Integrated Control of Mixed Traffic Networks using Model Predictive Control

M. van den Berg

Integrated Control of Mixed Traffic Networks using Model Predictive Control

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op donderdag 22 april 2010 om 15.00 uur
door Monique VAN DEN BERG,
Elektrotechnisch ingenieur,
geboren te Leiderdorp.

Dit proefschrift is goedgekeurd door de promotoren:
Prof.dr.ir. J. Hellendoorn
Prof.dr.ir B. De Schutter

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof.dr.ir. J. Hellendoorn	Technische Universiteit Delft, promotor
Prof.dr.ir. B. De Schutter	Technische Universiteit Delft, promotor
Prof.dr. H.J. van Zuylen	Technische Universiteit Delft
Prof.dr. K.A. Brookhuis	Technische Universiteit Delft
Prof.dr.ir. E.C. van Berkum	Universiteit Twente
Prof.ir. L.H. Immers	Katholieke Universiteit Leuven
Prof.Dr.-Ing. M. Papageorgiou	Technical University of Crete

The research described in this thesis was supported by the project NWO-CONNEKT project "Advanced Multi-Agent Control and Information for Integrated Multi-Class Traffic Networks (AMICI)", and by the BSIK project "Transition Sustainable Mobility (TRANSUMO)".

TRAIL Thesis Series T2010/4, The Netherlands TRAIL Research School
P.O. Box 5017
2600 GA Delft, the Netherlands
T: +31 (0) 15 278 6046
T: +31 (0) 15 278 4333
E: info@rstrail.nl

Published and distributed by: M. van den Berg
E-mail: vandenberg.monique@gmail.com

ISBN 978-90-5584-128-8

Keywords: traffic control, route choice control, model predictive control.

Printed in the Netherlands

Preface

Here it is... my Ph.D. thesis. It took me 6 years to create, including only 3.5 years of full-time work. How this could have happened? Well, I first did not want to take the job, then tried to follow a lot of DISC courses, got pregnant, twice, worked 3 days a week, rebuild my house, developed no programming skills, and started a new job too early. This means that I owe a lot of thanks to my supervisors Hans Hellendoorn and Bart De Schutter. They endured all my adventures, and even supported and helped me. They took care of the necessary administrative actions, provided personal support, and, most importantly, supported the research itself. Together they form a good team: Hans the person whom I could talk to about a lot of social questions, with a good overview of the research and its impact, and Bart focused on the topic, giving directions to the research, being a discussion partner, and having an eye for the details. Together they helped me discover the interesting parts of doing research, and encouraged me to continue with the less interesting parts. Without them the thesis would not have taken this form.

Further, I greatly appreciate the members of my Ph.D. committee Prof.Dr.-ing. Marcus Papageorgiou, Prof.ir. Ben Immers, Prof.dr.ir. Erik van Berkum, Prof.dr. Karel Brookhuis, and Prof.dr. Henk van Zuylen who took the time to review my thesis. Their constructing comments and remarks helped me to improve my thesis.

Off course, many others have contributed to the work as well. Most of the work described in this thesis is done together with Andreas Hegyi. We used to share our office, our interest in traffic flows and our supervisors. We differ with respect to experience, preferred room temperature, and working hours. However, Andreas changed from being my master thesis supervisor, to being my colleague, to being my friend.

Others which I would like to thank for their contributions to papers and parts of my thesis are Tom Bellemans, Mernout Burger, Bas Heutinck, Andrea Pinna, Antonio Nunziante, Federica Lamon, and Simona Cristaldi.

During the time at the Delft Center for Systems and Control, I enjoyed working with my colleagues, of whom I will especially mention Alina Tarau, Laksmi Baskar, Eric Trottemant, Diederik Joosten, Justin Rice, Jelmer van Ast, Redouane Halluzi, Jelmer Braaksma, Jorn van Doren, Gijs van Essen, Sara van der Hoeven, Ronald van Katwijk, Rudi Negenborn, and Kitty Dukker.

Last but not least, I would like to thank my family Rob, Naomi and Eline for their help in creating this thesis. Thanks to them, or should I say in spite of them, I have been able to perform the work necessary to complete this thesis.

Monique van den Berg,
Delft, January 2010.

Contents

Chapter 1

Introduction

In the near future the amount of traffic on the roads will keep increasing. The capacity of the available infrastructure cannot increase at the same pace, since creating new roads is expensive, time consuming, and requires free space that is difficult to find or not available at all. As a result the number and length of traffic jams will increase during the next years. Traffic jams result in economical costs due to the large delays that they cause, and they also have negative effects on the environment due to, e.g., increases of the noise levels and pollution. A possibility to prevent, or at least reduce, congestion is to make more efficient use of the available roads. This can be reached via traffic control measures, such as traffic signals, variable speed limits, ramp metering installations, and dynamic route information panels. The influence of these measures is largely depending on the values that they use for, e.g., green times or speed limits. These values are determined by control methods, which thereby are an important part of the total traffic control system. Different control methods result in different values for the green times and speed limits, and thus in different performances of the traffic network. To improve the traffic situation, control methods should be developed that select the values for the control measures in such a way that the performance of the traffic network is increased, taking into account the interests of, e.g., drivers, government, and the environment.

In this thesis we will develop advanced control methods that improve the performance of the traffic network. In particular, controllers for mixed urban/freeway networks, controllers that anticipate on route choice, and route choice control methods are developed. The developed control methods are based on model predictive control (MPC). MPC is a control method that uses a prediction model to determine the expected evolution of the traffic flows. Furthermore, within MPC an objective function is defined that describes the goals of the controller. Then an optimization algorithm is used to determine the optimal settings for the traffic control measures based on the model predictions and the objective function. If necessary, the method can also handle constraints on the variables and states in the network. Finally, the computed settings for the first control time step are applied by the control measures, and the next control time step the whole procedure is repeated. In this thesis, the objective function of most of the developed control methods is based on the total time the vehicles spend in the network.

The remainder of this chapter is organized as follows. First some background material about traffic networks is presented, including organizations involved with traffic networks, currently available traffic control measures, and traffic control methods. Next, the objectives of the research described in this thesis are formulated. Then the contributions of the thesis and the relevance of the research are presented. Finally, an outline of the whole thesis is given.

1.1 Background

In this section we will describe some aspects of general traffic networks, and in particular of the Dutch road network. We discuss different road types and the corresponding organizational structure, and describe different processes that occur within traffic flows. Further on, we describe the control measures that can be used to influence the traffic, and discuss available control methods that can be used to determine the settings of the control measures.

1.1.1 Road networks

Road networks evolve along with the development of transportation modes. The first roads where used by pedestrians, and during history horses, handhold carts, and horse carts were added. Later bicycles and cars where introduced, and separate railways where created for trains. The first roads where unpaved roads, while nowadays nearly all roads are hardened. The main road network is used by cars and trucks, while separate lanes are available for bikes and sidewalks for pedestrians.

The layout of the roads depends on the location and function of the roads. Freeways are used by long-distance traffic and are in principle uni-directional roads, without intersections. The speed on these roads is high, usually from 100 to 200 km/h. The capacity of these roads is around 2000 veh/h/lane. Highways are bi-directional roads, used for medium distances or for long routes with a low demand. The speed on this roads is usually lower than on freeways. In The Netherlands the maximum allowed speed on this kind of roads is 100 km/h. The capacity of these roads is about 1800 veh/h/lane. For short-distance traffic there are local roads, which are mainly located in urban areas. The capacity of these local roads is small, varying from 500 to 1500 veh/h/lane depending on the road layout. At these local roads there are many intersections, and the speed is around 50 km/h.

The historical evolution of the road network implies that the layout of the current network is not efficient with respect to the current origins and destinations of the traffic and the corresponding traffic flows. This results in a large mismatch between the demand and the available capacity on different routes. This mismatch is an important cause of congestion in the current network.

1.1.2 Traffic related organizations

In most countries general roads are the responsibility of the government. But when we consider, e.g., toll roads, they can also be privately owned. In general, each road type is managed by a different authority. Freeways are often managed by the government, while other roads are managed by municipalities. As an example, we will explain the situation as it is in The Netherlands.

In The Netherlands a distinction is made between local roads, provincial roads, and freeways. The local roads are managed by the municipalities or the cities. The larger cities develop and control their own road network, while smaller villages implement control measures developed by consultancy companies. The provincial roads are managed by the provincial government. Provincial roads connect villages and small cities with each other. These roads are bi-directional roads, with intersections controlled by traffic signals. The Dutch Ministry of Transport is responsible for the freeways in The Netherlands.

The freeway network in The Netherlands is dense, with many on-ramps and off-ramps on each freeway. This means that the different types of roads are strongly connected, which requires cooperation between the different authorities. However, until now the presence of different management authorities has prevented integral coordination and steering of traffic control measures. This has resulted in a situation where each management body solves its own problems by sending the traffic to the roads that are someone else's responsibility. It is easy to imagine that this leads to inefficient road use at the locations where the different road types are connected. Nowadays, most parties start to realize that a solution of the traffic problems can only be obtained when they cooperate. Intention declarations are written, procedures are described, and meetings are planned. The Dutch Ministry of Transportation has formulated an advice of how the whole process could be organized in a handbook, see [110]. This handbook provides guidelines to select important origins and destinations, and to define which roads should be used for the major connections between them. Then advise about possible traffic management measures is given. The handbook also presents the way in which the political process of developing the control management strategy can be organized.

Next to the traffic management bodies, other parties are also involved with the traffic road network. An important group are the road users. They actually experience the delay caused by congestion, and they have to adapt their behavior when control measures are applied. In The Netherlands, the drivers are united in the ANWB, the Dutch drivers' organization. This organization participates in consultations with the governments, informs drivers about the traffic network, participates in research groups, and has a road guard.

Further, consultancy companies play a large role in the development of traffic control measures. They use traffic models to predict the effects of the traffic measures, and give advice to the road managers about the use of the measures. The companies also develop new algorithms that can be used to determine the values of the control measures, see, e.g., [6, 24, 129].

The economical cost for the delays due to traffic jams is expressed in vehicle hours, and was equal to 49 million hours for 63 billion traveled kilometers during the year 2007 [113]. Since the costs of traffic jams are this high, reducing the length and frequency of the jams is of high importance for the government. The government tries to reduce the traffic jams on the short term by implementing available control measures, and it invests in traffic research to develop long-term solutions. The location of new roads is also a political issue, just as the taxes for road use. Furthermore, the government also sponsors the SWOV (Stichting Wetenschappelijk Onderzoek Verkeersveiligheid), an organization that investigates the safety of the road network in general.

1.1.3 Traffic processes

Road traffic has many different aspects. Each of this aspects can be captured with different simulation models.

First of all, the amount of vehicles that want to use the road network should be determined. This number of vehicles depends on the number of trips that the drivers would like to make, and on the transportation mode that the individual travelers select. Possible transportation modes are, e.g., train, bus, car, or airplane. When the mode choice is made, the number of vehicles that want to make a trip is known. This number of vehicles, combined with the corresponding origins and destinations, can be used to determine the origin-destination (OD) matrix. Such a matrix contains the demand from each origin to each destination in the network. Models that determine this OD matrix are based on questionnaires, historical traffic measurements, and on an investigation of the surroundings of the road network to determine the locations of origins and destinations.

The next decision of the drivers that should be modeled is the departure time. Drivers select their departure time based on, e.g., the desired arrival time, the current time, the expected travel time, the probability of a delay, and the expected length of the delay.

Another aspect is the route choice of the drivers. When the origins, destinations, and departure times of the vehicles are known, the route that they take should be determined. This can be determined in two ways: via a route choice model or with a dynamic traffic assignment model. A route choice model describes how drivers react at each location where the road splits. At these locations the drivers have to make a route choice, which they base on, e.g., earlier experiences, and on the current situation. A dynamic traffic assignment model assumes that all drivers select their route in such a way that a user equilibrium assignment appears. A user equilibrium assignment is the assignment that appears when all routes with the same origin and destination have equal travel times.

The next issue that should be considered is the behavior of vehicles on one road. How many vehicles are on each part of the road, how fast do they drive, what are the distances between the vehicles, do they change lanes, etc. Many models exist that describe this behavior. Different models consider different levels of detail, and describe the process in different ways. Microscopic models give a detailed description of the traffic, modeling the behavior of individual vehicles. Mesoscopic models consider probability distributions of the variables, while macroscopic models consider average values for speeds, densities, and flows. The computational effort for the detailed models is high, while the low-detail models require shorter computation times. Which model is the most suitable for a given application depends on a trade-off between model accuracy and available computation time. An overview of existing traffic models is given in [72, 124].

1.1.4 Control measures

Traffic flows can be influenced via traffic control measures, such as traffic signals, variable speed limits, ramp metering installations, and dynamic route information panels, [112]. Most of these measures are originally designed to improve the traffic safety, but they also have a significant influence on other aspects of the traffic flows. We will now give a short description of each of these measures.

Traffic signals are used to control the right of way at urban intersections, see Figure

Figure 1.1: Traffic signal installation.

1.1. They prevent accidents, and influence the throughput of intersections. At each link connected to an intersection a signal is located, with three lights: red, amber, and green. During the red period, the vehicles are not allowed to pass the intersection, during the green period they have to drive through. The amber period is used as a warning period between the green and red, vehicles that are able to stop should stop, while vehicles arriving with a high speed, unable to make a safe stop, can still cross the intersection. The first traffic signals had to be set manually by police agents, see [152]. Later computers were used, first to coordinate the lights at individual links at the intersections, next to allow the operator at a traffic control center to monitor and influence the measures, and finally for automated coordinated control. The coupling with available traffic detectors has led to the development of vehicle-actuated control, which is the main control method used in The Netherlands. Vehicle-actuated control uses detectors to determine whether there is a queue at each link, and base the green times on these measurements.

Variable speed limits, see Figure 1.2, can be used to control the speed on freeways. They are usually displayed on variable message signs along the road. The initial purpose of introducing the speed limits was to improve the safety. The speed limit values were lowered under bad weather conditions, and the limits were used to warn drivers when they approach a traffic jam. An example of such a warning system is the Dutch incident detection system, which has been developed to reduce the number of secondary accidents in congested areas. This is obtained by lowering the speed limits to warn the drivers that they approach an accident so they can lower their speed, which reduces the probability of head-tail collisions. Speed limits can further be used to improve the performance of the freeway network, as described in, e.g., [66, 145]. Reducing the speed upstream of a traffic jam lowers the amount of traffic that enters the jam, and thus allows the jam to dissolve faster, which increases the throughput of the freeway.

At on-ramps, ramp metering installations can be used, see Figure 1.3. Ramp metering installations are traffic signals located on on-ramps, and they allow one or two vehicles to

Figure 1.2: Variable speed limits.

pass during each green period. The purpose is to control the flow on the on-ramp in order to minimize the disturbance of the traffic on the freeway, see [122]. The ramp metering installation limits the flow that can enter the freeway, and it also reduces the peaks in the flow by dividing the flow equally over time. This reduction of the disturbances on the freeway traffic lowers the amount of congestion that is induced by the on-ramp flow. Another effect of ramp metering is that it can influence the route choice. The on-ramp installation causes a queue on the on-ramp, which generates extra travel time for the drivers. When this time becomes too large, drivers change their route in such a way that they enter the freeway at another location. A disadvantage of ramp metering installations is that they can have a negative impact on the urban traffic. When the queue that appears on the on-ramp spills back into the urban network, it can block urban intersections and also delay traffic that does not want to enter the freeway at all. The effect of this disadvantage can be reduced by applying queue management actions, which often consist of adapting the ramp metering rate when the queue length exceeds the available space. This however reduces the effect of the ramp metering on the freeway traffic.

A less direct control method is the use of dynamic route information panels (DRIPs). DRIPs are located at splitting nodes of the network, see Figure 1.4. The original purpose of the DRIPs is to inform the drivers about the current state of the traffic. Queue lengths or travel times on the different routes are presented, to allow drivers to take an informed route choice decision [18, 103]. The use of DRIPs within control systems are described in [48, 76]. In this thesis we illustrate how DRIPS can be used to persuade drivers to change their route choice in order to obtain a traffic assignment that gives a more optimal traffic performance from the system point of view.

Figure 1.3: On-ramp metering installation.

Figure 1.4: Dynamic route information panel.

1.1.5 Available control methods

The settings of the control measures are determined via control methods. The general idea of a control method is that it obtains some measurements from the traffic flows of the network, and uses these measurements to determine the settings for the control measures, see Figure 1.5. The number and type of measurements that is considered can be used to classify different control methods.

Fixed-time control methods are the oldest type of control methods, and in contrast to more recent methods, do not use real-time measurements at all. The settings of the control measures are constant or changing according to the time-of-day, often determined using historical measurements. Fixed-time control is still frequently used. For isolated intersections,

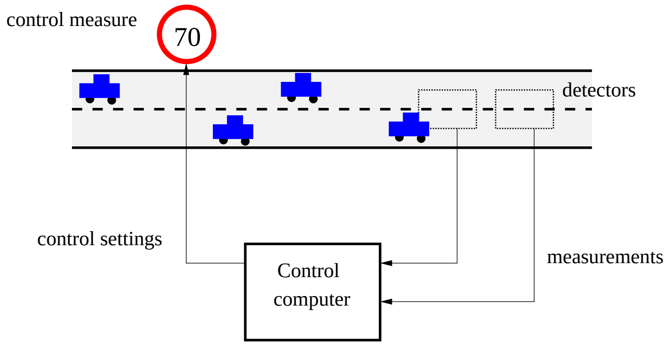


Figure 1.5: General principle of traffic control.

or intersections with a low demand the performance is good. The main disadvantage is that the method does not react on variations in the traffic flow, and that the settings can become out-dated, since the method does not take into account the increase of demand over time.

A second class of traffic control methods consist of vehicle-actuated systems. These systems use detectors in the immediate area of the control measure. Based on the obtained measurements, changes are applied to the fixed-time control scheme. Examples of such changes are switching between different schemes, or extending a green period [91, 183]. These methods react on variations on the traffic flow, and have proved to be very effective in urban areas. Also for ramp metering installations this kind of methods are used frequently. Based on measurements at the freeway and on the on-ramp the ramp metering rate can be determined, as described in [29, 122]. The Dutch algorithm for incident detection using variable speed limits is also vehicle-actuated.

The last class of control methods contains the traffic-responsive systems, e.g., [48, 129, 190]. These systems use traffic measurements of a large area around the traffic measure. The control methods optimize the actions of the control measures with respect to the whole network. These methods also allow for integrated control, using several measures. The coordination between the control actions of the different measures can further improve the performance of the traffic network.

In [173], an extensive overview of different control methods for traffic signals is given. A more general overview, also considering other control measures, is given in [124].

1.2 Research objectives

In this section we describe the objectives of this thesis. The thesis project is part of a larger research project, which will be described first. Next we formulate the research problem, and elaborate the research goals and the research approach.

1.2.1 Research project

The research described in this thesis is funded by the NWO-Connect projects AMICI, and the BSIK-TRANSUMO project ATMA-MODeRN.

The NWO-Connect research program AMICI considers Advanced Multi-agent Information and Control for Integrated multi-class traffic networks. The total program focuses on traffic congestion management in and around large cities, such as Beijing, Rotterdam, Amsterdam, and Shanghai. In particular, it aims at developing solutions to efficiently reduce traffic congestion by means of dynamic traffic management. Within the program five research topics have been defined:

Multi-class traffic flow theory for modeling of motorway and urban traffic networks This research focuses on the development of a multi-class macroscopic traffic flow model for freeways as well as urban roads [115, 116].

Impact of travel information and traffic control on travel behavior The effects of providing travel time information are investigated within this project. Based on this investigation, a route choice model is developed that includes the effects of providing travel time information [17, 18].

Optimal presentation of travel information based on personal preferences and needs The way in which information is provided influences the effect of the information. This project focuses on the relation between the way of presenting information and the corresponding reactions of the drivers [51].

Market for traveler information Generating traffic information costs money. This project investigates the amount of money that users are prepared to spend on different types of information. Further it investigates the relation between the penetration rate (the percentage of equipped vehicles) and the value of the provided information.

Development of advanced multi-agent control strategies for multi-class traffic networks This project investigates the development of advanced control strategies for multi-class traffic flows, using control measures as well as providing information to improve the performance of existing road networks.

This thesis describes results in the framework of the last project, in particular the development of control strategies to reduce congestion.

BSIK-TRANSUMO is a research program funded by the Dutch government in which universities, companies, the government, and research institutes perform research on sustainable mobility. Within BSIK-TRANSUMO there are several main projects, one of which is ATMA-MODeRN (Advanced Traffic MAnagement - Multi-Objective Decision aid for Regional Networks). The objective of the ATMA-MODeRN project is to develop support systems for the application of traffic management systems in regional road networks, consisting of freeways as well as local roads. Hereby, reliability and sustainability are explicitly taken into account as measures for the performance of the traffic network.

