road pricing will be difficult to implement the coming years, so for the Dutch Ministry of Transport, Public Works and Water Management traffic management is the key direction in which solutions for the increasing congestion problems have to be found (MinVenW, 2005). The reason for this is that traffic management is faster to implement and it faces less resistance than the other solution directions. In fact this is the situation since the nineties from the previous century. From 1989 on, a lot of traffic management measures were implemented, varying from a motorway traffic management system and ramp metering systems, to overtaking prohibitions for trucks and special rush hour teams of the traffic police. In a recent policy document the Dutch Ministry of Transport, Public Works and Water Management (MinVenW, 2008) estimates that traffic management reduced the increase of congestion (measured in vehicle hours delay) with 25% during the years 1996-2005.

In most cases traffic management in The Netherlands is used only on a local level. It lacks an integrated and network wide approach. The main reason for this is that different network types (e.g. motorways and urban roads) are operated and maintained by different road managers. In practise these road managers are only responsible for their own part of the network and normally they do not communicate or cooperate that much. To deal with this, The Netherlands

have adopted a different approach, described in the Handbook Sustainable Traffic Management (Rijkwaterstaat, 2003). The handbook gives a step-by-step method that enables policy makers and traffic engineers to translate policy objectives into specific measures. The method consists of clearly defined steps that can be summarised as: define policy objectives, assess current situation, determine bottlenecks and create solutions. The step-by-step plan helps to develop a network vision based on policy objectives, shared by all participating stakeholders. In addition, the handbook provides the stakeholders with a first indication of the measures required to achieve effective traffic management in line with the shared vision. To be able to say more about the effects, the Regional Traffic Management Explorer (RTME) was developed. This sketch and calculation tool supports the steps from the handbook and makes it possible to determine the effects of proposed traffic management services and measures. The effects can then be compared to the formulated policy objectives or other sets of measures. For more information on the method, the RTME and its applications, the reader is referred to Taale *et al.* (2004) and Taale and Westerman (2005).

To be able to calculate the effectiveness of traffic management, the Regional Traffic Management Explorer (RTME) uses a dynamic traffic assignment model. This Model for Assignment and Regional PoLicy Evaluation (MARPLE) is a result of the research described in this thesis. It consists of a control module, an assignment module and a network loading module. For the RTME the control module contains only local control strategies, because that is sufficient to represent the Dutch circumstances. For the research in this thesis the control module is much more advanced and contains a network wide, integrated and anticipatory control strategy. Later on more about this (which is the topic of this thesis), but first we discuss the assignment and network loading modules, which are the same ones for the RTME as we developed for our research and described in this thesis.

The dynamic traffic assignment (DTA) module contains three different assignment methods: deterministic, stochastic and system optimal. For the deterministic assignment problem the variational inequality (VI) approach is used to solve it. An important aspect of the VI-approach is the contraction factor which is used in the solution algorithm. In our research we use a dynamic contraction factor and for an example network it is shown that this leads to faster convergence than using a traditional fixed factor. To solve the stochastic assignment problem, the standard C-logit method is used, in combination with the method of successive averages (MSA) to smooth the flows. It is known that standard MSA has slow convergence. Therefore, adjusted MSA factors are used and it is shown, for the same example network, that using these adjusted MSA factors leads to faster convergence. For the convergence criterion of the iterative assignment process, two options are investigated: one based on flow and one based on the duality gap. Both criteria give the same equilibrium, but the flow criterion shows a faster convergence (again for the small example network). The system optimum assignment is the best possible distribution of traffic on available routes in terms of network indicators. In this thesis it is always used in combination with system optimum control. An evolutionary algorithm is used to solve both problems in one step. All assignment methods are route based. That means that route searching is important. The route enumeration process searches for the k-shortest routes using a Monte Carlo approach, with a stochastic variation of the free flow link travel times and

Dijkstra's shortest path algorithm.