1.2.2 Problem formulation

Congestion on traffic roads has different causes, e.g., accidents, incidents, and bottlenecks. The major cause of congestion is the difference between the demand and the capacity of the road. An investigation of recurrent congestion shows that this lack of capacity is nearly always the largest at locations near on-ramps, off-ramps, lane drops, or at intersections

[111]. This is due to the fact that the different roads merge, and that thus the inflow of such a location is often higher than the outflow capacity. These locations are called bottlenecks. Further, congestion can also appear in the middle of a freeway stretch. In this case it often starts at a bottleneck, but travels upstream over the freeway, causing a delay for all vehicles that travel on the freeway.

Many bottleneck locations are equipped with traffic control measures, which can be used to reduce the amount of congestion. However, the currently used control methods are in general not sufficient to solve the congestion. Part of the measures that are present have been in use for a long time without changing the settings, which means that they were designed for demands that were lower than the current demands, while others are only solving local problems, which usually only results in a re-location of the congestion. In dense road networks, where the number of traffic control measures is high, the influence of the measures on each other is large. This means that in dense networks solving local problems for each measure separately often does not lead to a good overall performance of the whole network, but only re-locates the congestion. Coordination of the available measures however can significantly improve the performance. For urban areas there already are systems that coordinate the control actions of traffic signals [49, 129, 140], and for freeways coordinated control systems for variable speed limits are available [2, 66]. However, for networks that contain freeways as well as urban roads, coordinated control methods are not available yet.

Within larger networks, control measures do not only affect the throughput of the network, but they also influence the route choice, see [148]. This is due to the fact that they change the travel times on different routes, which might make alternative routes more attractive for the drivers. As a result, the flows in the network change, and thus the demands at the bottleneck locations change. The change in route choice can also re-locate the congestion to undesired locations, creating, e.g., large queues in a nature reserve (which generates pollution), large queues in residential areas (which reduces the safety), or congestion at intersections (blocking crossing traffic). To prevent this undesired re-location of congestion, control measures should adapt their settings. However, the currently applied control methods do not take route choice into account. Thus the performance of these traffic control methods can be improved further when the methods do consider the change of route choice they can induce [10].

The last problem that we will consider in this thesis is related to the process of applying model-based control methods in practice. Before a control method can be implemented, a number of issues should be considered. The control objectives, control method, and infrastructure layout should be determined. Some guidelines for the selection process and for the implementation of dynamic traffic management are given in [110]. One of the important issues for the implementation is the selection of an averaging method for measurements. The measurements that are obtained from the network should be averaged over time. This can be done using different averaging methods. The influence of the selected averaging method on the performance of traffic control methods will be investigated in this thesis.

1.2.3 Research goals and approach

The main goal of the research described in this thesis is to develop control methods that can be used to increase the efficiency of road use, and to reduce the negative effects of congestion. Within the research, the main goal is divided into sub-goals, which correspond to the problems formulated above:

Control of mixed networks Mixed networks consist of freeways as well as urban roads. Bottlenecks that cause congestion often appear at locations where both road types are present. Therefore, we aim at developing control methods that coordinate the traffic control measures in such mixed networks. The control methods should integrate the different control measures and coordinate the control actions on freeways and urban roads, leading to control methods for network-wide integrated control. Challenges with respect to coupling freeway and urban networks are the different driver behavior (which results in different model types), the different time scales of the processes that appear, and the difference in available control measures.

Influencing route choice Control measures take control actions that influence the travel times on different routes, and thus influence the route choice of the drivers. We want to develop control methods that take this effect into account, and thus can determine more optimal settings for the control measures in networks where multiple routes are available. Further, we want to develop control methods that can re-locate congestion to locations where the impact is less severe, such as roads outside residential areas, or roads that are not on a main supply route. This implies that the control method should actively influence the route choice and change the traffic assignment.

Implementation aspects Before controllers can be applied in practice, many issues should be considered. We present an overview of the relevant implementation issues. With respect to implementing controllers much information is available; however, the structure of the whole implementation process is not clearly defined. Further, we focus on the effect of averaging methods for speed measurements when they are used in a control method. Differences in measurement methods can affect most steps in the controller design and implementation process, and thus can significantly influence the controller performance.

Each sub-goal is investigated in one or two separate studies. For most studies we follow the following approach. First, a literature study is performed in which an overview of the current research is obtained. Then the controller design is started with the selection of the model that is used in the controller. When no suitable model is found in literature, an existing model is adapted, or a new model is developed. This model is used for the remainder of the controller design process. When the design of the controller finished, a simulation study is performed to illustrate the performance of the developed control method.

1.3 Contributions of this thesis

This thesis contributes to the state-of-the-art with respect to advanced traffic control methods. The main contribution is the further development of advanced control methods based on model predictive control. These control methods can be used to increase the efficiency of road use, they can handle varying demands, and explicitly take hard constraints into account. The innovative contributions with respect to the state-of-the-art are the modelling and control of mixed networks, and the integrated control methods to influence route choice. The control methods are especially suitable for network control, and coordinate the actions of the various available control measures. The developed traffic controllers react to the traffic situation on the whole network, and optimize the control settings accordingly. Different controllers are designed, each using a different model and thus suitable for different situations:

- The first control method targets mixed networks, that contain freeways as well as urban roads. The control method reduces the total time the vehicles spend in these networks by coordinating the control actions of traffic signals, variable speed limits, and on-ramp metering installations.

- The second type of controllers that is developed influences the within-day route choice of drivers. The controllers use a traffic flow model to describe the evolution of the traffic flows, and a route choice or a traffic assignment model to determine the traffic assignment in the network. The predictions obtained with these models are used to select the values for the control measures in such a way that the route choice of drivers is steered actively.

- The third controller is designed to influence the day-to-day route choice. The traffic control measures changes the travel times on different routes in the network. Due to these changed travel times that the drivers experience at the current day, the drivers will select another route during the next day. By predicting the changes in route choice that will result from the control actions, the control method affects the route choice in such a way that the drivers select the most optimal route after a few days.

The controllers described above are all using prediction models. For the design of the controllers we have adapted or developed the following models:

- A model for mixed networks is created by coupling the macroscopic freeway flow model Metanet [106] with an urban queue length model developed by Kashani and Saridis [82]. The urban model is first extended to include horizontal queues, blocking effects, and a smaller time step, and next the on-ramps and off-ramps are modeled.

- A density-dependent route choice model based on Bayesian learning is developed. It is a combination of a day-to-day learning model based on experienced travel times, and of a fast look-up table containing turning rates based on densities for the within-day route process. The possible densities are divided into three groups, and based on the current density the model selects a value from the look-up table.

- A basic route choice model is developed that describes the day-to-day route choice. This model determines the travel times on different routes, and adapts the turning rate accordingly. The most basic version of the model allows for analytical solutions of the optimization problem within the controller, while the further developed variants lead to mixed integer linear optimization problems, for which efficient solvers are available.

Note that during the current research the models are developed, but not validated. Before the models could be used in practice, the validation must be performed, and thus this will be part of our future work.

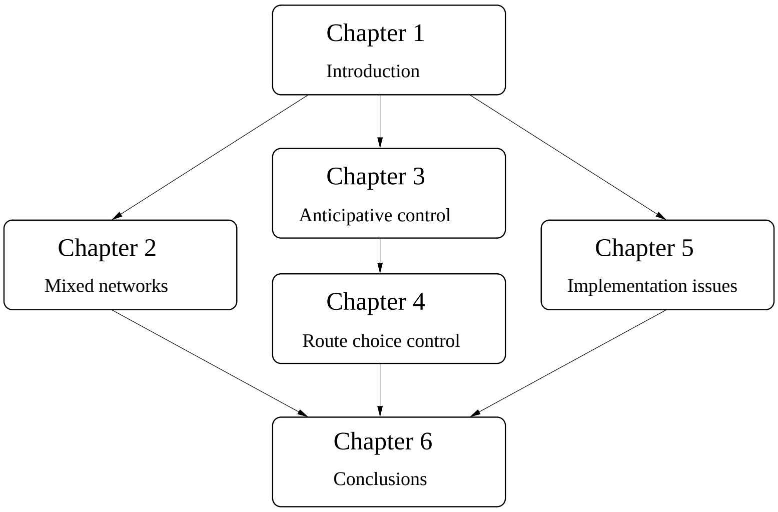


Figure 1.6: Overview of the relation between different chapters.

Another contribution of this thesis considers the effect of measurements on the controller performance. Different averaging methods for speed measurements are considered. Each of the averaging methods is used in a controller, and the performance of the controllers is compared.

The practical and social relevance of the thesis mainly consists of improving the efficiency of road use. The proposed control methods can reduce the amount of congestion, and thus reduce the experienced delays, economical costs, noise generated, and pollution caused by traffic jams. The route choice controller can be used to reduce the amount of traffic in urban areas or nature reserves, while the controller for mixed networks is a tool that can be used for integrated control when different road management bodies cooperate.

1.4 Thesis outline

In this section we give an outline of the thesis. The sub-goals of the research as formulated in Section 1.2.3 are described in a separate chapters. Each chapter is written in such a way that it can be read separately and independently. This means that readers of the whole thesis will encounter several repetitions, however, this can be justified by the large differences in the considered topics, which will attract many readers interested in only a part of the thesis. An overview of the relations between the chapters is presented in Figure 1.6. The arrows represent the relations between the chapters, and indicate how the chapters are divided over the three different sub-goals.

Chapter 2 considers the integrated network-wide control of mixed networks containing freeways as well as urban roads. A model is developed that describes the traffic flows on freeways and on urban roads. A model predictive control based controller is developed that uses this model for the control of mixed networks. The controller integrates the control actions of traffic signals, variable speed limits, and ramp metering installations.

In Chapter 3 anticipative route choice controllers are developed. Two different prediction models are used to determine the route choice within the controller: a dynamic traffic assignment algorithm using the method of successive averages, and a route choice model based on Bayesian learning. The results obtained with the anticipating controllers are encouraging, but the computation times are very large. This has led to the investigation of a less detailed route choice controller, which is described in Chapter 4. Three different versions of a basic route choice model are developed, which, when implemented in a controller, result in optimization problems that can respectively be solved analytically, with mixed integer linear programming, and with non-linear optimization algorithms.

Chapter 5 discusses the issues that should be considered when implementing the control strategies in practice. In particular, different averaging methods for speed measurements are considered, and their influence on the performance of the controller is investigated.

Finally, in Chapter 6 conclusions are drawn and topics for future research are given.

Parts of this thesis have already been published in journals or conference proceedings: Chapter 2 in [45, 152, 156, 158, 159, 164], Chapter 3 in [157, 160–162], Chapter 4 in [163, 165–167, 169, 170], and Chapter 5 in [168].

Chapter 2

Integrated traffic control for mixed urban and freeway networks

We develop a control method for networks containing both urban roads and freeways. These two types of roads are closely connected: congestion on the freeway often causes spill-backs leading to urban queues, slowing down the urban traffic. Urban queues can increase until they block off-ramps, causing traffic jams on the freeway. As a consequence, control measures taken in one of the two parts of the network can have a significant influence on the other area. We first develop a model that describes the evolution of the traffic flows in mixed networks. Next, we propose the control method that is used for the integrated control. This approach is based on model predictive control, in which the optimal control inputs are determined on-line using numerical optimization and a prediction model in combination with a receding horizon approach. We also compare our newly developed control method with systems that are similar to existing dynamic traffic control systems like SCOOT and UTOPIA/SPOT, in a qualitative as well as in a quantitative way via a case study. The results illustrate the potential benefits of the proposed approach and motivate further development and improvement of the proposed control method.

2.1 Introduction

The need for mobility is increasing, as can be seen from the growing number of road users as well as from the increasing number of movements per user [111]. This leads to an increase in the frequency, length, and duration of traffic jams, and to increasing queue lengths in the traffic network. The traffic jams cause large delays, resulting in higher travel costs and they also have a negative impact on the environment due to, e.g., noise and pollution. Due to these negative effects dealing with traffic jams has become an important issue.

To tackle the above congestion problems there exist different methods: construction of new roads, levying tolls, promoting public transport, or making more efficient use of the existing infrastructure. In this chapter we consider the last approach, implemented using dynamic traffic management or control, because this approach is effective on the short term, and inexpensive compared to constructing new infrastructure.

Current traffic control approaches usually focus on either urban traffic or freeway traffic. In urban areas traffic signals are the most frequently used control measures. Traditionally, they are controlled locally using fixed-time settings, or they are vehicle-actuated, meaning that they react on the prevailing traffic situation. Nowadays sophisticated, dynamic systems are also making progress. They coordinate different available control measures to improve the total performance. Systems such as SCOOT [140], SCATS [185], Toptrac [6], TUC [48], Mitrop [59], Motion [24], and UTOPIA/SPOT [129] use a coordinated control method to improve the urban traffic, e.g., by constructing green waves, or to improve the traffic circulation. Control on freeways is done using different traffic control measures. Ramp metering is applied on on-ramps, using systems like ALINEA [123]. Overviews of ramp metering methods and results are given in [121, 149]. The use of variable speed limits on freeways is described in [2, 66, 95, 145], and the use of route guidance in [46, 48, 81]. Several authors have described methods for coordinated control for freeways using different traffic control measures [10, 64, 86, 89].

Several authors have also investigated corridor control [48, 81, 186], where arterials are controlled together with freeways. In this chapter we go one step further, and we describe the coordinated and integrated control of networks that contain both freeways and urban roads, since the traffic flows on freeways are often influenced by traffic flows on urban roads, and vice versa. Freeway control measures like ramp metering or speed limits allow a better flow, higher speeds, and larger throughput but can lead to longer queues on on-ramps. These queues may spill back and block urban roads. On the other hand, urban traffic management policies often try to get vehicles on the freeway network as soon as possible, displacing the congestion toward neighboring freeways. The problems due to the mutual interactions between the two types of roads are often increased by the fact that in many countries urban roads and freeways are managed by different management bodies, each with their own policies and objectives.

By considering a coordinated control approach the performance of the overall network can be improved significantly. Therefore we develop a control approach for coordinated control of mixed urban and freeway networks that makes an appropriate trade-off between the performance of the urban and freeway traffic operations, and that prevents a shift of problems between the two. The new contributions of this chapter with respect to the state-of-the-art are a macroscopic model that describes networks that contain both urban roads and freeways, and an integrated control method that takes the traffic flows on both types of roads into account. In addition, this chapter contains a case study in which different control methods are compared in a qualitative and a quantitative way.

As control method we propose a model predictive control (MPC) approach [25, 100]. MPC is an on-line model-based predictive control approach that has already been applied successfully to coordinated control of freeway networks [10, 64, 89]. MPC optimizes the settings of the control measures over a certain prediction horizon. Using a receding horizon approach, only the first step of the computed control signal is applied, and next the optimization is started again with the prediction horizon shifted one time step further.

As MPC requires a model to predict the behavior of the traffic, we will first develop a traffic model for networks that contain both urban roads and freeways. Traffic flow models can be distinguished according to the level of detail they use to describe the traffic. In this chapter we use a macroscopic model. Macroscopic models are suited for on-line control since these models give a balanced trade-off between accurate predictions and computational efforts. The computation time for a macroscopic model does not depend on the number of vehicles in the network, making the model well-suited for on-line control, where the prediction should run on-line in an optimization setting, which requires that the model should run several times faster than real-time, and where the results should always be available within a specified amount of time. Examples of macroscopic models are the LWR model [96, 139], the models of Helbing [68] and Hoogendoorn [71], and METANET [106]. An overview of existing models is given in [72].

In particular, we use an extended version of the METANET traffic flow model to describe the freeway traffic, and a modified and extended model based on a queue length model developed by Kashani and Saridis [82] for the urban traffic. We also discuss how the freeway and the urban model have to be coupled. This results in a macroscopic model for mixed networks with urban roads and freeways, especially suited for the MPC-based traffic control approach developed in this chapter.
In a case study we illustrate how the developed MPC control method performs with respect to existing control systems. A simple benchmark network is simulated, and basic implementations of existing control systems are applied. The performance of these existing systems is compared with our theoretical MPC method. The results of this case study motivate the further development of the MPC method.

The remainder of the chapter is organized as follows. We first describe the model for mixed urban and freeway networks in Section 2.2. Next we develop the MPC-based traffic control method in Section 2.3, and in Section 2.4 we compare the developed method with systems similar to existing methods like SCOOT [140] and UTOPIA/SPOT [129].

2.2 Model development

As indicated above the model for mixed networks containing both urban roads and freeways that we develop is based on the METANET model [106] for the freeway part, and on a queue length model based on a model developed by Kashani [82] for the urban part. Since we want to use the model in an on-line control method, we have selected deterministic models, mainly because they require less computational efforts than probabilistic models. An overview of important symbols of these models is given in Appendix 2.A.

Note that we will explicitly make a difference between the simulation time step T_f for the freeway part of the network, the simulation time step T_u for the urban part of the network, and the controller sample time T_c. We will also use three different counters: k_f for the freeway part, k_u for the urban part, and k_c for the controller. For simplicity, we assume that T_u is an integer divisor of T_f, and that T_f is an integer divisor of T_c:

$$T_\mathrm{f} = N_\mathrm{fu} T_\mathrm{u}, \;\; T_\mathrm{c} = N_\mathrm{cf} T_\mathrm{f} = N_\mathrm{cf} N_\mathrm{fu} T_\mathrm{u},$$

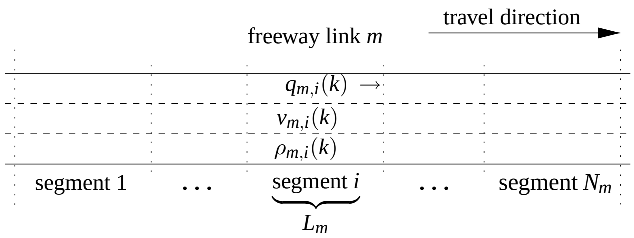


Figure 2.1: A freeway link in the METANET model divided in segments.

with N_{fu} and N_{cf} integers. The value for T_{f} must be selected in such a way that no vehicle can cross a freeway segment in one time step, which results in a typical value of 10 s for freeway segments of 0.5 km. The value of T_{u} is selected small enough to obtain an accurate description of the traffic, typically between 1 and 5 s, depending on the length of the roads. In our case study we will select T_{c} to be 120 s, because for an on-line controller T_{c} should be long enough to determine the new control signal, which depends on the required computation time, and short enough to deal with changing traffic conditions.

2.2.1 Freeway traffic model

In order to model traffic flows in the freeway part of the network we use the destination-independent version of the METANET model, developed by Papageorgiou and Messmer [106]. This model is also used in earlier work for the coordinated control of freeways [10, 64, 86, 89]. A disadvantage of the model is that the description of the transition between congestion and free flow does not completely correspond to what can be observed from real-life measurements. Since the control method that we will develop should have the largest influence at this transition moment, the mismatch might reduce the effect of the controller. This means that using a model that describes this effect more accurate, might improve the performance of the controller. However, the METANET model is sufficiently accurate for other traffic situations, forms a good trade-off between accuracy and computation time, and there is many knowledge about the model. In this chapter we add an extension to the model to obtain a better modeling of the outflow toward off-ramps when blocking phenomena on the off-ramp occur. For completeness we will first describe the original METANET model based on [106], and next present the extension.

Basic METANET model

In the METANET model the freeway network is divided into links. Each link m is further divided in segments, as illustrated in Figure 2.1. All the segments in a link have the same characteristics, e.g., number of lanes, capacity, length, etc.

The traffic state in each segment i of link m at time $t = k_{\mathrm{f}} T_{\mathrm{f}}$ is described with the macroscopic variables average density $\rho_{m,i}(k_{\mathrm{f}})$ in veh/km/lane, space mean speed $v_{m,i}(k_{\mathrm{f}})$ in km/h, and average flow $q_{m,i}(k_{\mathrm{f}})$ in veh/h.

The outflow of segment i of link m at time step k_{f} is given by:

$$q_{m,i}(k_{\mathrm{f}}) = \rho_{m,i}(k_{\mathrm{f}}) v_{m,i}(k_{\mathrm{f}}) n_m \tag{2.1}$$

where n_m denotes the number of lanes of link m. The density in each segment evolves as follows:

$$\rho_{m,i}(k_\mathrm{f}+1) = \rho_{m,i}(k_\mathrm{f}) + \frac{T_\mathrm{f}}{L_m n_m}\big(q_{m,i-1}(k_\mathrm{f}) - q_{m,i}(k_\mathrm{f})\big)$$

where L_m denotes the length of the segments in link m. This equation represents the law of conservation of vehicles: no vehicles appear or disappear within a link.