The dynamic network loading (DNL) model uses travel time functions to propagate traffic through the network. For different link types (normal links, signal controlled links, round-about links and priority links) different functions are used. The travel time is used to determine the outflow of links and with that the inflow of downstream links. Because of the assumption of a uniform distribution of traffic in the length of a link, differences in traffic flow at the beginning of a link are damped before they reach the end of the link, so it takes too much time before congestion starts or dissolves. To overcome this problem, a new step mechanism is introduced, which makes sure that differences in inflow reach the end of the link in time. Horizontal queuing and blocking back is modelled by the concept of 'available space'. The available space on a link determines how much traffic can enter the link and thus how much traffic is held back on the upstream links. This traffic can block other traffic and is distributed among the upstream links according to the number of lanes. At decision nodes traffic is distributed from the incoming to the outgoing links according to the splitting rates, which are calculated from the route flows using free flow travel times. Congestion is always caused by a capacity restriction and the resulting queue propagates upstream and horizontal, which means that blocking back is taken into account. The route travel times (needed for the assignment) are calculated from the link travel times using a trajectory method. The DNL model treats critical links (links for which traffic can travel the whole length within one time step) in a special way. Not the free flow travel time, but the actual travel time is used to determine the critical links. The determination of critical links and the adjustments for these links are done for every time step. Tests show that this way of handling critical links gives good results.

The DTA and DNL models are calibrated and validated for a motorway bottleneck and for a network with motorways and urban roads. For both situations real-life data is used to calibrate parameters and to see of model results and data are comparable. Although comments can be made concerning the data and the method of comparison, it appears that the DNL model is capable of simulating bottlenecks fairly accurate, and that the combination of the DTA and DNL models is capable of simulating medium-sized networks with good results.

Both the assignment and network loading modules are part of a framework for integrated anticipatory control. Integrated control means that the network is considered to be one multi-level network, consisting of motorways and urban roads. Anticipatory control means taking into account not only the current, but also future traffic conditions. For these future traffic conditions the focus is on long term behaviour of road users, such as route choice and choice of departure time. In this thesis the question is answered how traffic management should be designed and optimised and whether it is beneficial to anticipate route choice behaviour. To answer these questions the framework is extended with a control module and in this control module the traffic management measures are optimised in such a way that route choice behaviour is taken into account. The optimisation itself is done with an evolutionary algorithm. In this way the framework represents the Stackelberg game from game theory, in which one of the players knows the reaction of the other player(s) to his moves and therefore can anticipate this reaction.

The benefits of integrated anticipatory control are demonstrated with a number of case studies, representing different aspects. In all cases, with traffic signal control only, anticipatory control

gives better results in terms of the total delay in the network than local control strategies. This also is true for cases with integrated traffic management. Also for these cases anticipatory control gives the best results in terms of total delay. If traffic management is not integrated, but distributed among different road operators, we get cases with more than two players. Two cases with three players were simulated and also for these cases integrated anticipatory control was the best, especially if the number of control options increased.

We already mentioned that in many cases traffic management is reactive and local: it reacts on local traffic conditions and traffic management measures are taken to reduce congestion on that specific location. To come to an integrated and network-wide approach, the Handbook Sustainable Traffic Management describes a *process* for cooperation between the different road authorities and other stakeholders. This is a first and important step, but still a *methodological* approach to integrated traffic management is lacking. How can traffic management measures be operated to reduce congestion on a network level, taking network condition into account? In this thesis a framework for integrated and anticipatory traffic management is developed and demonstrated with good results. It can be used as a next step towards real network traffic management.