The equation for the evolution of the speed contains three main terms. The relaxation term expresses that the drivers try to achieve a desired speed $V(\rho)$ for the current density ρ. The convection term expresses that the speed changes due to the inflow of vehicles with a different speed, and the anticipation term expresses that drivers change their speed when the downstream density changes. The updated speed is then computed with:

$$v_{m,i}(k_\mathrm{f}+1) = v_{m,i}(k_\mathrm{f}) + \frac{T_\mathrm{f}}{\tau}\Big(V\big(\rho_{m,i}(k_\mathrm{f})\big) - v_{m,i}(k_\mathrm{f})\Big) + \frac{T_\mathrm{f}}{L_m} v_{m,i}(k_\mathrm{f})\big(v_{m,i-1}(k_\mathrm{f}) - v_{m,i}(k_\mathrm{f})\big)$$

$$- \frac{\nu T_\mathrm{f}}{\tau L_m}\frac{\rho_{m,i+1}(k_\mathrm{f}) - \rho_{m,i}(k_\mathrm{f})}{\rho_{m,i}(k_\mathrm{f}) + \kappa} \tag{2.2}$$

where τ, ν and κ are model parameters. They can be identified from data as described in [88]. The desired speed $V(\rho_{m,i}(k_\mathrm{f}))$ is given by:

$$V\big(\rho_{m,i}(k_\mathrm{f})\big) = v_{\mathrm{free},m} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k_\mathrm{f})}{\rho_{\mathrm{crit},m}}\right)^{a_m}\right] \tag{2.3}$$

where $v_{\mathrm{free},m}$ is the free flow speed on link m, $\rho_{\mathrm{crit},m}$ the critical density on this link, and a_m a model parameter.

Mainstream origins are modeled with a queue model:

$$w_o(k_\mathrm{f}+1) = w_o(k_\mathrm{f}) + T_\mathrm{f}\big(d_o(k_\mathrm{f}) - q_{m,o}(k_\mathrm{f})\big) \tag{2.4}$$

where w_o is the queue length at origin o connected to link m, d_o the demand at origin o, and $q_{m,o}$ the flow leaving origin o toward link m, which is determined by the number of available vehicles, the capacity of the freeway and the traffic conditions on the freeway:

$$q_{m,o}(k_\mathrm{f}) = \min\left(d_o(k_\mathrm{f}) + \frac{w_o(k_\mathrm{f})}{T_\mathrm{f}}, Q_{\mathrm{cap},m}\frac{\rho_{\mathrm{max},m} - \rho_{m,1}(k_\mathrm{f})}{\rho_{\mathrm{max},m} - \rho_{\mathrm{crit},m}}\right) \tag{2.5}$$

where $Q_{\mathrm{cap},m}$ is the capacity of link m and $\rho_{\mathrm{max},m}$ is the maximum possible density on the freeway link.

Freeway links are coupled via nodes, e.g., on-ramps, off-ramps, or intersections. Flows that enter a node p are distributed over the leaving nodes. They are first distributed according to the turning rates[1]:

$$Q_{\mathrm{tot},p}(k_\mathrm{f}) = \sum_{\mu \in I_p} q_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f})$$

$$q_{m,0}(k_\mathrm{f}) = \beta_{p,m}(k_\mathrm{f}) Q_{\mathrm{tot},p}(k_\mathrm{f}) \quad \text{for each } m \in O_p$$

[1] The index 0 in $q_{m,0}(k_\mathrm{f})$ corresponds to a virtual segment that is located upstream of the first segment of link m. This virtual segment is used to describe the traffic that will enter link m.

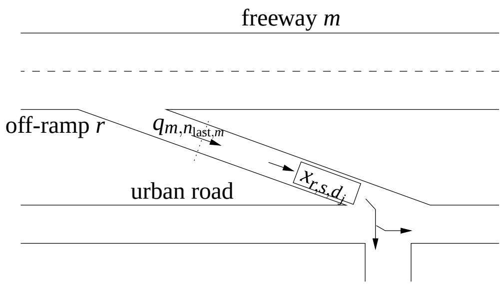


Figure 2.2: Layout of an off-ramp.

where $Q_{\text{tot},p}$ is the total flow entering node p, I_p is the set of all freeway links entering node p, $n_{\text{last},\mu}$ is the index of the last segment of link μ, $\beta_{p,m}$ is the turning rate from node p to leaving link m, and O_p the set of leaving links of node p.

When a node p has more than one leaving link, the virtual downstream density $\rho_{\mu,n_{\text{last},\mu}+1}(k_\text{f})$ of the link μ that enters the node is approximated with:

$$\rho_{\mu,n_{\text{last},\mu}+1}(k_\text{f}) = \frac{\sum\limits_{m\in O_p} \rho_{m,1}^2(k_\text{f})}{\sum\limits_{m\in O_p} \rho_{m,1}(k_\text{f})} \quad .$$

The virtual downstream density is used in the speed update equation (2.2) for the last segment $n_{\text{last},\mu}$ of link μ.

When a node p has more than one entering link, the virtual entering speed $v_{m,0}(k_\text{f})$ of leaving link m is given by:

$$v_{m,0}(k_\text{f}) = \frac{\sum\limits_{\mu\in I_p} v_{\mu,n_{\text{last},\mu}}(k_\text{f}) q_{\mu,n_{\text{last},\mu}}(k_\text{f})}{\sum\limits_{\mu\in I_p} q_{\mu,n_{\text{last},\mu}}(k_\text{f})} \quad .$$

The virtual entering speed is used in the speed update equation (2.2) to compute the speed of the traffic that enters the first segment of link m.

In a link or segment where weaving and/or merging effects are taking place extra terms are added to improve the description of these effects, as described in [89, 106].

Extension for off-ramp links

When the urban network is congested, it often happens that a nearby off-ramp is also blocked. This blockage will spill back onto the freeway. We propose an extension to the METANET model that more accurately models the behavior of off-ramp flows.

Consider an off-ramp r connected to a freeway link m as shown in Figure 2.2. The available space on off-ramp r limits the maximum flow that can enter it. This maximum flow $q_{r,1}^{\max}(k_\text{f})$ is seen as a boundary condition for the flow that leaves the freeway link $q_{m,n_{\text{last},m}}(k_\text{f})$ connected to the off-ramp:

$$q_{m,n_{\text{last},m}}(k_\text{f}) = \min\left(q_{m,n_{\text{last},m}}^{\text{normal}}(k_\text{f}), q_{r,1}^{\max}(k_\text{f})\right) \tag{2.6}$$

where $q_{m,n_{\mathrm{last},m}}^{\mathrm{normal}}(k_\mathrm{f})$ is the flow that would have exited the freeway if the off-ramp would not have been blocked. When the flow is indeed limited to $q_{r,1}^{\max}(k_\mathrm{f})$, the speed of the last segment of the freeway must be recalculated as follows in order to satisfy (2.1):

$$v_{m,n_{\mathrm{last},m}}(k_\mathrm{f}) = \begin{cases} v_{m,n_{\mathrm{last},m}}^{\mathrm{normal}}(k_\mathrm{f}) & \text{if } q_{m,n_{\mathrm{last},m}}^{\mathrm{normal}}(k_\mathrm{f}) < q_{r,1}^{\max}(k_\mathrm{f}) \\ v_{m,n_{\mathrm{last},m}}^{\mathrm{normal}}(k_\mathrm{f}) \dfrac{q_{r,1}^{\max}(k_\mathrm{f})}{q_{m,n_{\mathrm{last},m}}^{\mathrm{normal}}(k_\mathrm{f})} & \text{otherwise} \end{cases}$$

where $v_{m,n_{\mathrm{last},m}}^{\mathrm{normal}}(k_\mathrm{f})$ denotes the speed in the segment when no spill-back occurs, i.e. the speed computed with equation (2.2).

Further extensions describing, e.g., dynamic speed limits, and mainstream metering are given in [64, 66]. The effects of control measures such as ramp metering and variable speed limits will be described in Section 2.3.2.

The external inputs for a simulation of the freeway model are the initial state of the links and the origin queues, and the signals that describe the evolution over the entire simulation period of the turning rates[2] $\beta_{p,m}(k_\mathrm{f})$, the demands $d_o(k_\mathrm{f})$, the boundary conditions $q_{r,1}^{\max}(k_\mathrm{f})$, and the control signals such as the ramp metering rates and the variable speed limits.

2.2.2 Urban traffic model

Several authors have developed models to describe traffic in urban areas [49, 82, 129, 188]. Due to the fact that we want to model and control mixed networks under all conditions, the model we use should satisfy the following requirements:

1. It should be able to describe both light and congested traffic;

2. It should contain horizontal queues because queues often become long compared with buffer capacities, which can lead to blockage of intersections. When an intersection is blocked, no vehicles should be able to cross it.

There are many macroscopic urban traffic models that meet one or more of these requirements, such as the Kashani model [82] and the IN-TUC model [48, 49]. We will base our model on the Kashani model because it has the first of the above features, and because the model can easily be extended. A disadvantage of this model is that the time required to drive from the end of the queue to the intersection is not included. The influence of this approximation depends on the layout of the network, and can be reduced by selecting a proper value of the capacity of the intersection downstream of the link.

Extended urban model

Our model is based on the model developed by Kashani and Saridis [82], but to fulfill all the requirements given above we make the following extensions:

1. Horizontal, turning-direction-dependent queues,

[2]These turning rates can be given externally or they can be determined using a (dynamic) traffic assignment model (see, e.g., [33, 42, 143]).

2. Blocking effects, represented by maximal queue lengths and a flow constraint on flows that want to enter the blocked link, so no vehicle will be able to cross a blocked intersection,
3. A shorter time step[3], to get a more accurate description of the traffic flows.

The main variables used in the urban model are shown in Figures 2.3(a) and 2.3(b). The most important variables are the queue length x expressed in number of vehicles, the number of arriving vehicles m_{arr}, and the number of departing vehicles m_{dep}. Using these variables, the model is formulated as follows.

The number of vehicles that intend to leave the link $l_{o_i,s}$, connecting origin o_i and intersection s, toward destination d_j at time $t = k_{\mathrm{u}}T_{\mathrm{u}}$ is given by:

$$m_{\mathrm{dep,int},o_i,s,d_j}(k_{\mathrm{u}}) = \begin{cases} 0 & \text{if } g_{o_i,s,d_j}(k_{\mathrm{u}}) = 0, \\ \min\big(x_{o_i,s,d_j}(k_{\mathrm{u}}) + m_{\mathrm{arr},o_i,s,d_j}(k_{\mathrm{u}}), \\ \qquad S_{s,d_j}(k_{\mathrm{u}}), T_{\mathrm{u}}Q_{\mathrm{cap},o_i,s,d_j}\big) & \text{if } g_{o_i,s,d_j}(k_{\mathrm{u}}) = 1, \end{cases} \tag{2.7}$$

where $g_{o_i,s,d_j}(k_{\mathrm{u}})$ a binary signal that is 1 when the specified traffic direction has green, and zero otherwise. This means that $g_{o_i,s,d_j} = 0$ corresponds to a red traffic signal, and $g_{o_i,s,d_j} = 1$ to a green one,[4] T_{u} is the urban step with k_{u} as counter, $x_{o_i,s,d_j}(k_{\mathrm{u}})$ is the queue length consisting of vehicles coming from origin o_i and going to destination d_j at intersection s, $m_{\mathrm{arr},o_i,s,d_j}(k_{\mathrm{u}})$ is the number of vehicles arriving at the end of this queue, $S_{s,d_j}(k_{\mathrm{u}})$ is the free space in the downstream link expressed in number of cars, and $Q_{\mathrm{cap},o_i,s,d_j}$ is the saturation flow[5].

The free space $S_{\sigma,s}$ in a link $l_{\sigma,s}$ expresses the maximum number of vehicles that can enter the link. It can never be larger than the length $L_{\sigma,s}$ of the link expressed in number vehicles, and is computed as follows:

$$S_{\sigma,s}(k_{\mathrm{u}}+1) = S_{\sigma,s}(k_{\mathrm{u}}) - m_{\mathrm{dep},\sigma,s}(k_{\mathrm{u}}) + \sum_{d_j \in D_s} m_{\mathrm{dep},\sigma,s,d_j}(k_{\mathrm{u}})$$

where $m_{\mathrm{dep},\sigma,s}(k_{\mathrm{u}})$ is the number of vehicles departing from intersection σ towards link $l_{\sigma,s}$, and D_s is the set of destinations connected to intersection s.

The number of vehicles departing from intersection s towards link l_{s,d_j} can be computed as

$$m_{\mathrm{dep},s,d_j}(k_{\mathrm{u}}) = \sum_{o_i \in O_s} m_{\mathrm{dep},o_i,s,d_j}(k_{\mathrm{u}}).$$

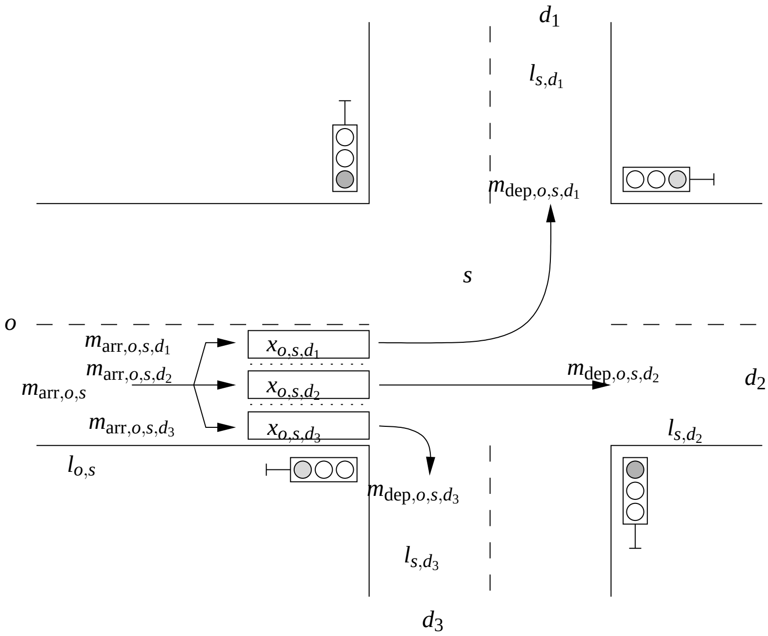
These vehicles drive from the beginning of the link l_{s,d_j} toward the tail of the queue waiting on the link. This gives a time delay $\delta_{s,d_j}(k_{\mathrm{u}})$ which is approximated as:

$$\delta_{s,d_j}(k_{\mathrm{u}}) = \mathrm{ceil}\left(\frac{S_{s,d_j}(k_{\mathrm{u}})L_{\mathrm{av,veh}}}{v_{\mathrm{av},s,d_j}}\right) \tag{2.8}$$

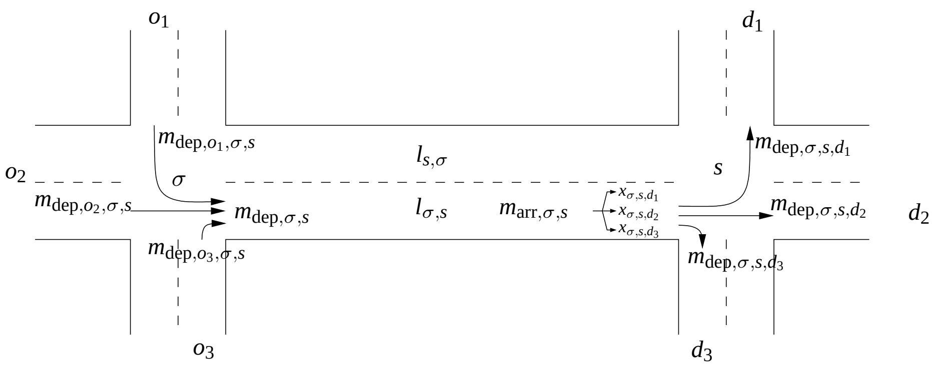
[3] Kashani and Saridis use the cycle time as time step, which restricts the model to effects that take longer than the cycle time. For MPC-based traffic control the other effects can also be interesting, and one might also want to control the cycle times as part of the control measures.

[4] The computed green time is the effective green time. The exact signal timing including the amber time can easily be derived from this effective green time.

[5] The saturation flow is the maximum flow that can cross the intersection under free-flow conditions.



(a) Variables for an urban intersection.



(b) Variables for an urban link.

Figure 2.3: Overview of urban network variables.

where $L_{\text{av,veh}}$ is the average length of a vehicle, and v_{av,s,d_j} the average speed on link l_{s,d_j}.

The time instant at which the vehicle enters the link and the vehicles' delay on the link result in the time instant at which the vehicle will arrive at the end of the queue. It can happen that vehicles that have entered the link at different instants reach the end of the queue during the same time step. To take this into account the variable $m_{\text{arr},s,d_j}(k_\text{u})$ that describes the vehicles arriving at the end of the queue is updated accumulatively every time step. This results in:

$$m_{\text{arr},s,d_j}(k_\text{u}+\delta_{s,d_j}(k_\text{u}))_{\text{new}} = m_{\text{arr},s,d_j}(k_\text{u}+\delta_{s,d_j}(k_\text{u}))_{\text{old}} + m_{\text{dep},s,d_j}(k_\text{u})$$

where $m_{\text{arr},s,d_j}(k_\text{u}+\delta_{s,d_j}(k_\text{u}))$ is the number of vehicles arriving at the end of the queue at time $k_\text{u}+\delta_{s,d_j}(k_\text{u})$, and $m_{\text{dep},s,d_j}(k_\text{u})$ the number of vehicles entering link l_{s,d_j}.

The traffic flow reaching the tail of the queue in link l_{s,d_j} divides itself over the subqueues according to the turning rates $\beta_{o_i,s,d_j}(k_\text{u})$:

$$m_{\text{arr},o_i,s,d_j}(k_\text{u}) = \beta_{o_i,s,d_j}(k_\text{u})m_{\text{arr},o_i,s}(k_\text{u}).$$

The subqueues are then updated as follows:

$$x_{o_i,s,d_j}(k_\text{u}+1) = x_{o_i,s,d_j}(k_\text{u}) + m_{\text{arr},o_i,s,d_j}(k_\text{u}) - m_{\text{dep},o_i,s,d_j}(k_\text{u}).$$

The total flow entering a destination link consists of several flows from different origins. The available space in the destination link should be divided over the entering flows, since the total number of vehicles entering the link may not exceed the available space. We divide this available space equally over the different entering flows. When one flow does not fill its part of the space, the remainder is proportionally divided over the rest of the flows. To illustrate how the effective values of $m_{\text{dep},o_i,s,d_j}(k_\text{u})$ can be computed let us assume that there are two origins, and so two queues from which vehicles want to drive into the same link. Let $m_{\text{dep,int},1}(k_\text{u})$ and $m_{\text{dep,int},2}(k_\text{u})$ denote the number of vehicles that intend to enter the link l_{s,d_j} from respectively origin 1 and origin 2. If we assume without loss of generality that $m_{\text{dep,int},1}(k_\text{u}) \leq m_{\text{dep,int},2}(k_\text{u})$, then the effective values for $m_{\text{dep},1}(k_\text{u})$ and $m_{\text{dep},2}(k_\text{u})$ can be computed as follows:

- if $m_{\text{dep,int},1}(k_\text{u}) + m_{\text{dep,int},2}(k_\text{u}) \leq S_{s,d_j}(k_\text{u})$, then

$$m_{\text{dep},1}(k_\text{u}) = m_{\text{dep,int},1}(k_\text{u}) \quad \text{and} \quad m_{\text{dep},2}(k_\text{u}) = m_{\text{dep,int},2}(k_\text{u}) \quad ,$$

- if $m_{\text{dep,int},1}(k_\text{u}) + m_{\text{dep,int},2}(k_\text{u}) \geq S_{s,d_j}(k_\text{u})$, then

$$\begin{cases} \begin{aligned} m_{\text{dep},1}(k_\text{u}) &= m_{\text{dep,int},1}(k_\text{u}) \\ m_{\text{dep},2}(k_\text{u}) &= S_{s,d_j}(k_\text{u}) - m_{\text{dep,int},1}(k_\text{u}) \end{aligned} & \text{if } m_{\text{dep,int},1}(k_\text{u}) < \frac{1}{2}S_{s,d_j}(k_\text{u}), \\ \\ m_{\text{dep},1}(k_\text{u}) = m_{\text{dep},2}(k_\text{u}) = \frac{1}{2}S_{s,d_j}(k_\text{u}) & \text{if } m_{\text{dep,int},1}(k_\text{u}) \geq \frac{1}{2}S_{s,d_j}(k_\text{u}). \end{cases}$$

The extension to a situation with more upstream queues is straightforward.

The external inputs for a simulation of the urban model are the initial state of the queues, the number of arriving vehicles, and the free space, and the signals that describe the evolution over the entire simulation period of the turning rates $\beta_{o_i,s,d_j}(k_\text{u})$ and of the green/red indicators $g_{o_i,s,d_j}(k_\text{u})$.

2.2.3 Interface between the models

The urban part and the freeway part are coupled via on-ramps and off-ramps. In this section we present the formulas that describe the evolution of the traffic flows on these on-ramps and off-ramps. The main problems are the different simulation time steps T_f and T_u and the boundary conditions that the models create for each other. We assume that the time steps are selected such that $T_\text{f} v_{\text{free},m} < L_m$.