Samenvatting

Geïntegreerd anticiperend regelen van verkeersnetwerken
Een benadering met speltheorie

Henk Taale

Verkeersmanagement is één van de belangrijkste pijlers van het verkeers- en vervoerbeleid in Nederland. Het bouwen van nieuwe wegen is òf te duur òf het kost teveel tijd om het te realiseren door de ruimtelijke randvoorwaarden en leefbaarheidsaspecten. Daarnaast is het betalen voor mobiliteit de komende jaren technisch en politiek lastig in te voeren. Dus voor het Ministerie van Verkeer en Waterstaat is verkeersmanagement de belangrijkste richting waar binnen oplossingen voor het steeds maar groter wordende fileprobleem gezocht moeten worden (MinVenW, 2005). De redenen hiervoor zijn dat verkeersmanagement veel sneller is te implementeren en dat de maatschappelijke weerstand tegen verkeersmanagement kleiner is dan tegen de andere oplosrichtingen. In feite is dit de situatie sinds de negentiger jaren van de vorige eeuw. Vanaf 1989 is een groot aantal verkeersmanagement maatregelen gerealiseerd, variërend van het signaleringssysteem en toeritdosering tot het inhaalverbod voor vrachtverkeer en de spitsteams van het Korps Landelijke Politiediensten. In een recent beleidsdocument schat het Ministerie van Verkeer en Waterstaat (MinVenW, 2008) dat verkeersmanagement de toename van de files (gemeten in voertuiguren vertraging) heeft gereduceerd met 25%, gedurende de periode 1996-2005.

In de meeste gevallen wordt in Nederland verkeersmanagement toepast op een lokaal niveau. Het ontbreekt aan een geïntegreerde en netwerk brede aanpak. De belangrijkste reden hiervoor is dat de verschillende soorten wegen (bijvoorbeeld snelwegen en stedelijke wegen) door

verschillende wegbeheerders worden beheerd en onderhouden. In de praktijk zijn deze wegbe-
heerders alleen verantwoordelijk voor hun eigen deel van het wegennet en normaal gesproken
is er niet veel communicatie en onderlinge samenwerking. Om hier iets aan te doen, is een
nieuwe aanpak ontwikkeld, beschreven in het Werkboek Gebiedsgericht Benutten (Rijkwater-
staat, 2003). Dit werkboek bevat een stappenplan waarmee bestuurders en verkeerskundigen
samen beleidsdoelen vertalen in concrete maatregelen. Samengevat staan in het werkboek de
volgende duidelijke stappen beschreven: definieer beleidsdoelen, beschrijf de huidige situatie,
bepaal de knelpunten en verzin oplossingen voor die knelpunten. Het plan van aanpak met de
uitwerking van deze stappen is een hulpmiddel om een breed gedragen netwerkvisie te ontwik-
kelen. Bovendien geeft het werkboek een eerste indicatie van de maatregelen die nodig zijn
voor effectief verkeersmanagement, in lijn met de overeengekomen netwerkvisie. Om meer
over de effecten te kunnen zeggen, is de Regionale BenuttingsVerkenner (RBV) ontwikkeld.
Dit sketch-model ondersteunt de stappen uit het handboek en maakt het mogelijk om de ef-
fecten van de voorgestelde services en maatregelen te bepalen. Meer informatie over de GGB-
methode, de RBV en toepassingen daarvan, kan de lezer vinden in Taale *et al.* (2004) en Taale
and Westerman (2005).

Om de effectiviteit van verkeersmanagement te bepalen, gebruikt de Regionale Benuttings-
Verkenner een dynamisch toedelingsmodel. Dit model heeft de naam 'Model for Assignment
and Regional PoLicy Evaluation' (MARPLE) gekregen en is het resultaat van het onderzoek dat
in dit proefschrift beschreven wordt. Het model bestaan uit een module voor regelingen, een
module voor de toedeling en een simulatiemodule. Voor de RBV bestaat de regelmodule alleen
uit lokale regelstrategieën, omdat dat voldoende is voor de weergave van de Nederlandse situatie
met voldoende nauwkeurigheid. Voor het onderzoek in dit proefschrift is de regelmodule veel
geavanceerder en bevat het een netwerk brede, geïntegreerde en anticiperende regelstrategie.
Dit is het eigenlijke onderwerp van dit proefschrift en wordt later in deze samenvatting uit-
gebreider beschreven, maar eerst volgt een beschrijving van de toedelings- en simulatiemodule.
Deze zijn hetzelfde voor zowel de RBV als voor het promotieonderzoek.

De dynamische toedelingsmodule bevat drie methoden voor de toedeling: deterministisch,
stochastisch en systeem-optimum. Voor het oplossen van het deterministische toedelingspro-
bleem, wordt de 'variational inequality' (VI) aanpak gebruikt. Een belangrijk aspect van de VI-
aanpak is de contractiefactor die in het algoritme gebruikt wordt. Voor ons onderzoek gebruiken
we een dynamische contractiefactor en voor een klein netwerk kan aangetoond worden dat met
deze dynamische factor convergentie sneller wordt bereikt dan met een vaste factor. Voor
het oplossen van het stochastische toedelingsprobleem wordt de standaard C-logit methode
gebruikt, in combinatie met MSA (method of successive averages) om de intensiteiten af te
vlakken. Het is bekend dat met de standaard MSA factoren de toedeling niet zo snel convergeert.
Om het convergentieproces te versnellen, zijn daarom aangepaste MSA factoren gebruikt. Voor
hetzelfde voorbeeld netwerk kunnen we dit ook laten zien. Voor het convergentie criterium van
de iteratieve evenwichtstoedeling zijn twee mogelijkheden bekeken: een criterium gebaseerd
op het verschil in intensiteiten en een criterium dat kijkt hoe ver het product van intensiteit
en reiskosten afzit van de ideale situatie. Beide criteria leiden tot hetzelfde evenwicht, maar
het intensiteitscriterium convergeert sneller naar dit evenwicht (voor het gebruikte voorbeeld

netwerk). De systeem-optimum toedeling is die verdeling van het verkeer over de beschikbare routes die leidt tot de minste totale vertraging. In dit proefschrift wordt dit type toedeling altijd gebruikt in combinatie met een systeem-optimale regeling. Een evolutionair algoritme wordt gebruikt om beide problemen tegelijkertijd op te lossen. Alle toedelingstypen zijn gebaseerd op de verdeling van verkeer over routes. Dit betekent dat het kunnen beschikken over een goede set routes belangrijk is. Het proces om deze set te genereren, zoekt naar een opgegeven aantal kortste routes en maakt gebruik van een Monte Carlo aanpak, waarbij een stochastische term bij de vrije reistijden op de links wordt opgeteld en het Dijkstra algoritme wordt gebruikt om de dan geldende kortste route te bepalen.

Het simulatiemodel gebruikt reistijd functies om het verkeer door het netwerk heen te bewegen. De reistijd functies zijn verschillend voor verschillende typen links (gewone, geregelde, rotonde en voorrang links). De berekende reistijd wordt gebruikt om de uitstroom van een link te bepalen en met de uitstroom is dan ook weer de instroom bekend. Omdat wordt aangenomen dat verkeer uniform over de lengte van een link verdeeld is, worden verschillen in intensiteit bij de ingang van een link uitgedempt, voordat het einde van de link bereikt wordt. Dit leidt ertoe dat wachtrijen te langzaam opbouwen en weer verdwijnen. Om dit probleem te verhelpen, is een nieuw mechanisme geïntroduceerd dat er voor zorgt dat verschillen in de instroom op tijd het einde van een link bereiken. Horizontale wachtrijen en terugslag wordt gemodelleerd met de zogeheten 'beschikbare ruimte'. De beschikbare ruimte op een link bepaalt hoeveel verkeer een link kan binnen stromen en dus ook hoeveel verkeer er wordt tegengehouden op links stroomopwaarts. Dit verkeer blokkeert ook het andere verkeer en wordt over de inkomende links verdeeld naar rato van het aantal rijstroken. In knopen met routekeuze wordt het verkeer verdeeld met behulp van splitsfracties. Deze fracties worden berekend aan de hand van de intensiteiten voor de routes. Files worden altijd veroorzaakt door een beperkte capaciteit en de resulterende wachtrij bouwt zich stroomopwaarts (dus horizontaal) op. Dit betekent dat rekening wordt gehouden met de terugslag van files. De route reistijden (die nodig zijn voor de toedeling) worden berekend aan de hand van de link reistijden met een soort trajectorie methode. Het model gaat op een speciale manier om met de zogeheten kritische links. Dat zijn links die zo kort zijn dat het verkeer ze in één tijdstap passeert. Het bepalen van deze kritische links en de aanpassingen die dan worden gedaan, gebeurt elke tijdstap. Testen wijzen uit dat deze manier om met kritische links om te gaan, goede resultaten oplevert.