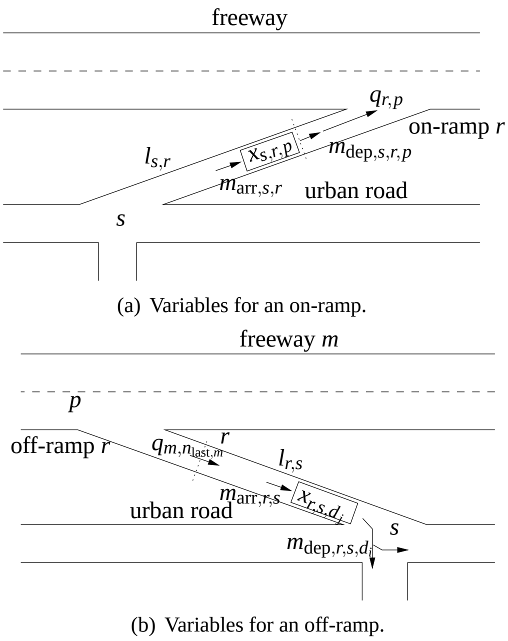


Figure 2.4: Overview of variables on on-ramps and off-ramps.

On-ramps

Consider an on-ramp r that connects intersection s of the urban network to node p of the freeway network, as shown in Figure 2.4(a). The number of vehicles that enter the on-ramp from the urban network is given by $m_{\text{arr},s,r}(k_\text{u})$. These vehicles have a delay $\delta_{s,r}(k_\text{u})$ similar to (2.8). The evolution of the queue length is first described with the urban model. At the end of each freeway time step, the queue length as described in the urban model is then translated to the queue length for the freeway model as explained below.

Now consider the freeway time step k_f corresponding to the urban time step $k_\text{u} = N_\text{fu}k_\text{f}$. In order to get a consistent execution of the urban and freeway models the computations should be done in the following order:

1. Determine the on-ramp departure flow $q_{r,p}(k_\text{f})$ during the period $[k_\text{f}T_\text{f}, (k_\text{f}+1)T_\text{f})$ using (2.5).

2. Assume that these departures spread out evenly over the equivalent urban simulation period $[k_\text{u}T_\text{u}, (k_\text{u}+N_\text{fu})T_\text{u})$. Compute the departures for each urban time step in this period using $m_{\text{dep},s,r,p}(k) = \dfrac{q_{r,p}(k_\text{f})T_\text{f}}{N_\text{fu}}$ for $k = k_\text{u}, \ldots, k_\text{u}+N_\text{fu}-1$ (note that $T_\text{u} = T_\text{f}/N_\text{fu}$).

3. The number of arriving vehicles, the free space, and the queue length $x_{s,r,p}$ at link $l_{s,r}$ can now be computed using the equations for the urban traffic model given in Section 2.2.2.

4. When the queue length $x_{s,r,p}(k_\text{u}+N_\text{fu})$ is computed, we set $w_o(k_\text{f}+1) = x_{s,r,p}(k_\text{u}+N_\text{fu})$. It is easy to verify that this is equivalent to (2.4).

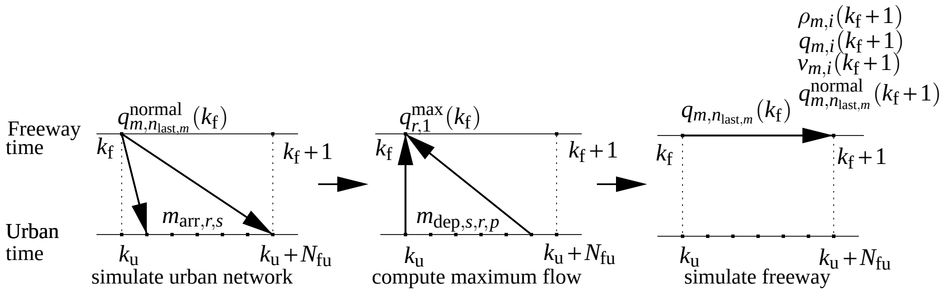


Figure 2.5: Order of computations during a simulation.

Off-ramps

The evolution of the traffic flows on an off-ramp r is computed for the same time steps as for the on-ramp, starting at time step $k_\mathrm{u} = N_\mathrm{fu} k_\mathrm{f}$. The variables are shown in Figure 2.4(b). The following steps are required to simulate the evolution of the traffic flows, in order to get a consistent execution of the urban and freeway models:

1. Determine the number of departing vehicles from link $l_{r,s}$ at intersection s during the period $[k_\mathrm{u} T_\mathrm{u}, (k_\mathrm{u}+N_\mathrm{fu})T_\mathrm{u})$ using the urban traffic flow model.

2. Compute the maximal allowed flow $q_{r,1}^{\max}(k_\mathrm{f})$ that can leave the freeway and enter the off-ramp in the period $[k_\mathrm{f} T_\mathrm{f}, (k_\mathrm{f}+1)T_\mathrm{f})$ based on the available storage space in the link $l_{r,s}$ at the end of the period. We have

$$q_{r,1}^{\max}(k_\mathrm{f}) = \frac{1}{T_\mathrm{f}} S_{r,s}(k_\mathrm{u}) + \sum_{k=k_\mathrm{u}}^{k_\mathrm{u}+N_\mathrm{fu}-1} \sum_{d_j \in D_s} m_{\mathrm{dep},r,s,d_j}(k).$$

   The effective outflow $q_{m,n_{\mathrm{last},m}}(k_\mathrm{f})$ of freeway link m between node p and off-ramp r is then given by (2.6).

3. Now the METANET model can be updated for simulation step $k_\mathrm{f}+1$.

4. We assume that the outflow of the off-ramp is distributed evenly over the period $[k_\mathrm{f} T_\mathrm{f}, (k_\mathrm{f}+1)T_\mathrm{f})$ such that

$$m_{\mathrm{arr},r,s}(k+\delta_{r,s}) = \frac{q_{m,n_{\mathrm{last},m}}(k_\mathrm{f}) T_\mathrm{f}}{N_\mathrm{fu}} \quad \text{for } k = k_\mathrm{u}, \ldots, k_\mathrm{u}+N_\mathrm{fu}-1.$$

   The corresponding urban queue lengths $x_{r,s,d_j}(k)$ for $k = k_\mathrm{u}+1, \ldots, k_\mathrm{u}+N_\mathrm{fu}$ can be updated using the urban traffic flow model.

In summary, the model for the off-ramp as well as the model for the on-ramp require a special order in which the computations are done. For simulating the whole network this means the computations should be done in the order shown in Figure 2.5. At the bottom each subfigure shows the urban time steps, at the top the freeway time steps. The first subfigure shows that with the flow at time step k_f the number of arriving vehicles in the urban network can be computed for time steps $k_\mathrm{u}+1,\ldots,k_\mathrm{u}+N_\mathrm{fu}$. Next, as shown in the second subfigure, the urban variables at time steps $k_\mathrm{u},\ldots,k_\mathrm{u}+N_\mathrm{fu}-1$ are used to adapt the flows at freeway time step k_f. Last, the freeway variables at time step $k_\mathrm{f}+1$ are computed with the variables at time step k_f.

2.3 Coordinated control for mixed networks

In the previous section we have developed a model that describes traffic networks that contain both urban roads and freeways. This model forms the basis for our model predictive control-based method. In this section we first give a general description of model predictive control (MPC). Next we formulate a traffic controller for mixed urban and freeway networks that is based on MPC.

We have selected MPC because it has the following features and advantages:

1. It can easily handle multi-input multi-output systems,
2. Only a few parameters have to be tuned,
3. It can handle constraints on inputs and outputs in a systematic way.

One of the first applications of MPC for traffic control is described in [58]. Other publications that deal with MPC or MPC-like approaches for traffic control are [49, 89, 129]. As described in [10, 64] MPC can be extended to coordinated control of freeway networks.

2.3.1 Model Predictive Control

Model predictive control (MPC) [25, 100] is a control method that has its origins in the process industry, where it is widely implemented due to its ability to deal effectively with increasing productivity demands, environmental regulations, and tighter product quality specifications. MPC is also suited for traffic control because it can easily handle changes in demands and in external conditions.

MPC approach

The goal of MPC is to minimize a cost function over a given prediction period. This cost function should give an indication for the performance of the system.

Figure 2.6 gives an overview of the operation of MPC. Assume that we are at time $t=k_\mathrm{c}T_\mathrm{c}=k_\mathrm{f}T_\mathrm{f}=k_\mathrm{u}T_\mathrm{u}$ where T_c is the controller time step. The current state of the system is measured, and fed into the controller. Now the current state and a prediction model are used to predict the behavior of the traffic during the period $[k_\mathrm{c}T_\mathrm{c},(k_\mathrm{c}+N_\mathrm{p})T_\mathrm{c})$, where N_p is called the prediction horizon. Note that in principle any traffic model can be used, but in this chapter we use the model described in Section 2.2 because it provides a good trade-off between accuracy and efficiency. With the obtained prediction the value $J(k_\mathrm{c})$ of the cost function for this period is computed.

The cost function should be minimized by selecting the optimal control signal sequence $c^*(k_\mathrm{c}),c^*(k_\mathrm{c}+1),\ldots,c^*(k_\mathrm{c}+N_\mathrm{p}-1)$. In order to reduce the number of optimization variables (and thus the computational complexity) usually a control horizon N_c (with $N_\mathrm{c}\leq N_\mathrm{p}$) is

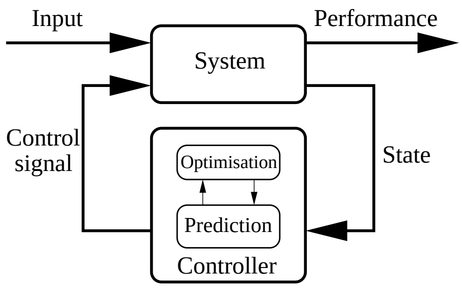


Figure 2.6: Working principle of MPC.

introduced and the control sequence is only allowed to vary over the period $[k_c T_c, (k_c + N_c)T_c)$ and is set constant afterwards, i.e. we have $c^*(k_c+k) = c^*(k_c+N_c-1)$ for $k = N_c, N_c + 1, \ldots, N_p - 1$.

From the optimal control signal sequence only the first sample $c^*(k_c)$ is applied to the real system. At the next control time step, a new optimization is performed with a prediction window that is shifted one control time step further. Of the resulting control signal again only the first sample is applied, and so on. This is called a receding horizon approach. This approach allows for updating the (estimated) system state from measurements every iteration, which introduces a feedback mechanism. In addition, it allows for adaptive control by regularly updating the model parameters using system identification.

Control signal, constraints, cost function, and prediction model

The MPC method requires defining the control signal c, the cost function J, and the constraints. Further, a suitable prediction model should be selected. Below we describe these elements for a general setting. In Section 2.3.2 they will be made specific for traffic control for mixed networks.

The control signal contains the settings for the control measures that are able to influence the system.

The constraints can contain upper and lower bounds on the control signal, but also linear or non-linear equality and inequality constraints on the control inputs and the states of the system. The constraints are used, e.g., to keep the system working within safety limits or to avoid unwanted situations.

The cost function J represents the performance of the network. Different performance criteria are possible. In practice, cost functions are often a combination of the different performance indicators:

$$J_{\text{total}}(k_c) = \upsilon_1 J_1(k_c) + \upsilon_2 J_2(k_c) + \upsilon_3 J_3(k_c) + \ \ldots$$

where the weights $\upsilon_i \geq 0$ of each term can be determined by the user of the controller.

MPC uses a model of the system to make predictions. Note that to be able to make a prediction of the traffic flows, the current state of the network should be known. This current state can be obtained via direct measurements or by using a state estimator, e.g., based on Kalman filtering [79].

MPC is an on-line control approach, and thus requires prediction models that give a balanced trade-off between accurate predictions and computational efforts. These models should be able to run several times faster than real-time, to ensure that the optimization algorithm can have results available within a specified amount of time (e.g., the sampling time of the system).

Optimization algorithms

At each control step MPC computes an optimal control sequence over a given prediction horizon. In general, this optimal control sequence is the solution of a non-linear, non-convex optimization problem in which the cost function is minimized subject to the model equations and the constraints. To solve this optimization problem different numerical optimization techniques can be applied, such as multi-start sequential quadratic programming (SQP) [126, Chapter 5] or pattern search [133] for real-valued problems, and genetic algorithms [44], tabu search [61], or simulated annealing [53] for mixed integer problems arising when discrete control measures are included.

2.3.2 MPC-based traffic control for mixed urban and freeway networks

The principle of MPC has been explained above. In this section we describe how MPC can be used to design a traffic controller for mixed urban-freeway networks. Note that the elements of an MPC controller (model, control signal, cost function, constraints, optimization algorithm) can be selected separately, and that the elements that we select here are just an example of a possible implementation.

The model requirements for MPC lead to the selection of a macroscopic traffic flow model to predict the behavior of the traffic. Macroscopic models are suited since the computation time is relatively low and does not depend on the number of vehicles in the network, and since they offer a good trade-off between accuracy and computational efforts. In Section 2.2 we have developed such a model, which we now include in our controller.

The control signal c can contain traffic signals for urban networks, presenting the green times and off-sets for each intersection. For freeway networks it can contain, e.g., ramp metering rates, variable speed limits, or lane closure settings. The MPC controller is often used as a higher-level controller. In this case the control signal contains control profiles and set-points for the local controllers. The low-level local controllers translate these profiles and set-points in the red/green signals for the real traffic control measures, as illustrated in Figure 2.7.

The traffic signals work as given in (2.7). The green time indicator signals are included in the global control signal via the cycle time T_{cyc}, the green offsets $o_{\text{green},o_i,s,d_j}$ (expressed as a percentage of the cycle time), and the green times $\pi_{\text{green},o_i,s,d_j}$ (also expressed as a percentage of the cycle time). This is done to prevent mixed-integer optimization problems, and to reduce the number of variables in the control signal. The percentages are translated into the binary red/green signal g_{o_i,s,d_j} as follows:
Suppose that we have to compute the control signals over the period $[t^0, t^{\text{end}}]$ with $t^0 = k^0 T_{\text{u}}$

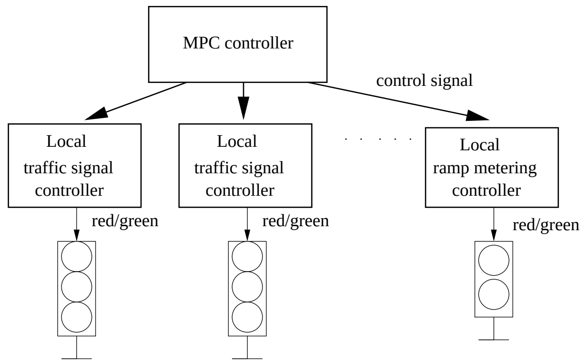


Figure 2.7: MPC-based control implemented with local controllers.

and $t^{\text{end}} = k^{\text{end}} T_{\text{u}}$. In this period the number of cycles is equal to $N_{\text{cyc}} = \text{ceil}\left(\frac{t^{\text{end}}-t^0}{T_{\text{cyc}}}\right)$. For each cycle $\ell = 0, 1, \ldots, N_{\text{cyc}}-1$ the vehicles coming from origin o_i and going to destination d_j at intersection s have green from time instant $t^0 + \ell T_{\text{cyc}} + o_{\text{green},o_i,s,d_j}(\ell) T_{\text{cyc}}$ up to time instant[6] $t^0 + \ell T_{\text{cyc}} + o_{\text{green},o_i,s,d_j}(\ell) T_{\text{cyc}} + \pi_{\text{green},o_i,s,d_j}(\ell) T_{\text{cyc}}$. So we have

$$g_{o_i,s,d_j}(k) = \begin{cases} 1 & \text{if } kT_{\text{u}} \in \bigcup_{\ell=1}^{N_{\text{cyc}}-1} [t^0 + \ell T_{\text{cyc}} + o_{\text{green},o_i,s,d_j}(\ell) T_{\text{cyc}}, \\ & \qquad t^0 + \ell T_{\text{cyc}} + o_{\text{green},o_i,s,d_j}(\ell) T_{\text{cyc}} + \pi_{\text{green},o_i,s,d_j}(\ell) T_{\text{cyc}}] \\ 0 & \text{otherwise} \end{cases}$$

for $k = k^0, k^0+1, \ldots, k^{\text{end}}$.
This implies that the actual urban control inputs computed by the MPC controller consist of the cycle times T_{cyc}, the offset percentages $o_{\text{green},o_i,s,d_j}$, and the green time percentages $\pi_{\text{green},o_i,s,d_j}$ for each traffic cycle in the given prediction period. The number of green times and offsets depends on the lengths of the cycles, which results in a variable length of the control signal. However, optimization algorithms that can handle inputs of variable size do not exist. Several options are available to avoid this problem: select a fixed cycle time, first optimize the cycle times for fixed offsets and green times and then optimize the offsets and green times with the obtained cycle time, or perform a bi-level optimization of cycle times, offsets, and green times.

Ramp metering installations limit the flows that leave the on-ramps. The ramp metering rates will be computed every controller time step T_{c}. For ease of notation we define the set of freeway time steps k_{f} that correspond to a given interval $[k_{\text{c}}^a, k_{\text{c}}^b]$ of controller time steps as follows:

$$\mathcal{K}_{\text{f}}(k_{\text{c}}^a, k_{\text{c}}^b) = \left[k_{\text{c}}^a \frac{T_{\text{c}}}{T_{\text{f}}}, k_{\text{c}}^a \frac{T_{\text{c}}+1}{T_{\text{f}}}, \ldots, k_{\text{c}}^b \frac{T_{\text{c}}}{T_{\text{f}}} - 1\right] .$$

The ramp metering rates in freeway time steps are then given by:

$$r^{\text{ramp}}(k_{\text{f}}) = r_{\text{c}}^{\text{ramp}}(k_{\text{c}}) \;\; \forall \;\; k_{\text{f}} \in \mathcal{K}_{\text{f}}(k_{\text{c}}, k_{\text{c}}+N_{\text{p}}) \tag{2.9}$$

[6]Note that in fact time instants beyond t^{end} do not have to be considered.

where $r^{\text{ramp}}(k_\text{f})$ is the ramp metering rate in freeway time steps, and $r_\text{c}^{\text{ramp}}(k_\text{c})$ is the metering rate in control time steps.

The flow that leaves the on-ramp is then determined by:

$$q_{r,p}^{\text{onramp,metering}}(k_\text{f}) = \min(q_{r,p}^{\text{onramp,no metering}}(k_\text{f}), r^{\text{ramp}}(k_\text{f})Q_{\text{cap},r})$$

where $q_{r,p}^{\text{onramp,no metering}}(k_\text{f})$ is the flow on the on-ramp when no metering is applied (cf. (2.5)), and $Q_{\text{cap},r}$ is the capacity of the on-ramp.

The last part of the control signal contains the freeway speed limits, which are also determined via zero order interpolation, as in (2.9). Speed limits influence the speed of the drivers by changing the speed that they try to approximate [64]:

$$v_{m,i}^{\text{desired,limits}}(k_\text{f}) = \min(v_{m,i}^{\text{desired,no limits}}(k_\text{f}), (1+\alpha_m)v_{m,i}^{\text{limit}}(k_\text{f})) \tag{2.10}$$

where $v_{m,i}^{\text{desired,no limits}}(k_\text{f})$ is the desired speed of the drivers when there are no speed limits applied (cf. (2.3)), $v_{m,i}^{\text{limit}}(k_\text{f})$ is the value of the applied speed limits, and α_m is a parameter which represents the fact that drivers will freely interpret and adhere to the speed limits. When enforcement is used α_m will typically be around -0.1, but without enforcement drivers will tend to drive too fast, and α_m can be around 0.1.

The subsequent values of the ramp metering rates and the variable speed limits over the prediction period form the freeway part of the control signal.

Furthermore, we can impose constraints for the controller. For traffic control such constraints can consist of, e.g., maximum queue lengths at intersections, at on-ramps or at off-ramps, minimum and maximum green times, minimum and maximum speed limits, maximum flows on roads, constraints that the traffic signal plans should be conflict-free, etc. These constraints could be prescribed by regulations, or they could express a policy selected by the traffic management authorities.

The cost function can be determined by the traffic management authorities, to represent their traffic management policies. The cost function could contain the total time that the vehicles spend in the network, the average queue length, the number of stops, the total delay, the throughput, vehicle loss hours, variation in the travel times, the total fuel consumption, the emission levels, the noise production, etc., or a combination of them. The cost functions for the urban and freeway parts of the network are often computed separately, to allow a trade-off between the two:

$$J_{\text{total}}(k_\text{c}) = \upsilon_\text{f} J_{\text{freeway}}(k_\text{c}) + \upsilon_\text{u} J_{\text{urban}}(k_\text{c})$$

where υ_f and υ_u are a weight factors to determine the relative influence of the freeway and urban traffic.

A cost function that is often used in literature (see, e.g., [10, 64, 86, 89]) is the total time spent (TTS) by all vehicles in the network. We will also use this objective function for our case study in Section 2.4. Therefore, we will now expand somewhat on this specific objective function. To compute the TTS for the urban part of the network the number of vehicles in each urban link $n_{\text{vehicles},l_{\sigma,s}}$ is required:

$$n_{\text{vehicles},l_{\sigma,s}}(k_\text{u}) = L_{\sigma,s} - S_{\sigma,s}(k_\text{u})$$

where $L_{\sigma,s}$ is the maximum number of vehicles that the link can contain. Using this equation the number of vehicles for all urban links, on-ramps, and off-ramps can be computed.