Het toedelingsmodel en het simulatiemodel zijn gekalibreerd voor een knelpunt op een snelweg en voor een netwerk van snelwegen en stedelijke wegen. Voor beide situaties is meetdata gebruikt om de modelparameters te kalibreren en om te kijken of de modeluitkomsten vergelijkbaar zijn met de gemeten data. Ondanks dat er zeker opmerkingen gemaakt kunnen worden over de data en de methode van vergelijken, is duidelijk dat het simulatiemodel in staat is om een knelpunt op een snelweg redelijk nauwkeurig te simuleren. De combinatie van het toedelingsmodel en het simulatiemodel is in staat om een middelgroot netwerk te simuleren met goede resultaten.

Zowel het toedelingsmodel als het simulatiemodel maken deel uit van een raamwerk voor geïntegreerd en anticiperend regelen. Geïntegreerd regelen betekent dat het netwerk beschouwd wordt als één netwerk, bestaande uit snelwegen, provinciale en stedelijke wegen. Anticiperend

regelen betekent dat niet alleen gereageerd wordt op de huidige toestand, maar dat ook met toekomstige verkeersomstandigheden rekening wordt gehouden. Voor toekomstige situaties ligt de nadruk op het lange termijn gedrag van weggebruikers, zoals routekeuze en keuze van vertrektijdstip. In dit proefschrift wordt de vraag beantwoord hoe verkeersmanagement ontworpen en geoptimaliseerd moet worden en wat de voordelen zijn om rekening te houden met routekeuze gedrag. Om deze vraag te kunnen beantwoorden, is het raamwerk uitgebreid met een regelmodule en in deze regelmodule worden verkeersmanagement maatregelen dusdanig geoptimaliseerd dat rekening wordt gehouden met routekeuze gedrag. De optimalisatie zelf wordt uitgevoerd met een evolutionair algoritme. Het raamwerk representeert het Stackelberg evenwicht uit de speltheorie, waarbij één of meerdere spelers kennis hebben van de reactie van andere spelers op hun zetten en daardoor kunnen anticiperen op deze reactie.

De voordelen van geïntegreerd en anticiperend regelen worden gedemonstreerd met een aantal case studies, die verschillende aspecten duidelijk maken. In alle netwerken met verkeerslichtenregelingen leidt anticiperend regelen tot betere resultaten (minder vertraging) dan lokale regelingen. En dat is ook waar voor netwerken met geïntegreerd verkeersmanagement. Als er meerdere wegbeheerders in het spel zijn, krijgen we gevallen met meer dan twee spelers. Twee voorbeelden met drie spelers zijn gesimuleerd en ook voor die voorbeelden kan geconcludeerd worden dat geïntegreerd en anticiperend regelen de beste resultaten geeft, zeker als de regelmogelijkheden groter worden.

We noemden al dat in veel gevallen verkeersmanagement reactief en lokaal is. Verkeersmanagement reageert op lokale omstandigheden en maatregelen worden genomen om files te verminderen op een specifieke locatie. Om tot een netwerk brede en geïntegreerde aanpak te komen, beschrijft het Werkboek Gebiedsgericht Benutten een *proces* om de samenwerking tussen de verschillende wegbeheerders en andere belanghebbenden tot stand te brengen. Echter, het ontbreekt nog steeds aan een *methodische* aanpak voor geïntegreerd verkeersmanagement. Hoe kunnen verkeersmanagement maatregelen worden ingezet om de files op netwerkniveau te verminderen? In dit proefschrift is een raamwerk voor geïntegreerd en anticiperend verkeersmanagement ontwikkeld en is aangetoond dat deze aanpak voordelig kan zijn. Het kan daarom gebruikt worden als een volgende stap naar echt verkeersmanagement op netwerkniveau.

About the Author

Henk Taale was born on May 29th, 1967 in Middelharnis, a small town situated on the island Goeree-Overflakkee, south of Rotterdam. He still lives there with his wife and two sons. He studied Applied Mathematics at the Delft University of Technology, where he received a MSc degree in 1991 on the topic of classification of dynamical systems of differential equations. After his study he started working as a project manager for Rijkswaterstaat, the Dutch motorway authority and he is still employed by this government agency.

He has almost 18 years of experience in the fields of traffic management, traffic models and evaluation and was project manager for numerous projects in those fields. To name a few: the assessment of tools for dynamic traffic management, such as ramp metering; the assessment of the UTC-system SCOOT in the city of Nijmegen; the further development, validation and maintenance of the microscopic simulation tool FLEXSYT-II-; several assessment studies, including road maintenance on the ring road of Amsterdam, cross-border management, a pilot on the use of floating car data and measures taken by the traffic police; the development of guidelines for assessment studies and model validation studies.

Currently, he is responsible for the further development and application of a tool for sustainable traffic management (called the Regional Traffic Management Explorer), for the design of a national monitoring and evaluation plan and for ITS Edulab, a cooperation between Rijkswaterstaat and the Delft University of Technology to enable master students to graduate on a research topic which is relevant for the operational practise of Rijkswaterstaat. He is also a member of the management committee of the International Benefits, Evaluation and Costs (IBEC) Working Group and a member of several conference committees.

As a part-time assignment he started his PhD in 1999 on the subject of anticipatory control of road networks. Anticipatory control deals with the interaction between road users and road

managers and with the question how to find an integrated control plan for networks taking into account the route choice behaviour of road users. The research was done under the supervision of prof.dr. Henk van Zuylen and was finished with a PhD thesis in 2008.

TRAIL Thesis Series

A series of The Netherlands TRAIL Research School for theses on transport, infrastructure and logistics. See for a complete overview our website: www.rsTRAIL.nl.

Taale, H., *Integrated Anticipatory Control of Road Networks. A game theoretical approach*, T2008/15, TRAIL Thesis Series, the Netherlands

Li, M., *Robustness Analysis for Road Networks. A framework with combined DTA models*, T2008/14, TRAIL Thesis Series, the Netherlands

Yu, M., *Enhancing Warehouse Performance by Efficient Order Picking*, T2008/13, TRAIL Thesis Series, the Netherlands

Liu, H., *Travel Time Prediction for Urban Networks*, T2008/12, TRAIL Thesis Series, the Netherlands

Kaa, E.J. van de, *Extended Prospect Theory. Findings on Choice Behaviour from Economics and the Behavioural Sciences and their Relevance for Travel Behaviour*, T2008/11, TRAIL Thesis Series, the Netherlands

Nijland, H., *Theory and Practice of the Assessment and Valuation of Noise from Roads and Railroads in Europe*, T2008/10, TRAIL Thesis Series, the Netherlands

Annema, J.A., *The Practice of Forward-Looking Transport Policy Assessment Studies*, T2008/9, TRAIL Thesis Series, the Netherlands

Ossen, S.J.L., *Theory and Empirics of Longitudinal Driving Behavior*, T2008/8, TRAIL Thesis Series, the Netherlands