The TTS will be computed over the period $[k_c^0 T_c, (k_c^0 + N_p)T_c)$ when we are at time $t = k_c^0 T_c$. Now define k_u^0 and k_f^0 such that $k_c^0 T_c = k_u^0 T_u = k_f^0 T_f$ and $k_u^{0,\text{end}}$ and $k_f^{0,\text{end}}$ such that $(k_c^0 + N_p)T_c = (k_u^{0,\text{end}} + 1)T_u = (k_f^{0,\text{end}} + 1)T_f$. The total time spent in the urban part of the network during the period $[k_c^0 T_c, (k_c^0 + N_p)T_c)$ is then given by:

$$\text{TTS}_\text{urban}(k_c^0) = T_u \sum_{k=k_u^0}^{k_u^{0,\text{end}}} \Big(\sum_{l_{o_i,s} \in I} n_{\text{vehicles},l_{o_i,s}}(k) + \sum_{l_{s,r} \in R_\text{on}} n_{\text{vehicles},l_{s,r}}(k) + \sum_{o \in O_\text{urban}} n_{\text{vehicles},o}(k) \\ \sum_{l_{r,s} \in R_\text{off}} n_{\text{vehicles},l_{r,s}}(k) \Big)$$

where $\text{TTS}_\text{urban}(k_c^0)$ denotes the total time spent in the urban part of the network during the period $[k_c^0 T_c, (k_c^0 + N_p)T_c)$, I the set of all urban links, O_urban the set of all urban origins o, R_on the set of links urban $l_{r,s}$ connected to the on-ramps, and R_off the set of urban links $l_{s,r}$ connected to the off-ramps.

The TTS in the freeway part of the network is computed using the density on the segments:

$$\text{TTS}_\text{freeway}(k_c^0) = \sum_{k=k_f^0}^{k_f^{0,\text{end}}} \sum_{m \in M} \Big(L_m n_m \sum_{i \in I_m} \rho_{m,i}(k) + \sum_{o \in O_\text{freeway}} w_o(k) \Big)$$

where $\text{TTS}_\text{freeway}(k_c^0)$ is the total time spent in the freeway part of the network during the period $[k_c^0 T_c, (k_c^0 + N_p)T_c)$, M the set of all freeway links m, I_m the set of all segments i in link m, and O_freeway the set of all freeway origins.

The total cost function is given by the weighted sum of the urban and freeway total time spent:

$$\text{TTS}(k_c^0) = \upsilon_f \text{TTS}_\text{freeway}(k_c^0) + \upsilon_u \text{TTS}_\text{urban}(k_c^0) \quad .$$

2.4 Case study

To illustrate the performance of the MPC method we will present a simple case study. The case study concentrates on urban control but in a network that also contains a freeway. We have done this to be able to make a comparison with existing dynamic control systems, which have mainly been developed for urban control measures.

2.4.1 Set-up of the case study

For the case study a simple network is used, as shown in Figure 2.8. The network consists of two freeways (freeway 1 and 2) each with two lanes, two on-ramps, and two off-ramps (ramp 1 to 4). Furthermore, there are two urban intersections (A and C), which are connected to the freeway and to each other. Between these intersections and the freeways there are some crossing roads (B, D, and E), where there is only crossing traffic that does not turn into other directions, e.g., pedestrian traffic, bicycles, etc. We have selected this network because it contains most of the essential elements from mixed networks. There are freeways with

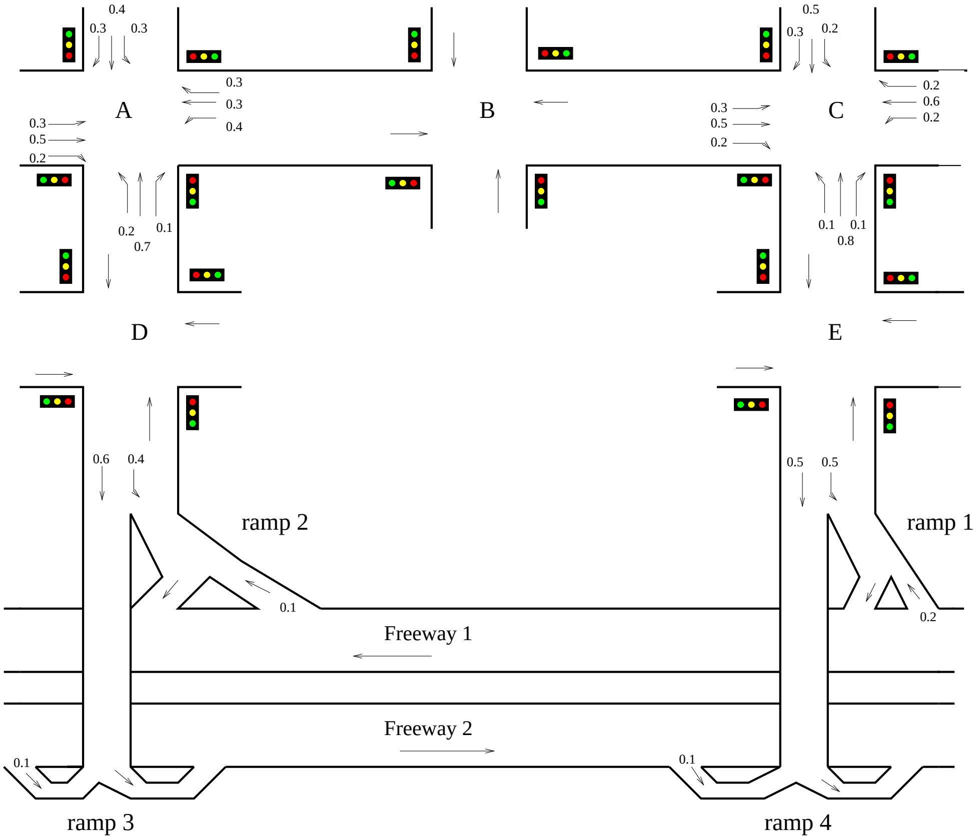


Figure 2.8: Network used in the case study.

on-ramps and off-ramps and controlled intersections not too far away from the freeways resulting in a strong relation between the traffic on the two types of roads. The network is small enough to use intuition to analyze and interpret the results, but large enough to make the relevant effects visible.

The performance of the control systems will be shown for different traffic scenarios. Three of them are scenarios with different traffic situations, while the fourth is a control-related scenario. We have selected these scenarios because they clearly show the influence of the urban traffic on the freeway traffic and vice versa, because this influence occurs frequently, and because some properties of the control methods will become clearly visible. The ‘basic’ scenario has a demand of 3600 veh/h for freeway origins and 1000 veh/h for urban origins, and turning rates as shown in Figure 2.8. Each of the scenarios is a variation on this ‘basic’ scenario, with one variable or parameter changed or a constraint added. The total simulated time is 30 minutes. These are the four scenarios:

Scenario 1: congestion on the freeway A traffic jam exists at the downstream end of freeway 1. This congestion grows into the upstream direction and blocks the on-ramps, causing a spill-back leading to urban queues. The congestion is created by imposing a downstream density of 65 veh/km/lane for the last segment of the freeway.

Scenario 2: blockage of an urban intersection On intersection D an incident has occurred, and the whole intersection is blocked. The queues spill back to neighboring intersections, and also block the off-ramps of the freeways. This incident is simulated by setting the saturation flow of all links leaving the intersection to 0 veh/h.

Scenario 3: rush hour In this scenario the demand at the origins becomes larger during a short period. We have selected a flow of 500 veh/h with a peak of 2000 veh/h for the urban origins, and a flow of 2000 veh/h with a peak of 4000 veh/h for freeway origins. The duration of the peak is 10 minutes.

Scenario 4: maximum queue length Here, the queue on the link from intersection A toward intersection B may not become longer than 20 vehicles. This can be a management policy, e.g., when the link is in a residential area.

2.4.2 Simulation set-up

For all control systems the implementation of the simulations and the controller is completely done in the mathematical computation environment Matlab. We use the model described in Section 2.2 both as real-world model and as prediction model. With this set-up we can give a proof-of-concept of the developed control method, without introducing unnecessary side-effects.

In our case study the MPC optimization problem is a non-convex, non-linear problem with real-valued optimization variables. To solve this problem we have selected multi-start SQP [126] as optimization algorithm. This algorithm is implemented in the `fmincon` function of the Matlab optimization toolbox [154].

As cost function we select the total time spent (TTS). The parameters of the METANET model are selected according to [87]: $v_{\text{free},m} = 106$ km/h, $\rho_{\text{crit},m} = 33.5$ veh/km/lane, $\rho_{\text{max},m} = 180$ veh/km/lane, $Q_{\text{cap},m} = 4000$ veh/h, $\tau = 18$ s, $\nu = 65$ km^2/h, $\kappa = 40$ veh/km/lane, and $a_m = 1.867$. The parameters of the urban model are $Q_{\text{cap},o,s,d} = 1000$ veh/h, $L_{\text{av,veh}} = 6$ m, and $v_{\text{av},l_{s,d}} = 50$ km/h.

We have selected the following time steps: $T_\text{c} = 120$ s, $T_\text{f} = 10$ s, and $T_\text{u} = 1$ s. A small value is selected for the urban time step to obtain detailed information. The freeway time step of 10 s forms a trade-off between computational effort and accuracy.

There are three parameters that can be tuned for the MPC controller. We have selected $N_\text{p} = 8$ and $N_\text{c} = 3$ as horizons, and $\alpha = 1$ as trade-off between urban and freeway performance in the cost function.

2.4.3 Alternative control methods

Many dynamic traffic control systems have already been implemented in the real world. Some of these systems are SCATS [185], Toptrac [6], SCOOT [140], UTOPIA/SPOT [129], MOTION [24], and IN-TUC [48]. Here we will use approximations of SCOOT and UTOPIA/SPOT to make a comparison between the developed MPC control method and some existing systems. Both SCOOT and UTOPIA/SPOT target the urban traffic, and they optimize intersections independently of the neighboring freeway. We have selected these methods because they are good representatives of this kind of dynamic traffic control systems. Note however that these systems are commercial systems, meaning that real specifications are not publicly available. This means that we can only approximate their functioning as follows:

System 1: a SCOOT-like system SCOOT [140] has a controller on each intersection. These controllers estimate the arriving traffic flows using a cyclic flow profile, that is updated via measurements taken at the beginning of each link. Every control time step the cycle time is updated. This is done according to the ratio between the current queue length and the maximum allowed queue length. When more than 90% of the maximum queue length is reached, the cycle time is increased. The time differences between the beginning of the green times of different intersections are called the offsets. At the beginning of each cycle the offsets are optimized locally by adapting them to the expected demands. A prediction of the traffic flows for the next cycle is used to determine the optimal values for each intersection separately, using predictions obtained from neighboring intersections during the previous time step. The green times are updated every time step (1 s). A prediction of the traffic during the next cycle is made to determine whether it is useful to increase or decrease the green times with 4 s. The model used for the predictions is a simple queue length model. It describes the number of vehicles arriving at the beginning of the link, the delay due to the travel time on the link, the length of the queue, and the number of vehicles leaving the link.

System 2: an UTOPIA/SPOT-like system UTOPIA/SPOT [104, 129] has been developed in Turin, Italy. It is a hierarchical system with a local controller at each intersection, and a central controller. The central controller computes an optimal control signal consisting of setpoints for the local controllers, using a prediction of the traffic in the whole urban network over a period of 15 minutes. These setpoints are sent to the local controllers. Each of these local controllers communicates with its neighbors to obtain their measurements and expected control scheme. With this information the local controllers compute a locally optimal green times and offsets, using predictions of the traffic on the local intersection during the next cycle, including the arriving traffic. In the cost function used by the local controllers a penalty is added for deviations from the signal computed by the central controller. In this way the central controller can influence the local controllers. A queue length model is used to obtain the predictions of the traffic state.

In both systems constraints like maximum queue lengths are introduced by adding a penalty term to the cost function. This penalty term must become relatively large when the maximum queue length is reached. This results in a very high value of the cost when the maximum queue length is violated. While the purpose of the control is to minimize the cost function, a trade-off will have to be made between minimizing the original cost and violating the queue length constraint.

2.4.4 Qualitative comparison

The main difference between the MPC-based system proposed in this chapter and the existing systems is that the new system takes the influences and interactions between the urban and freeway parts of the network into account. By simulating the effect of one measure on both kinds of roads, control settings can be found that provide a trade-off between improving traffic conditions on the freeway and delaying traffic on the urban roads and vice versa.

Furthermore, the MPC-based system we have developed can handle hard constraints on both the control signals and the states of the system. All systems can handle constraints that are directly linked to the control signals, e.g., maximum and minimum green times and cycle time constraints. But the MPC-based system can also handle more indirect constraints such as maximum queue lengths, maximum delays, etc. These constraints are included as hard constraints in the MPC optimization problem, which is subsequently solved using a constrained optimization algorithm (e.g., SQP). In the other systems such a constraint is implemented by adding a penalty term that penalizes the constraint violation to the performance function. This can lead to either satisfying the constraints with a degraded performance, or violating the constraints and obtaining a better performance. Which of the two occurs depends on the weight that is given to the penalty term. Figure 2.9(a) shows a queue on the link from A to B when MPC-based control is applied. Figure 2.9(b) shows the same queue, but now with MPC-based control with a queue length constraint of 12 vehicles. Whereas in Figure 2.9(a) the queue repeatedly exceeds 12 vehicles, in Figure 2.9(b) the queue stays around 10 vehicles. This is due to the fact that the controller predicts that the queue will exceed the limit during the prediction horizon, which causes the controller to change the value of the control signals. When the first step of this control signal is then applied on the real system, the queue stays lower than the value of the constraint.

The three control methods are also characterized by different communication requirements. **System 1** is based on local controllers, each with their own detectors and control algorithms. **System 2** uses different levels: local controllers that communicate with their neighbors, and a centralized control computer that communicates with each local controller, mainly sending set-points for the local control algorithms. The MPC method is in principle a centralized method in which the control algorithm runs on a central computer, and only the results of the optimization are communicated toward the low-level controllers. In this way an optimum for the total network is found, possibly at the cost of large computation times in the case of large networks (in Section 2.4.6 we will sketch some ways to address this issue).

2.4.5 Quantitative comparison

We have applied the three different control methods to the case study network. The results are shown in Table 2.1. Each traffic scenario is simulated with each control method. The table shows the total time spent for the freeway part of the network, for the urban part, and for the whole network. The last column of the table shows the improvement of the MPC method compared to **System 1** (first number) and to **System 2** (second number). This makes it possible to determine in which part of the network the largest improvements are obtained. For the fourth scenario the largest attained queue length is also shown.

The first two scenarios show that the MPC method can improve the performance for the urban as well as for the freeway part of the network when a problem arises in one of the two. The immediate negative effects of such a problem are reduced just as the negative influence on the rest of the network.

*Table 2.1: Results of the case study: total time spent for the freeway part of the network, for the urban part, and for the total network; and also the improvement of the MPC-based method compared to **System 1** and **System 2** respectively.*

Scenario 1: congestion on the freeway

	System 1	**System 2**	**MPC**	**improvement**
freeway	595.4	565.1	563.9	5.3 / 0.3%
urban	313.6	335.7	305.7	3.0 / 9.0%
total	909.0	900.8	869.6	4.4 / 3.5%

Scenario 2: blockage of an urban intersection

	System 1	**System 2**	**MPC**	**improvement**
freeway	498.0	526.2	495.0	0.7 / 6.0%
urban	665.9	672.3	620.3	6.9 / 7.8%
total	1163.9	1198.5	1115.3	4.2 / 7.0%

Scenario 3: rush hour

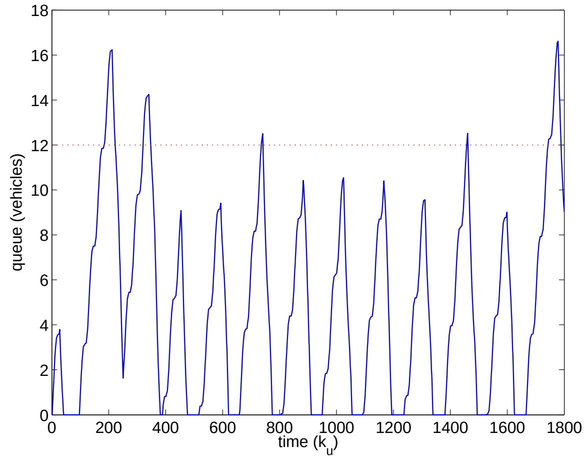
	Sytem 1	**System 2**	**MPC**	**improvement**
freeway	244.6	280.1	253.3	-3.5 / 9.6%
urban	409.0	383.5	386.8	5.5 / -1.6%
total	653.6	663.6	640.1	2.1 / 3.5%

Scenario 4: maximum queue length of 20 vehicles with large weight

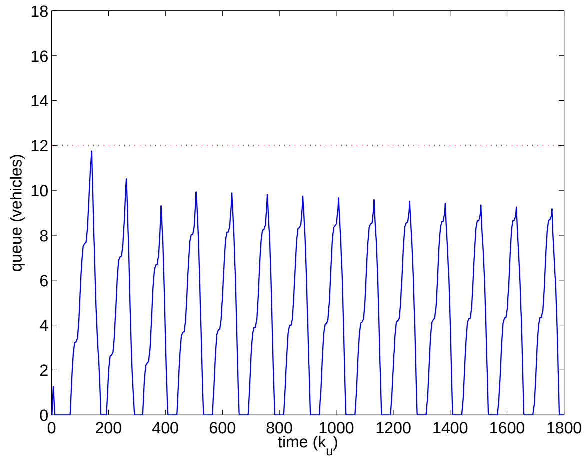
	System 1	**System 2**	**MPC**	**improvement**
freeway	367.2	510.3	373.9	-1.8 / 26.8%
urban	309.7	435.4	264.4	15.7 / 39.3%
max. queue	19	19	21	
total	676.9	945.7	638.3	6.8 / 32.6%

Scenario 4: maximum queue length of 20 vehicles with small weight

	System 1	**System 2**	**MPC**	**improvement**
freeway	367.1	428.1	373.9	-1.8 / 13.7%
urban	303.0	360.5	264.5	13.8 / 26.7%
max. queue	93	43	21	
total	670.1	788.6	638.3	5.8 / 19.1%



(a) Queue length at the link from A to B, for MPC without queue length constraint.



(b) Queue length at the link from A to B, for MPC with a queue length constraint of 12 vehicles.

Figure 2.9: The effect of a queue length constraint.

The third scenario shows that the MPC method can control the traffic slightly better when a large peak in the demand occurs. In this scenario the trade-off between the freeway and urban parts of the network can clearly be seen. A reduction of the performance on the urban network can lead to an improvement of the performance on the freeway network, and vice versa. This can be used to obtain a better performance for the total network.

The maximum queue length constraint is implemented in **System 1** and **System 2** by adding an extra penalty term in the cost function. This term has a relative weight that allows a trade-off between the performance of the network and the importance of the maximum queue length constraints. When the weight is high the queue length constraint is satisfied but the performance is low, as shown in the first simulations done for the fourth scenario. In the second set of simulations the weighting term for the queue constraint is low, resulting in a better performance, but now the maximum queue length is exceeded. The values for MPC are the same for both simulation sets because the queue length constraint is implemented as a hard constraint for the optimization algorithm[7].

2.4.6 Discussion

Although the MPC-based method gives good results, some parts of it have to be investigated more extensively.

The most important problem at the moment is the required computational effort. The run time for the MPC-based method is larger than for the other methods. This is due to the use of one central computer and to the fact that a larger network is optimized at once. This can be solved by, e.g., using faster computers, by using the method in a distributed setting, or by using better special, dedicated solvers implemented in object code[8].

The optimization technique also forms an important factor in relation to the computation time and the computed optimal control signal. Different optimization algorithms can have different run times, and yield different solutions. To select the best algorithm extensive simulations should be done for a wide range of set-ups and scenarios to compare the various algorithms.

When hard constraints are implemented, it is possible that the optimization problem becomes infeasible. When this occurs, one or more constraints have to be relaxed (see [25, 100] for more details). This can in reality mean that the constraints are violated for a short period.

The effects of selecting different cost functions should also be investigated, just as the influence of the weighting parameters υ_{f} and υ_{u}, which determine the trade-off between the urban and the freeway costs.

2.5 Conclusion

Congestion on urban roads and congestion on freeways cannot be seen as separate problems: the traffic on urban roads influences the traffic on freeways and vice versa. As a result, control measures taken on one of the two types of roads have influence on both types of roads. We have developed a control approach that takes this influence into account when the control signals are determined. The approach is suitable for integrated control, and yields a balanced trade-off between the urban and the freeway parts of the network.

We have first developed a model that describes the evolution of traffic flows on mixed urban-freeway networks. For the freeway part the METANET model is used, and for the urban roads a queue length model based on Kashani's model is developed. We have made the connection between the urban and freeway parts of the network by modeling on-ramps and off-ramps.

The mixed network model has been used to develop a coordinated control method using MPC. In MPC the evolution of the traffic flow is predicted over a certain period, and this

[7]The MPC-based method violates the constraint with 1 vehicle at the start of the simulation. This is due to infeasibility problems during the optimization, related to the initial state of the network at the start of the simulation. This issue can be solved by increasing the horizons N_{p} and N_{c}.

[8]The current simulations are programmed in Matlab, which is basically an interpreted language.

prediction is then used to optimize the control settings, using numerical algorithms. MPC uses a receding horizon approach: only the first step of the optimized control settings is applied, and then the procedure is started all over again. This makes that the controlled system can also cope with changes in the traffic demand, and with model mismatches.

We have performed a case study to compare the MPC method with existing control methods. In particular, for the comparison we have selected methods that are an approximation of SCOOT and UTOPIA/SPOT. Different traffic scenarios have been simulated and the result of the three systems have been compared qualitatively and quantitatively. The MPC method performs between 2% and 7% better than the other two systems, and can guarantee bounds on the queue lengths without a large decrease in performance.

The results of the simulation are promising: they can be seen as a proof-of-concept for the proposed approach, they show its potential benefits, and encourage further research. This research could include the following steps. First, additional case studies, with several different traffic scenarios and set-ups including larger networks should be performed. Next, case studies should be done where different models are used to model the 'real-world' traffic flows (for the prediction model we would keep on using the macroscopic model proposed in this chapter). Then, the efficiency of the algorithm should be improved by, e.g., selecting/developing other optimization algorithms, or adapting the model as in [97]. Attention should be payed to the robustness and sensitivity of the control method. Last, a real-life test should be done. Other topics that should be investigated are the validation and calibration of the model. Furthermore, for the simulation of larger networks, it is useful to investigate MPC for distributed control in which different adjacent network regions are defined and optimized separately (but with some coordination to avoid negative influences of the control actions of one region on the other regions).

2.A List of symbols

Freeway model

T_f	freeway time step (h)
k_f	freeway time step counter
$v_{m,i}(k_\mathrm{f})$	space mean speed on segment i of freeway link m at freeway time step k_f (km/h)
$\rho_{m,i}(k_\mathrm{f})$	average density on segment i of freeway link m at freeway time step k_f (veh/km/lane)
$q_{m,i}(k_\mathrm{f})$	average outflow of segment i of freeway link m at freeway time step k_f (veh/h)
n_m	number of lanes of freeway link m
L_m	length of the segments of freeway link m
$V(\rho_{m,i}(k_\mathrm{f}))$	desired speed corresponding to the density at segment i of freeway link m during freeway time step k_f (km/h)
$v_{\mathrm{free},m}$	free flow speed of freeway link m (km/h)

$\rho_{\mathrm{crit},m}$	critical density of freeway link m (veh/km/lane)
$w_o(k_\mathrm{f})$	queue length at origin o at freeway time step k_f (veh)
$d_o(k_\mathrm{f})$	demand at origin o at freeway time step k_f (veh/h)
$Q_{\mathrm{cap},m}$	capacity of freeway link m (veh/h)
$\rho_{\mathrm{max},m}$	maximum density at freeway link m (veh/km/lane)
$Q_{\mathrm{tot},p}(k_\mathrm{f})$	total flow entering node p at freeway time step k_f (veh/h)
I_p	set of all freeway links entering node p
$n_{\mathrm{last},m}$	index of the last segment of freeway link m
$\beta_{p,m}$	turning rate from node p to leaving freeway link m
O_p	set of leaving links of node p
$r^{\mathrm{ramp}}(k_\mathrm{f})$	ramp metering rate at freeway time step k_f
$q_{m,n_{\mathrm{last},m}}(k_\mathrm{f})$	flow that can enter the off-ramp r connected to freeway link m at freeway time step k_f (veh/h)

Urban model

T_u	urban time step (h)
k_u	urban time step counter
$l_{o_i,s}$	link connecting origin o_i with intersection s
$m_{\mathrm{arr},s,d_j}(k_\mathrm{u})$	number of vehicles arriving at the end of the queue in link l_{s,d_j} at urban time step k_u (veh)
$m_{\mathrm{dep},s,d_j}(k_\mathrm{u})$	number of vehicles departing from intersection s towards link l_{s,d_j} at urban time step k_u (veh)
$m_{\mathrm{arr},o_i,s,d_j}(k_\mathrm{u})$	number of vehicles from origin o_i going to destination d_j arriving at the end of the partial queue at intersection s at urban time step k_u (veh)
$m_{\mathrm{dep},o_i,s,d_j}(k_\mathrm{u})$	number of vehicles leaving queue at urban time step k_u (veh)
D_s	set of destinations connected to intersection s
$x_{o_i,s,d_j}(k_\mathrm{u})$	queue length consisting of vehicles coming from origin o_i going to destination d_j at intersection s at urban time step k_u (veh)
$g_{o_i,s,d_j}(k_\mathrm{u})$	binary signal that is 1 when the direction from origin o_i to destination d_j at intersection s has green at urban time step k_u
$S_{s,d_j}(k_\mathrm{u})$	free space in the link l_{s,d_j} connecting intersection s and destination d_j at urban time step k_u (veh)
$Q_{\mathrm{cap},o_i,s,d_j}$	the saturation flow at intersection s for traffic from origin o_i with destination d_j
$L_{l_{o_i,s}}$	length of link $l_{o_i,s}$ from origin o_i to intersection s
$\delta_{s,d_j}(k_\mathrm{u})$	delay experienced in the link l_{s,d_j} by a vehicle that enters at urban time step k_u (expressed as a multiple of T_u)
$L_{\mathrm{av,veh}}$	average length of a vehicle (m)
$\beta_{o_i,s,d_j}(k_\mathrm{u})$	percentage of the traffic on link $l_{o_i,s}$ with destination d_j at urban time step k_u

Controller

T_c	control time step (h)
k_c	control time step counter
N_p	prediction horizon
N_c	control horizon
c	general control signal
$J(k_\mathrm{c})$	generalized total costs for the period $[k_\mathrm{c}, k_\mathrm{c}+N_\mathrm{p})$
I	set of all urban links
$n_{\mathrm{vehicles},l_{o_i,s}}(k_\mathrm{u})$	number of vehicles in link $l_{o_i,s}$ at urban time step k_u (veh)
$\mathrm{TTS}(k_\mathrm{c})$	total time spent in the network at control time step k_c (veh·h)
$\mathrm{TTS}_\mathrm{urban}(k_\mathrm{c})$	total time spent in the urban part of the network at control time step k_c (veh·h)
$\mathrm{TTS}_\mathrm{freeway}(k_\mathrm{c})$	total time spent in the freeway part of the network at control time step k_c (veh·h)
T_cyc	global cycle time
N_cyc	number of cycles
$\pi_{\mathrm{green},o_i,s,d_j}(k_\mathrm{u})$	green time for the direction from origin o_i to destination d_j at intersection s at urban time step k_u (percentage of cycle time)
$o_{\mathrm{green},o_i,s,d_j}(k_\mathrm{u})$	offset of the green for the direction from origin o_i to destination d_j at intersection s at urban time step (percentage of cycle time)

Chapter 3

Traffic control strategies based on different assignment models

In this chapter we develop three coordinated control strategies that take re-routing effects into account. All strategies use a model-based predictive control approach to determine optimal settings for traffic control measures. The prediction model used in the controllers consists of two parts; a traffic flow model, and a traffic assignment model. For the traffic flow model the controllers use METANET [106]. The assignment models differ for the controllers developed in this chapter.

The first strategy that we develop considers the within-day route choice process, and uses an explicit equilibrium-based dynamic traffic assignment model based on the Method of Successive Averages (MSA) to describe this process.

The second and third strategies use a route-choice-based traffic assignment model, which describes the within-day route choice as well as the day-to-day route choice. The second strategy considers only the within-day part of the route choice model, while the third strategy includes the within-day and day-to-day route choice.

The strategies use the route choice models for different purposes. The first an second control strategies only take the effect of the route choice process into account when they determine the settings for the control variables, while the third strategy actively influences the route choice via information provided on dynamic route information panels.

The performance of the control strategies is illustrated with several small case studies, in which the methods that are developed are compared with existing control methods.

3.1 Introduction

When there are different routes from origin to destination in a network, the traffic flows divide themselves over these routes. This process is called dynamic traffic assignment. When a control measure is present in the network, its control actions influence the traffic flows, and thus implicitly influence the traffic assignment [148]. This change in the traffic assignment may require a change in the control actions. However, the effect of control actions on the traffic assignment is usually not included in the current traffic control frameworks.

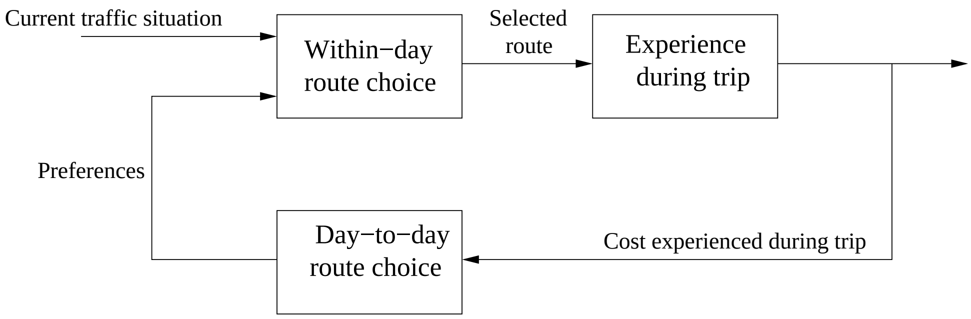


Figure 3.1: Overview of the route choice process.

In this chapter we focus on including this dynamic traffic assignment process in the control method[1].

A traffic assignment is the result of the route choices of the drivers. The route choice process has two different time scales, see Figure 3.1. There is the day-to-day change in route choice, and the within-day route choice. The day-to-day route choice depends on preferences of the drivers with respect to, e.g., nice surroundings, absence of traffic signals, wider lanes, and experiences encountered during the last day, such as queue lengths, travel times, delays, etc. The day-to-day preferences are updated at the end of each day.

The within-day route choice is based on the situation in the network at the moment the driver has to make this route choice. This situation can be qualified using, e.g., the instantaneous travel time, instantaneous queue lengths, flows, densities, or speeds at the current moment. When the process converges, the day-to-day and within-day route choice together lead to a dynamic equilibrium traffic assignment.

The first paper that describes the effects concerning route choice is [182]. Other early works are [54, 144], which describe the reasoning of the drivers while selecting their route, and which claim that the route choice will lead to a so called static user equilibrium traffic assignment. This is the assignment in which all routes have the same costs for the drivers. Several authors have developed methods to compute this static equilibrium assignment [50, 102]. But with varying demands the equilibrium assignment will also vary, and so dynamic traffic assignment algorithms have been developed, see, e.g., [15, 56].

In this chapter we will consider two types of assignment models: equilibrium-based models that assume that the traffic flows are always in an (dynamic) equilibrium, and en-route route-choice-based models that describe the route choice behavior of drivers at individual intersections. Equilibrium-based models are relatively easy to validate since real-life travel time data can be gathered with on-line measurements, but require large computation times since they lead to bi-level optimization problems when they are used within an optimal control setting, as described in [10, 148]. En-route route-choice-based models require less computation time than equilibrium-based models, but they are more difficult to validate due to the lack of route choice data.

All traffic assignment models are using the costs for different routes to determine the

[1] The actions of the control measures cannot only influence the route choice but also the departure times of the drivers. Including the departure time choice in a control method can be done in a similar way as including the route choice: by embedding a departure time model in the controller.

assignment. The cost that the drivers experience is for a large part based on the travel times on the routes [18, 42, 142]. These travel times are influenced by the control measures, and thus the control measures influence the traffic assignment. This means that current traffic control methods can be improved by taking into account that the change in the route choice generated by their control actions requires a change in the control strategy to maintain optimal performance. Another reason to include route choice behavior in the control strategy is that some control measures, e.g., DRIPs, are designed to explicitly influence the traffic assignment. The effects of these measures on the route choice of drivers are described in, e.g., [18, 46, 103, 147].

In this chapter we develop three control strategies that include re-routing effects into the controller. The strategies that we propose are all based on model predictive control (MPC) [25, 100], which has already been applied for different freeway networks [10, 22, 65], and which can handle hard constraints. MPC requires a model that predicts the evolution of the traffic flows on the network. For the control strategies that we develop we will use the macroscopic traffic flow model, METANET [106], combined with a traffic assignment model.

The main differences between the control strategies that will be developed are the timing of the route choice processes that is considered, the assignment algorithm that is selected, and the function of the assignment model within the controller. The first strategy considers only within-day route choice, uses an equilibrium-based dynamic traffic assignment (DTA) algorithm [11], and uses the DTA algorithm to anticipate on changes in the route choice due to the control actions. The equilibrium-based DTA algorithm determines the traffic assignment given the present values for the control input in such a way that an equilibrium appears in which the costs for all routes between a specific origin and destination are equal. This is done according to the following procedure. First, the METANET model is used to predict the evolution of the traffic flows with these inputs. Based on these predictions, the travel times for the routes in the network are obtained. Then the method of successive averages (MSA) [134] is used to determine the corresponding traffic assignment. This new assignment is used in the optimization process to select the best settings for the coordinated traffic control measures. We illustrate the first strategy with a case study involving a small network with two routes and one on-ramp, using ramp metering as control measure.

The second control strategy uses an en-route route-choice-based assignment model that implicitly determines the traffic assignment. The model describes within-day as well as day-to-day route choice, but the second control strategy uses only the within-day route choice. The model does not assume an equilibrium assignment, but predicts the route choice of drivers based on previously experienced travel times. The control strategy uses the DTA algorithm to anticipate on changes in the route choice. The performance of this control strategy is illustrated with a case study involving anticipative on-ramp metering. Also, a short investigation of off-ramp metering using this control strategy is performed.

The third control strategy considers within-day and day-to-day route choice using the en-route route-choice-based assignment model. The strategy uses the DTA model to to actively steer the route choice, which is also done in [47, 81, 147]. Further, the third control method integrates existing control measures, e.g., ramp metering installations and variable speed limits with dynamic route information panels (DRIPs) to be able to influence the route choice. A case study on a network with two routes is performed, to illustrate the possibilities of the control method that is developed.

The remainder of this chapter is organized as follows. Since we use ramp metering to illustrate the different control strategies, we introduce ramp metering and available control methods for ramp metering in Section 3.2. This section also contains a short description of the Metanet model, which is used for all MPC-based control strategies in this paper. Then, Section 3.3 describes an anticipative ramp metering strategy using equilibrium-based dynamic traffic assignment, while Section 3.4 introduces anticipative ramp metering based on route-choice-based DTA. Next, Section 3.5 presents the integrated control method using DRIPs in combination with variable speed limits. At last, in Section 3.6 conclusions are drawn.

3.2 Ramp metering

In this section we first describe the general principles of on-ramp metering and off-ramp metering, and next we present three different control methods that can be used to determine the ramp metering rate: fixed-time control, ALINEA, and model predictive control (MPC). The first two methods are the most well known methods that are used in practice, and MPC-based control is considered since it is the basis of the control methods developed in the remainder of this chapter. Later on, we will compare the performance of the different control methods in simulation case studies.

3.2.1 Principles of ramp metering

Ramp metering is a control measure that has two goals: minimizing the disturbances caused by the merging behavior near the ramp, and limiting the flow on ramps connected to a freeway. Minimizing the disturbances is done by releasing vehicles from the ramp with a constant rate, which allows for a smooth merging into the freeway traffic. Limiting the flow that leaves an on-ramp leads to a lower flow on the freeway, but can result in a queue on the ramp. Off-ramp metering limits the flow that exits from the freeway, which results in lower flows in the urban area, but increases the density on the freeway. Ramp metering is implemented with traffic signals, that usually allow only one vehicle to drive through during each green period.

Remark 3.1 Although the traffic is a continuous process, we will consider discrete time steps in this chapter. The time step used for the freeway simulation, T_{f}, is counted using the freeway time step counter k_{f}. The settings of the control measures will be computed every controller time step T_{c}. Here, the index k_{c} is used to denote the controller time step index. We assume that T_{c} is an integer multiple of T_{f}. For ease of notation we define the set of simulation steps k_{f} that correspond to a given interval $[k_{\mathrm{c}}^a T_{\mathrm{c}}, k_{\mathrm{c}}^b T_{\mathrm{c}}]$ of controller time steps as follows:

$$\mathscr{K}_{\mathrm{f}}(k_{\mathrm{c}}^a, k_{\mathrm{c}}^b) = \left\{ k_{\mathrm{c}}^a \frac{T_{\mathrm{c}}}{T_{\mathrm{f}}}, k_{\mathrm{c}}^a \frac{T_{\mathrm{c}}+1}{T_{\mathrm{f}}}, \ldots, k_{\mathrm{c}}^b \frac{T_{\mathrm{c}}}{T_{\mathrm{f}}} - 1 \right\} .$$

□

On-ramp metering

On-ramp metering installations are located at on-ramps of freeways, as illustrated in Figure 3.2. During the green period, only 1 or 2 vehicles are allowed to leave the on-ramp. In

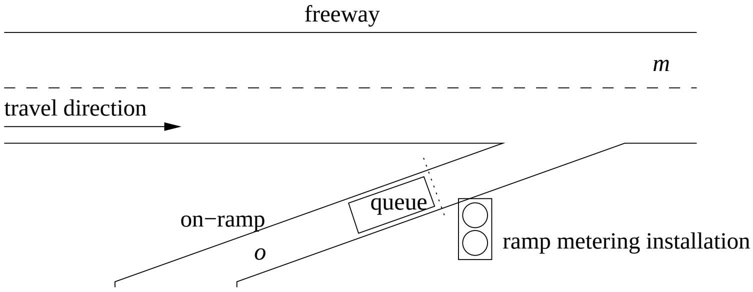


Figure 3.2: On-ramp with ramp metering installation.

this way the flow that can enter the freeway is limited, which prevents high densities and disturbances of the traffic flows on the freeway link downstream of the on-ramp. We will now describe the effects of applying ramp metering at a high level. The resulting ramp metering rate should be translated into green times and red times by lower level controllers.

The ramp metering rate $r_o^*(k_c)$ at control time step k_c gives the fraction of the capacity flow $Q_{\mathrm{cap},o}$ of the on-ramp o that is allowed to drive on. Before the sample $r_o^*(k_c)$ can be used, it must be translated to the corresponding freeway simulation time steps, e.g., using a zero-order-hold method:

$$r_o(k_f) = r_o^*(k_c) \text{ for } k_f \in \mathscr{K}_f(k_c, k_c+1) \quad . \tag{3.1}$$

The ramp metering rate can vary between a maximum and a minimum value: $r_{\min} \leq r_o(k_f) \leq r_{\max}$.

The flow that can enter the freeway via on-ramp o is then given by:

$$q_o^{\mathrm{real}}(k_f) = \min\Big(\min\big(r_o(k_f)Q_{\mathrm{cap},o},\ Q_{\lim}(\rho_{m,1}(k_f))\big),\ q_o^{\mathrm{int}}(k_f)\Big) \tag{3.2}$$

where $q_o^{\mathrm{int}}(k_f)$ is the flow that intends to enter freeway link m via on-ramp o during time interval $[k_f T_f, (k_f+1)T_f)$, $r_o(k_f)$ is the ramp metering rate, $Q_{\lim}(\rho_{m,1}(k_f))$ is the maximum flow that can enter the freeway taking into account limiting effect of the current density at the freeway, and $q_o^{\mathrm{real}}(k_f)$ is the flow that actually enters the freeway.

Originally, on-ramp metering installations were designed to prevent congestion on freeways, as described in [29, 35, 121, 190]. When the on-ramp metering installations were actually implemented, it became clear that they also influenced the route choice of the drivers [63, 149]. The explanation for this phenomenon is that ramp metering changes the travel times on the routes, and since route choice is mainly based on these travel times, some drivers will select another route when a ramp metering installation is present. These ideas have led to research on corridor control where on-ramp metering installations are used to influence route choice, and to prevent rat-running [10, 81, 187]. Rat-running describes the phenomenon that when a freeway is congested, drivers leave this freeway via an off-ramp, travel over a local road, and then again enter the freeway via an on-ramp.

A disadvantage of on-ramp metering is that it causes queues on the on-ramps. These queues can become so long that they block intersections near the freeway, which decreases the traffic efficiency, and additionally can lead to noise and pollution in the urban network.