Tu, H., *Monitoring Travel Time Reliability on Freeways*, T2008/7, April 2008, TRAIL Thesis Series, the Netherlands

DAriano, A., *Improving Real-Time Train Dispatching: Models, Algorithms and Applications*, T2008/6, April 2008, TRAIL Thesis Series, the Netherlands

Quak, H.J., *Sustainability of Urban Freight Transport. Retail Distribution and Local Regulations in Cities*, T2008/5, March 2008, TRAIL Thesis Series, the Netherlands

Hegeman, G., *Assisted Overtaking. An assessment of overtaking on two-lane rural roads*, T2008/4, February 2008, TRAIL Thesis Series, the Netherlands

Katwijk, R.T. van, *Multi-Agent Look-ahead Traffic Adaptive Control*, T2008/3, January 2008, TRAIL Thesis Series, the Netherlands

Argiolu. R., *Office Location Choice Behaviour and Intelligent Transport Systems*, T2008/2, January 2008, TRAIL Thesis Series, the Netherlands

Houtenbos, M., *Expecting the Unexpected. A study of interactive driving behaviour at intersections*, T2008/1, January 2008, TRAIL Thesis Series, the Netherlands

Negenborn, R.R., *Multi-Agent Model Predictive Control with Applications to Power Networks*, T2007/14, December 2007, TRAIL Thesis Series, the Netherlands

Nederveen, A.A.J., *Ruimtelijke Inpassing van Lijninfrastructuur. Een onderzoek naar de geschiktheid van inspraakreacties voor het ontwerpen van lijninfrastructuur*, T2007/13, December 2007, TRAIL Thesis Series, the Netherlands

Nuttall, A.J.G., *Design Aspects of Multiple Driven Belt Conveyors*, T2007/12, November 2007, TRAIL Thesis Series, The Netherlands

Gietelink, O.J., *Design and Validation of Advanced Driver Assistance Systems*, T2007/11, November 2007, TRAIL Thesis Series, the Netherlands

Driel, C.J.G. van, *Driver Support in Congestion: an assessment of user needs and impacts on driver and traffic flow*, T2007/10, November 2007, TRAIL Thesis Series, the Netherlands

Warffemius, P.M.J., *Modeling the Clustering of Distribution Centers around Amsterdam Airport Schiphol. Location Endowments, Economies of Agglomeration, Locked-in Logistics and Policy Implications*, T2007/9, September 2007, TRAIL Thesis Series, the Netherlands

Joksimovic, D., *Dynamic Bi-Level Optimal Toll Design Approach for Dynamic Traffic Networks*, T2007/8, September 2007, TRAIL Thesis Series, the Netherlands

Vlist, P. van der, *Synchronizing the retail supply chain*, T2007/7, June 2007, TRAIL Thesis Series, the Netherlands

Fiorenzo-Catalano, M.S., *Choice Set Generation in Multi-Modal Networks*, T2007/6, June 2007, TRAIL Thesis Series, the Netherlands

Bok, M.A. de, *Infrastructure and Firm Dynamics: A micro-simulation approach*, T2007/5, May 2007, TRAIL Thesis Series, the Netherlands

Zondag, B., *Joined Modeling of Land-use, Transport and Economy*, T2007/4, April 2007, TRAIL Thesis Series, the Netherlands

Weijermars, W.A.M., *Analysis of Urban Traffic Patterns Using Clustering*, T2007/3, April 2007, TRAIL Thesis Series, the Netherlands

Chorus, C., *Traveler Response to Information*, T2007/2, February 2007, TRAIL Thesis Series, the Netherlands

Miska, M., *Microscopic Online Simulation for Real time Traffic Management*, T2007/1, January 2007, TRAIL Thesis Series, the Netherlands

Makoriwa, C., *Performance of Traffic Networks. A mosaic of measures*, T2006/10, December 2006, TRAIL Thesis Series, The Netherlands