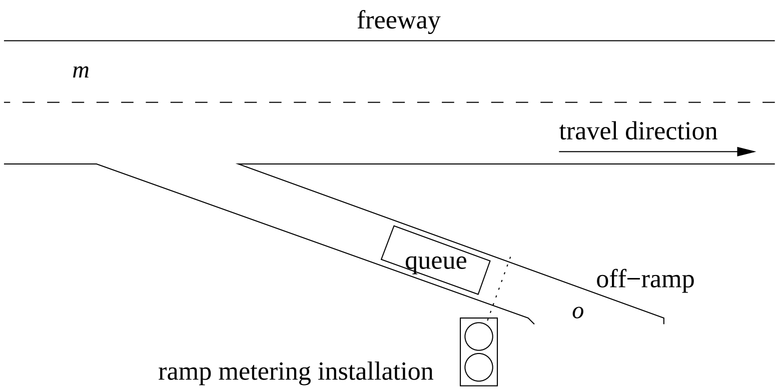


Figure 3.3: Off-ramp with off-ramp metering installation.

To prevent this, the selected control method must be able to take into account hard constraints on the queue lengths.

Off-ramp metering

Figure 3.3 shows the set-up of an off-ramp metering installation. Off-ramp metering limits the flow that can leave the freeway, and creates a visible queue at the off-ramp, which will further discourage drivers to exit the freeway. To the author's best knowledge, off-ramp metering has not yet been investigated in earlier research, and also not applied in practice. In this chapter we however shortly investigate the idea of off-ramp metering [69, 162], since we expect that it might be able to prevent rat-running, and to prevent a gridlock on the local roads. Since the off-ramp metering keeps the traffic on the freeway, the long distance traffic does not form long queues on the urban network. And since the number of vehicles that leaves the freeway is reduced, the number of vehicles in the urban network is decreased, which improves the traffic condition in the whole urban network.

Off-ramp metering limits the flow that can leave the freeway as follows:

$$q_o^{\mathrm{real}}(k_\mathrm{f}) = \min\left(r_o(k_\mathrm{f})Q_{\mathrm{cap},o}\ ,\ Q_{\mathrm{lim}}(\rho_o(k_\mathrm{f}))\ ,\ q_o^{\mathrm{int}}(k_\mathrm{f})\right) \tag{3.3}$$

where $q_o^{\mathrm{real}}(k_\mathrm{f})$ is the flow that really leaves freeway m toward off-ramp o, $q_o^{\mathrm{int}}(k_\mathrm{f})$ the flow that intends to leave the freeway, and $Q_{\mathrm{lim}}(\rho_o(k_\mathrm{f}))$ is the flow limit due to the density on the urban network, determined by (3.9).

A disadvantage of off-ramp metering is that the queues can become too long for the off-ramp, and spill back on the freeway. This can decrease the throughput on the freeway, and impair the safety since large speed-differences are created between the traffic on the first lane waiting to enter the off-ramp and on the second lane of the freeway. This can be solved by creating more space for the queue on the off-ramp itself, and by informing drivers upstream of the off-ramp about the active off-ramp metering installation, e.g., via a dynamic route information panel. In this way drivers can select another route before entering the queue.

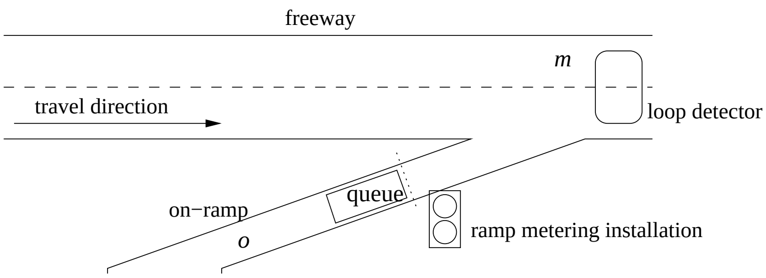


Figure 3.4: Location of the measurements and ramp metering installation for ALINEA.

3.2.2 Methods for ramp metering control

There are different methods to control ramp metering installations. A fixed ramp metering rate can be selected off-line. Other methods determine the ramp metering rate on-line, like ALINEA [123] and the method described in [149]. Cooperation of different ramp metering installations is also possible, see [10, 87]. Objectives of this kind of systems are, e.g., maximizing throughput on the freeway, maximizing the mean speed, reducing the shock waves to improve safety, minimizing the queue length and waiting time on the ramps, or minimizing the total time spent in the network.

Below we will describe three different methods: fixed-time control, ALINEA, and MPC-based control using the METANET model as prediction model. Fixed-time control is the most basic method for ramp metering, and ALINEA is a well-known algorithm which is applied at many locations. In this chapter we use the MPC-based control strategy for the ramp metering controller.

Fixed-time control

With fixed-time control the ramp metering installations are operating with a fixed ramp metering rate. This rate is determined off-line, and can be selected based on historical measurements. Based on these historical demands, a prediction of the future traffic flows can be made. Then, using, e.g., calculation models, tuning methods, or optimization algorithms, the ramp metering rate that leads to the most optimal performance of the controller can be selected. This ramp metering rate can have different values for different time periods, e.g., morning peak, evening peak, or during the rest of the day. The optimal ramp metering rate that is obtained off-line is then applied in practice via the real ramp metering installation.

ALINEA

ALINEA is a method for on-ramp metering developed in [63, 123]. It uses occupancy measurements on the freeway downstream of the on-ramp, as shown in Figure 3.4. The ramp metering rate is determined based on the occupancy downstream of the on-ramp:

$$r_o(k_\mathrm{f}) = r_o(k_\mathrm{f}-1) + K_o(\sigma_\mathrm{setpoint} - \sigma_{m,1}(k_\mathrm{f}))$$

The controller tries to keep this density near the set-point value σ_setpoint, which is selected in such a way that the density stays a little lower than the critical density $\rho_{\mathrm{crit},m}$, to allow as

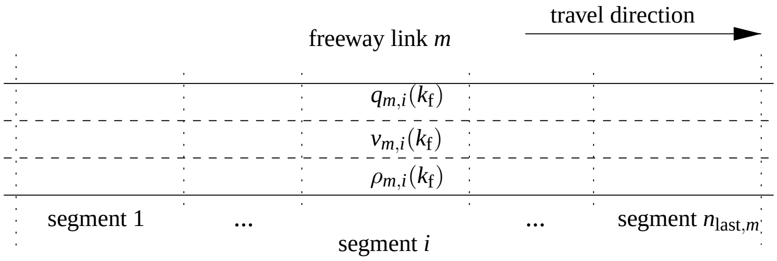


Figure 3.5: A link divided in segments.

much flow as possible without creating a traffic jam. K_{r_o} is a positive constant, and $\sigma_{m,1}(k_\mathrm{f})$ is the occupancy on the freeway segment downstream of the on-ramp.

Model Predictive Control using METANET as prediction model

Ramp metering installations can also be controlled using MPC [100], which works as follows [10]. At a given time $t = k_\mathrm{c}T_\mathrm{c} = k_\mathrm{f}T_\mathrm{f}$ the MPC controller uses a prediction model and numerical optimization (e.g., sequential quadratic programming (SQP) [19]) to determine the optimal ramp metering rate sequence $r_o^*(k_\mathrm{c}),\ldots,r_o^*(k_\mathrm{c}+N_\mathrm{p}-1)$ that minimizes a given performance indicator $J(k_\mathrm{c})$ over the time period $[k_\mathrm{c}T_\mathrm{c},(k_\mathrm{c}+N_\mathrm{p})T_\mathrm{c})$ based on the current state of the traffic network and on the expected demands over this period, where N_p is called the prediction horizon. Furthermore, within MPC a receding horizon approach is used in which at each control step only the first ramp metering input sample $r_o^*(k_\mathrm{c})$ is applied to the system during the period $[k_\mathrm{c}T_\mathrm{c},(k_\mathrm{c}+1)T_\mathrm{c})$. When the first sample has been applied, the horizon is shifted, new measurements are made, and the process is repeated all over again.

In this chapter we use the METANET model developed in [106] as prediction model within the MPC-based control method. Here we will shortly present the model, a more detailed description is given in Chapter 2 and in [106, 153]. In METANET the freeway network is represented as a graph with nodes and links, where the links correspond to freeway stretches with uniform characteristics; the nodes are placed at on-ramps and off-ramps, where two or more freeways connect, or where the characteristics of the freeway change. Links are divided into one of more segments with a length of about 500 m, as illustrated in Figure 3.5. The evolution of the traffic flows is characterized by the average density $\rho_{m,i}(k_\mathrm{f})$, flow $q_{m,i}(k_\mathrm{f})$, and space mean speed $v_{m,i}(k_\mathrm{f})$ for each segment i of each link m at time $t = k_\mathrm{f}T_\mathrm{f}$:

$$\rho_{m,i}(k_\mathrm{f}+1) = \rho_{m,i}(k_\mathrm{f}) + \frac{T_\mathrm{f}}{L_m n_m}\left[q_{m,i-1}(k_\mathrm{f}) - q_{m,i}(k_\mathrm{f})\right] \tag{3.4}$$

$$q_{m,i}(k_\mathrm{f}) = \rho_{m,i}(k_\mathrm{f})v_{m,i}(k_\mathrm{f})n_m$$

$$\begin{aligned} v_{m,i}(k_\mathrm{f}+1) = {} & v_{m,i}(k_\mathrm{f}) + \frac{T_\mathrm{f}}{\tau}\left(V(\rho_{m,i}(k_\mathrm{f})) - v_{m,i}(k_\mathrm{f})\right) + \\ & \frac{T_\mathrm{f}}{L_m} v_{m,i}(k_\mathrm{f})\left[v_{m,i-1}(k_\mathrm{f}) - v_{m,i}(k_\mathrm{f})\right] - \\ & \frac{\nu T_\mathrm{f}[\rho_{m,i+1}(k_\mathrm{f}) - \rho_{m,i}(k_\mathrm{f})]}{\tau L_m[\rho_{m,i}(k_\mathrm{f}) + \kappa]} \end{aligned} \tag{3.5}$$

where L_m, $V(\rho_{m,i}(k_\mathrm{f}))$, and n_m are respectively the length of the segments of freeway link m, the desired speed of the drivers on segment i of freeway link m, and the number of lanes of freeway link m, while τ, ν and κ are model parameters. The desired speed $V(\rho_{m,i}(k_\mathrm{f}))$ is computed as follows:

$$V(\rho_{m,i}(k_\mathrm{f})) = v_m^\mathrm{free} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k_\mathrm{f})}{\rho_{\mathrm{crit},m}}\right)^{a_m}\right] \tag{3.6}$$

where a_m is a model parameter, v_m^free is the free flow speed and $\rho_{\mathrm{crit},m}$ is the critical density in link m, i.e. the density where congestion starts to appear.

Freeway links are coupled via nodes, e.g., on-ramps, off-ramps, or intersections. Flows that enter a node n are distributed over the leaving nodes according to the turning rates as follows:

$$Q_{\mathrm{tot},n}(k_\mathrm{f}) = \sum_{\mu \in L_n^\mathrm{enter}} q_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f}) \tag{3.7}$$

$$q_{m,0}(k_\mathrm{f}) = \beta_{n,m}(k_\mathrm{f}) Q_{\mathrm{tot},n}(k_\mathrm{f}) \text{ for each } m \in L_n^\mathrm{leave} \tag{3.8}$$

where $Q_{\mathrm{tot},n}$ is the total flow entering node n, L_n^enter is the set of all freeway links entering node n, $n_{\mathrm{last},\mu}$ is the last segment of link μ, $\beta_{n,m}$ is the turning rate from node n to leaving link m which is determined with one of the traffic assignment models that will be described in Sections 3.3 and 3.4, and L_n^leave the set of leaving links of node n. The virtual downstream density $\rho_{\mu,n_{\mathrm{last},\mu}+1}(k_\mathrm{f})$ of the links μ that enter node n is approximated with:

$$\rho_{\mu,n_{\mathrm{last},\mu}+1}(k_\mathrm{f}) = \frac{\sum_{m \in L_n^\mathrm{leave}} \rho_{m,1}^2(k_\mathrm{f})}{\sum_{m \in L_n^\mathrm{leave}} \rho_{m,1}(k_\mathrm{f})} .$$

The virtual downstream density is used in the speed update equation (3.5) for the last segment $n_{\mathrm{last},\mu}$ of link μ. The virtual entering speed $v_{m,0}(k_\mathrm{f})$ of leaving link m of node n is given by:

$$v_{m,0}(k_\mathrm{f}) = \frac{\sum_{\mu \in L_n^\mathrm{enter}} v_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f}) q_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f})}{\sum_{\mu \in L_n^\mathrm{enter}} q_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f})} .$$

The virtual entering speed is used in the speed update equation (3.5) for the first segment of link m.

The effect of on-ramp metering is described by (3.2). The maximum flow $Q_\mathrm{lim}(\rho_{m,1}(k_\mathrm{f}))$ that can enter the freeway taking into account the limiting effect of the density on the freeway, is computed as follows:

$$Q_\mathrm{lim}(\rho_{m,1}(k_\mathrm{f})) = \frac{\rho_{\mathrm{max},m} - \rho_{m,1}(k_\mathrm{f})}{\rho_{\mathrm{max},m} - \rho_{\mathrm{crit},m}} Q_{\mathrm{cap},o}$$

where m is the freeway link to which on-ramp o is connected, $Q_{\mathrm{cap},o}$ is the capacity of the on-ramp, and $\rho_{\mathrm{max},m}$ is the maximum density at freeway link m. The traffic that is not able to enter the freeway forms a queue with length $w_o(k_\mathrm{f})$ at the on-ramp:

$$w_o(k_\mathrm{f}+1) = w_o(k_\mathrm{f}) + T_\mathrm{f}\left(q_o^\mathrm{int}(k_\mathrm{f}) - q_o^\mathrm{real}(k_\mathrm{f})\right) \quad .$$

The flow that has the intent to leave the on-ramp, as used in (3.2), is given by

$$q_o^\mathrm{int}(k_\mathrm{f}) = q_o^\mathrm{dem}(k_\mathrm{f}) + \frac{w_o(k_\mathrm{f})}{T_\mathrm{f}}$$

where q_o^dem is the demand at on-ramp o.

The effect of the off-ramp metering is described by (3.3). Within the METANET model an off-ramp can be seen as a splitting node. The flow $q_o^\mathrm{int}(k_\mathrm{f})$ that intends to leave the freeway can be computed with the node equations (3.7) and (3.8):

$$Q_{\mathrm{tot},n}(k_\mathrm{f}) = q_{m,n_{\mathrm{last},m}}(k_\mathrm{f})$$

$$q_o^\mathrm{int}(k_\mathrm{f}) = \beta_{n,o}(k_\mathrm{f}) Q_{\mathrm{tot},n}(k_\mathrm{f}) \quad .$$

This flow is however limited by the density on the off-ramp, which results in a maximum flow that can enter the urban network:

$$Q_\mathrm{lim}(\rho_o(k_\mathrm{f})) = \frac{\rho_{\mathrm{max},o} - \rho_o(k_\mathrm{f})}{\rho_{\mathrm{max},o} - \rho_{\mathrm{crit},o}} Q_{\mathrm{cap},o} \quad . \tag{3.9}$$

When the flow that enters the off-ramp is determined, the flow that leaves the last segment of freeway link m toward the freeway link μ that leaves node n can be computed with:

$$q_{m,n_{\mathrm{last},m}}^\mathrm{real}(k_\mathrm{f}) = q_o^\mathrm{real}(k_\mathrm{f}) + \beta_{n,\mu}(k_\mathrm{f}) Q_{\mathrm{tot},n}(k_\mathrm{f})$$

Remark 3.2 Due to the properties of the METANET model a part of the vehicles that intends to leave the freeway but that is not able to do so will change its route and continue to travel on the freeway. As a result the density on the freeway segment upstream of the off-ramp, as determined with the model, will be lower than expected. This can be solved by using the destination dependent version of the METANET model. This however increases the required computational effort. □

The reaction on the speed limits is modeled by changing the desired speed of the drivers [64]. To include the reaction of the drivers (3.6) is replaced by:

$$V(\rho_{m,i}(k_\mathrm{f})) = \min\left(v_m^\mathrm{free} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k_\mathrm{f})}{\rho_{\mathrm{crit},m}}\right)^{a_m}\right],\ (1+\alpha) v_{m,\mathrm{control}}(k_\mathrm{f})\right) \tag{3.10}$$

where $v_{m,\mathrm{control}}(k_\mathrm{f})$ is the applied speed limit, and α is a compliance factor that expresses to which extent the speed limits are obeyed. The value of α can change depending on the kind of drivers on the road or depending on the level of enforcement.

3.3 Anticipative control using equilibrium-based dynamic traffic assignment

The first control strategy that we develop uses dynamic traffic assignment to determine the within-day route choice of the drivers [157, 160]. The controller uses the DTA model to anticipate on the change in route choice due to the control actions. In this section we first describe the equilibrium-based dynamic traffic assignment (DTA) algorithm, and then the application to ramp metering. Next, the control strategy is illustrated with a case study, where we make a comparison with the ALINEA ramp metering strategy.

3.3.1 Equilibrium-based dynamic traffic assignment

Dynamic traffic assignment assumes that every driver assigns a cost to every possible route r between his origin and intended destination, and selects the route with the lowest cost. This will result in a user equilibrium assignment, where the costs of alternative routes are equally high [182]. The cost considered by the drivers can contain many different terms, such as the length of the road, the number of intersections, the environment, or the travel time. In this chapter we only use the travel time to describe the cost of a route, as suggested in [42, 142]. We use the instantaneous travel time since it is easy to compute and since it provides a relatively good approximation of the real travel time that the drivers will experience [175]. Note however that in literature many other algorithms are available, which can obtain similar or even better approximations of the experienced travel time, but these algorithms often require more measurements or have a larger computation time, which makes them less suitable for the use in on-line controllers. The instantaneous travel time for route r is computed as follows:

$$\tau_r(k_\mathrm{f}) = \sum_{(m,i)\in M_r^{\mathrm{link}}} \frac{L_m}{v_{m,i}(k_\mathrm{f})} \quad , \tag{3.11}$$

where M_r^{link} is the set of pairs of indexes (m,i) of all links and segments belonging to route r.

With the costs the dynamic traffic assignment can be obtained, using a DTA algorithm. In literature several methods exist to compute dynamic traffic assignments based on a cost function (see [15, 33, 177]). A disadvantage of these methods is often that they require much computation time. For the use in real-time model-based controllers the assignment must be computed every controller time step and so these models cannot be used. The method that we use in this paper requires less computation time, which also means that it looses accuracy. The model that we develop consists of two parts. First the static *perceived equilibrium assignment* is determined, which is seen as the equilibrium assignment that should appear based on the current demand in the network. Next, to approximate a dynamic traffic assignment, the current assignment is adapted incrementally in such a way that it converges toward this *perceived equilibrium assignment*.

Perceived equilibrium assignment

We determine the *perceived equilibrium assignment* based on the drivers' perceived knowledge about the current state of the network and the current demands. We assume that the state of the network is not exactly known by the drivers, but that the drivers have been gathering information about the traffic for some time span τ_{info}. They use the average of the information that is gathered during the period $[k_{\text{f}} - \tau_{\text{info}}, k_{\text{f}})$ to determine their estimation of the current state of the network and their estimation of the expected demands. These estimations are used by the drivers to determine their perceived equilibrium assignment. The larger τ_{info}, the slower the response of the route choice behavior of the drivers to varying traffic demands and metering rates will be. A typical value for τ_{info} is 30 min. Further, we assume that the drivers do not update their perceived equilibrium assignment at every time step k_{f}. Therefore we introduce the update time step T_{a} of the perceived equilibrium assignment, which is counted with index k_{a}, and which is an integer multiple of T_{f}.

To compute an equilibrium assignment there exist several methods, some of them described in [142, 176, 179, 184]. In this paper we use the 'Method of the Successive Averages' (MSA) [134] since it is a relatively simple and fast algorithm. MSA is an iterative method that computes the cost (in this case the travel times) of different routes according to the flow $q_{r,j}^{\text{MSA}}(k_{\text{a}})$ on each route r in iteration j departing during time step k_{a}. These flows are determined by making a prediction of the traffic flows over the period $[k_{\text{a}}T_{\text{a}}, (k_{\text{a}}+N_{\text{MSA}})T_{\text{a}}) = [k_{\text{f}}T_{\text{f}}, (k_{\text{f}}+N_{\text{MSA}}\frac{T_{\text{a}}}{T_{\text{f}}})T_{\text{f}})$ with the METANET model using the turning rates corresponding to the flows obtained in iteration $j-1$, where N_{MSA} is the prediction horizon for the MSA algorithm. Then for iteration j, all flows are directed to the route with the lowest cost, which results in the all-or-nothing (AON) assignment flows $q_{r,j}^{\text{AON}}(k_{\text{a}})$. With these AON-flows, the flows for the next iteration are updated:

$$q_{r,j+1}^{\text{MSA}}(k_{\text{a}}) = \left(1 - \frac{1}{j}\right) q_{r,j}^{\text{MSA}}(k_{\text{a}}) + \left(\frac{1}{j}\right) q_{r,j}^{\text{AON}}(k_{\text{a}})$$

where $q_{r,j}^{\text{MSA}}$ is the flow towards route r during iteration j of the MSA algorithm, and $q_{r,j}^{\text{AON}}$ the flow towards route r determined by the all-or-noting assignment after iteration j. The stopping criterion is based on a maximum value for the difference between two successive iteration flows: when the difference is below this specified value the algorithm terminates. To prevent long computation times the algorithm will also exit when a maximum number of iterations is reached.

The resulting flows are used to determine the equilibrium turning rates. For a network with non-overlapping routes (as used in the case study below) this results in:

$$\beta_{n,m}^{\text{MSA}^*}(k_{\text{a}}) = \frac{\sum_{\zeta \in R_{n,m}^{\text{link}}} q_{\zeta,j_{\text{final}}}^{\text{MSA}}(k_{\text{a}})}{Q_{\text{tot},n}(k_{\text{a}})}$$

with $q_{r,j_{\text{final}}}^{\text{MSA}}(k_{\text{a}})$ the equilibrium flow determined with the MSA algorithm at MSA time step T_{a}, $\beta_{n,m}^{\text{MSA}}(k_{\text{a}})$ the corresponding turning rate on node n toward link m, and $R_{n,m}^{\text{link}}$ the set of all routes that passes through node n toward link m.

Dynamic traffic assignment

Now we assume that the static *perceived equilibrium assignment* as formulated above is not present in the network yet, but that the drivers try to reach this assignment dynamically. We assume that the current traffic assignment will change toward the *perceived equilibrium* *assignment* in an exponential way, with time constant τ_{reac}. This results in an adaptation of the turning rates according to:

$$\beta_{n,m}(k_{\text{f}}+1) = \beta_{n,m}(k_{\text{f}}) + (\beta_{n,m}^{\text{MSA}}(k_{\text{a}}) - \beta_{n,m}(k_{\text{f}}))(1 - e^{-\frac{T_{\text{f}}}{\tau_{\text{reac}}}}) \qquad (3.12)$$
$$\forall k_{\text{f}} \in \{k_{\text{a}}\frac{T_{\text{a}}}{T_{\text{f}}}, \ldots, (k_{\text{a}}+1)\frac{T_{\text{a}}}{T_{\text{f}}}\} \ .$$

The parameter τ_{reac} influences how fast the current assignment converges toward the presumed equilibrium assignment. This swiftness in practice depends on the time that elapses before a congestion can be noticed by drivers that still have to make their route choice. Note that this time can be shortened by providing travel time information, since then the drivers can be informed about the congestion before they actually experience it.

3.3.2 Ramp metering using equilibrium-based DTA

The control method for the ramp metering controller that we develop in this section is based on MPC, as described in Section 3.2.2.

The controller uses the METANET model as prediction model for the evolution of the traffic flows, and the equilibrium-based DTA model of Section 3.3.1 to determine the traffic assignment. Since the controller has to be able to determine the change in route choice that it induces, the perceived equilibrium should be updated at least once within the prediction horizon. This means that N_{p} should be selected such that

$$NT_{\text{a}} \leq N_{\text{p}}T_{\text{c}}$$

with N an integer larger than 1. The most suitable value depends on the re-routing dynamics in the network, which depend on the topology of the network [150].

As performance indicator we will consider the total time spent (TTS) by all vehicles in network (but note that the proposed approach also works for other performance indicators). The TTS can be computed as:

$$\text{TTS}(k_{\text{c}}) = T_{\text{f}} \sum_{k_{\text{f}} \in \mathscr{K}_{\text{f}}(k_{\text{c}}, k_{\text{c}}+N_{\text{p}})} \left(\sum_{(m,i)\in M} L_m n_m \rho_{m,i}(k_{\text{f}}) + \sum_{o\in O} w_o(k_{\text{f}}) \right) \qquad (3.13)$$

where M is the set of pairs of indexes (m,i) of all links in the network, and O the set of all origins.

The MPC strategy can handle hard constraints. This makes it possible to prevent blocking of urban intersections or too long waiting times by setting constraints on the queue length or the metering rates:

$$w_o(k_{\text{f}}) \leq w_o^{\max}$$
$$r_o^{\min} \leq r_o(k_{\text{f}}) \leq r_o^{\max} \ .$$

3.3.3 Case study

A simple network will be used to illustrate the effects of the MPC-based method for anticipative ramp metering control using equilibrium-based DTA. The layout of the network is

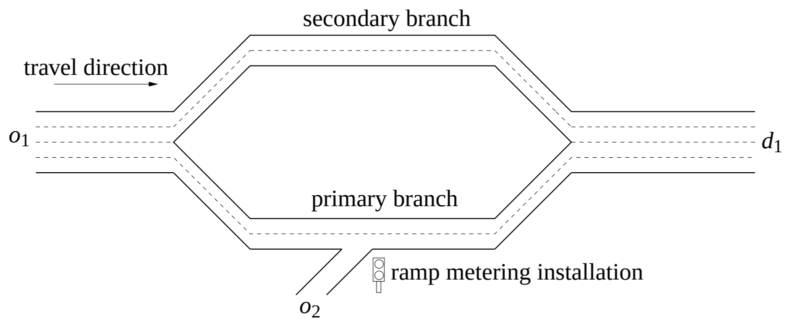


Figure 3.6: Schematic representation of the freeway network of the case study described in Section 3.3.3.

shown in Figure 3.6, where the arrow gives the direction of the traffic flows. The network consists of a freeway with four lanes that bifurcates into two branches of two lanes each. Downstream both branches join in a four-lane freeway. Both four-lane freeway links are 3 km long. The lower two-lane branch is the primary branch. The primary branch is 6 km long and an on-ramp is present in the middle of the branch. The secondary branch is longer than the primary branch and is 8 km long. Route 1 follows the primary branch, and route 2 the secondary. The traffic originating from the mainstream origin distributes over the two branches using the route choice mechanism described in Section 3.3.1. We perform two different experiments. First we compare the MPC-based controller developed in this section with ALINEA, and next we investigate the effects of the MPC-based controller when maintenance works are performed in the network.

Comparison with ALINEA

Simulations have been performed to compare the developed control strategy with the existing ramp metering strategy ALINEA [123].

In this case study the following parameter settings are used: $\tau_{\text{info}} = 30\,\text{min}$, $\tau_{\text{reac}} = 45\,\text{min}$, $T_{\text{a}} = 5\,\text{min}$, $T_{\text{f}} = 10\,\text{s}$, $T_{\text{c}} = 1\,\text{min}$, $N_{\text{p}} = 15$, $r_o^{\max} = 1$, $r_o^{\min} = 0.1$, $K_o = 0.01$, and σ_{setpoint} is selected such that it corresponds to a density of 34 veh/km/lane. A period of four hours is simulated. We simulate a traffic scenario with a constant traffic demand at the mainstream origin o_1 equal to 4500 veh/h. The traffic demand on the on-ramp o_2 is 100 veh/h at the start of the simulation at 6.00 a.m., increases to 800 veh/h after one hour, decreases again to 100 veh/h after two hours, and stays constant until the end of the simulation at 10.00 a.m., see Figure 3.7.

The results of the simulation with ALINEA ramp metering are shown in Figure 3.8. At 7 a.m. the peak in the on-ramp demands starts. Figure 3.8(a) shows the increase in density on the segment downstream of the on-ramp at that moment. As a reaction on this high density the metering signal becomes active and goes to zero, as can be seen in Figure 3.8(b). This low metering rate causes a drop in the density below the original density (Figure 3.8(a)). This drop has two effects: more drivers choose the first route so the turning rates change (Figure 3.8(c)), and the metering rate becomes higher, so more traffic can enter the freeway,

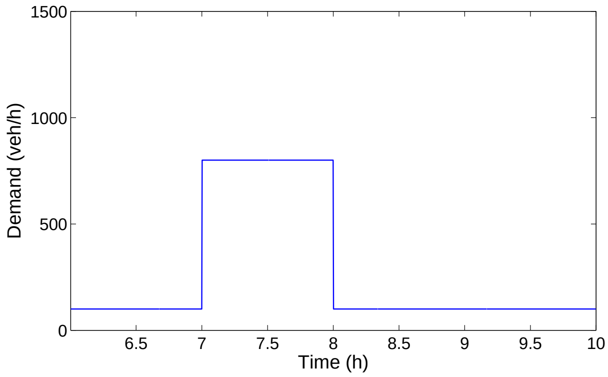
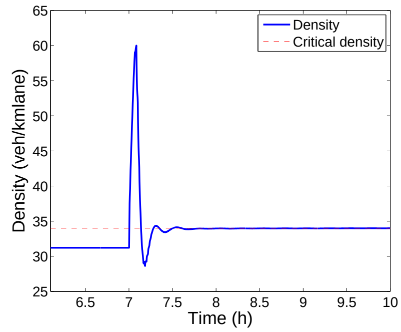


Figure 3.7: Traffic demand on the on-ramp for the case study in which we compare the developed DTA-based control method with ALINEA.

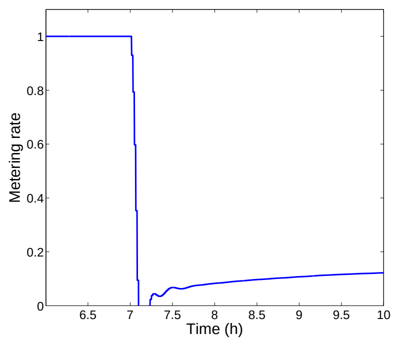
as shown in Figure 3.8(d). Then the density increases again, and converges to the critical density after 7.45 a.m. The controller will try to keep the density on this value, in which it succeeds for the rest of the simulation. In Figure 3.8(e) the travel times on the two routes are shown. Since at 7.30 a.m. the travel time on route 1 is longer than on route 2, more traffic starts taking route 2. This leads to a slowly decreasing amount of traffic on route 1, resulting in a decreasing travel time on route 1 until the travel times converge to the equilibrium values. Figure 3.8(f) shows the queue on the on-ramp. At 7.10 a.m. the ramp metering signal becomes active, and the queue starts to grow. After 8.00 a.m. the peak of the on-ramp demand ends, and the queue starts to clear. The total time spent in the network is 3618.9 veh·h for the ALINEA method.

The results obtained with the anticipative MPC strategy are shown in Figure 3.9. As the peak in the on-ramp demand starts, the density on the segment downstream of the on-ramp increases (Figure 3.9(a)). When this density becomes too high, the metering rate decreases, as shown in Figure 3.9(b). But the control method keeps the density above the critical density, on 40 veh/lane/km. This means that the travel time on the first route, which is shown in Figure 3.9(e), stays high, resulting in more drivers selecting route 2 (see Figure 3.9(c)). On the on-ramp a queue starts to grow, until 8 a.m. At this moment the peak in the on-ramp demand ends, and now the queue starts to empty. Many vehicles enter the freeway from the on-ramp, as can be seen in Figure 3.9(d), resulting in a longer travel time on route 1, and thus more traffic turning toward route 2. But then the density becomes lower and the travel times decrease, leading to more drivers selecting route 1. After 9 a.m. the flow from the origin becomes more stable, the queue empties, resulting in longer travel times, and more traffic selecting route 2. The total time spent in the network is 3300.5 veh·h for the MPC-based control method, which is an improvement of 9% compared to ALINEA. The improvement obtained with the MPC-based controller is mainly obtained by diverting more vehicles to route 2, so the queue on the on-ramp stays shorter, and can clear faster.

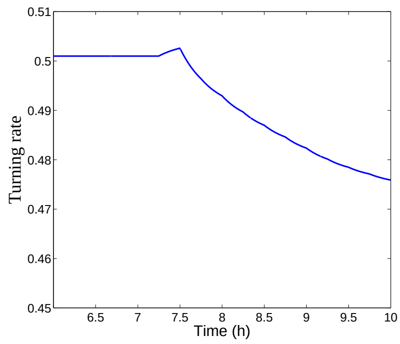
Remark 3.3 Note that the fluctuations in the control signal, as shown in Figure 3.9(b), are due to the many local minima in the cost function. Each control time step the optimization algorithm can determine a local minimum that differs from the minimum that has been



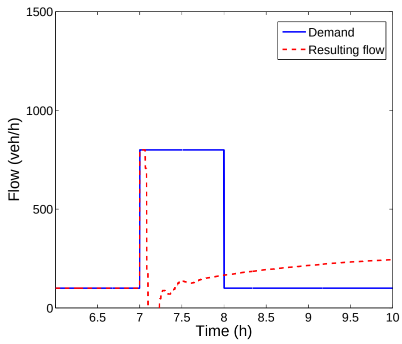
(a) Density on the segment downstream the on-ramp.



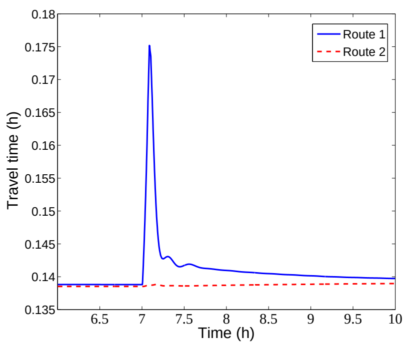
(b) Metering rates.



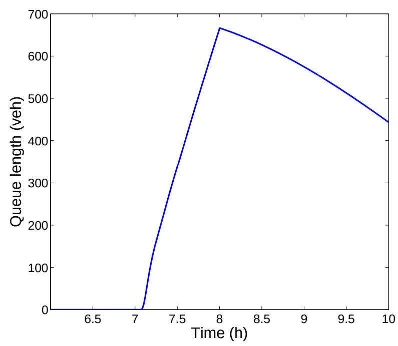
(c) Turning rate toward route 1.



(d) Flow on the on-ramp.

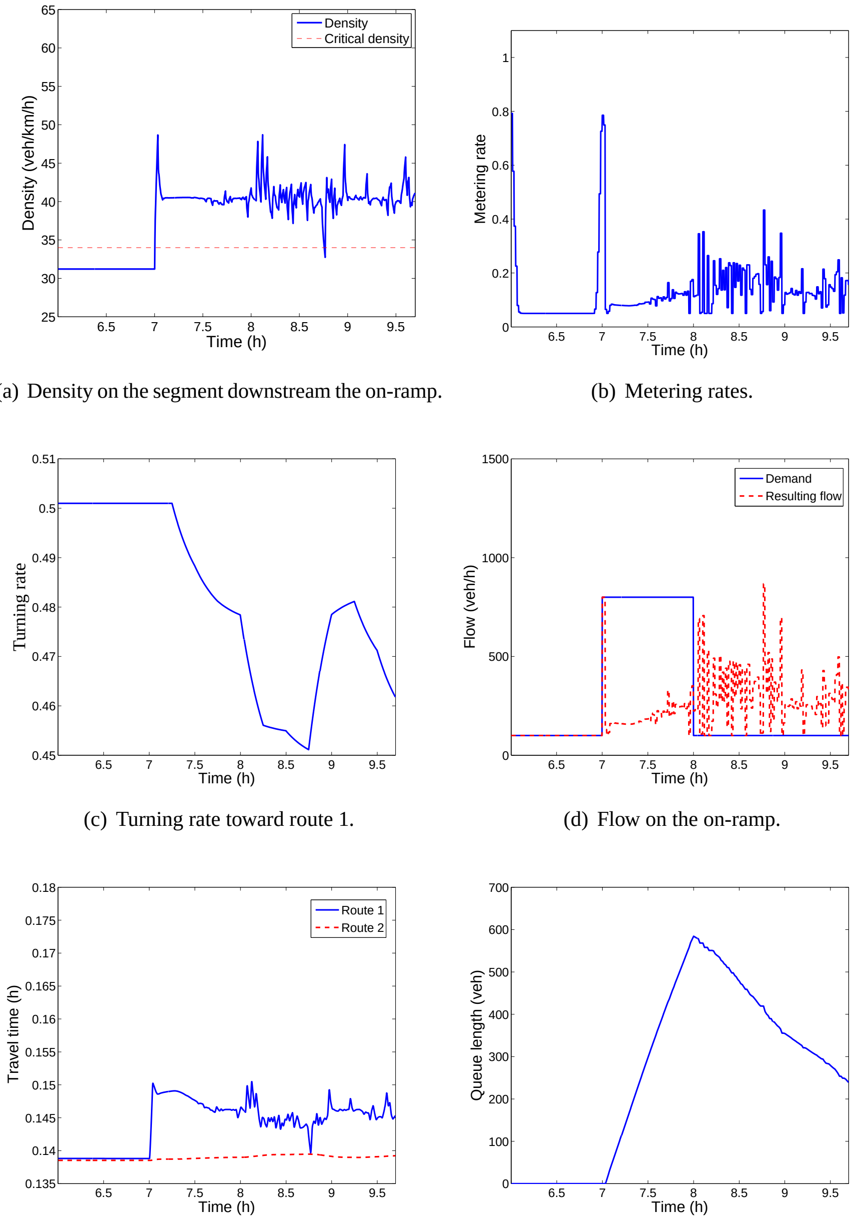


(e) Travel times on the two routes.



(f) Queue length on the on-ramp.

Figure 3.8: Simulation results with ALINEA ramp metering.



(a) Density on the segment downstream the on-ramp.

(b) Metering rates.

(c) Turning rate toward route 1.

(d) Flow on the on-ramp.

(e) Travel times on the two routes.

(f) Queue length on the on-ramp.

Figure 3.9: Simulation results with MPC-based ramp metering.

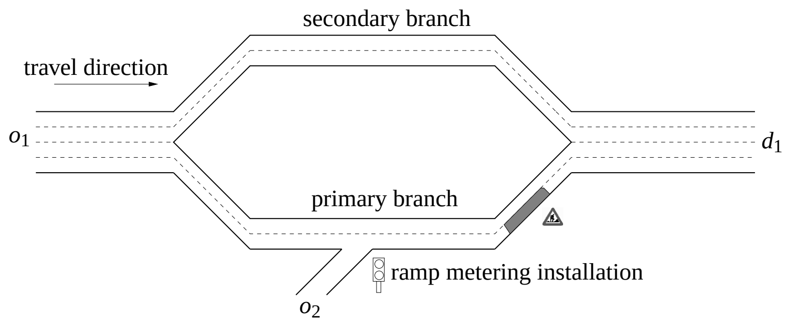


Figure 3.10: Schematic representation of the freeway network with maintenance works used in Section 3.3.3.

obtained during the previous control time step, which will lead to different values for the optimal control signal. The fluctuations in the control signal can be reduced by adding a penalty on the variations in the control signal to the cost function, or by increasing the number of runs performed by the optimization algorithm which will increase the probability of obtaining the same optimum at each control time step. □

Maintenance works

As second experiment we simulate the selected network when maintenance works are performed on route 1 as shown in Figure 3.10. We simulate the traffic from 4.00 a.m. until 11.00 a.m. The maintenance works result in a reduction of the number of lanes from 2 to 1 in the last segment (i.e., the last 500 m) of the primary branch. The maintenance works start at 5.00 a.m. and persist during the remainder of the simulation. The traffic demand on the mainstream is considered constant and equal to 4500 veh/h in this simulation. The traffic demand on the on-ramp is equal to 200 veh/h which starts to increase at 7.30 a.m. to a peak demand of 500 veh/h from 8.00 a.m. until 8.30 a.m., and decreases to 200 veh/h at 8.45 a.m., as given in Figure 3.11. We assume that the demand is known by the controller. The model and controller parameters are selected as follows: $\tau_{\text{info}} = 30$ min, $\tau_{\text{reac}} = 45$ min, $T_{\text{a}} = 5$ min, $T_{\text{f}} = 10$ s, $T_{\text{c}} = 1$ min, $N_{\text{p}} = 15$, $r_o^{\max} = 1$, $r_o^{\min} = 0.05$, $w_o^{\max} = 100$ veh.

To show the effects of ramp metering we have performed two simulations: one without ramp metering and one with ramp metering. The first is used to illustrate the functioning of the dynamic traffic assignment in the absence of control, and the second illustrates the change in route choice and the improved travel times due to the ramp metering.

The results for the simulation without ramp metering are shown in Figure 3.12. At the beginning of the simulation an equilibrium situation exists: the two travel times have the same value, see Figure 3.12(a). At 5.00 a.m. the maintenance works start. The travel times become different, resulting in a change in the turning rates (Figure 3.12(b)). From 6.00 a.m. until 7.30 a.m. the exponential convergence to the equilibrium turning rates can be seen. At 8 a.m. the traffic on the on-ramp increases, which causes a change in the travel time of route 1, and thus again a change in the turning rates. After 9.00 a.m. the exponential behavior again can be seen. Figure 3.12(c) shows the density on the segment downstream of the on-

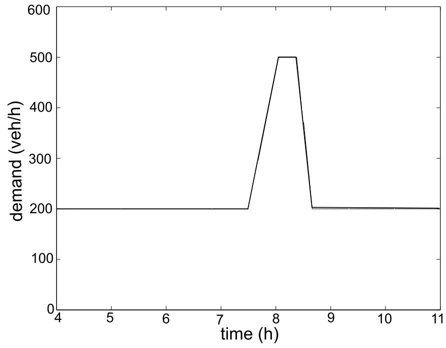
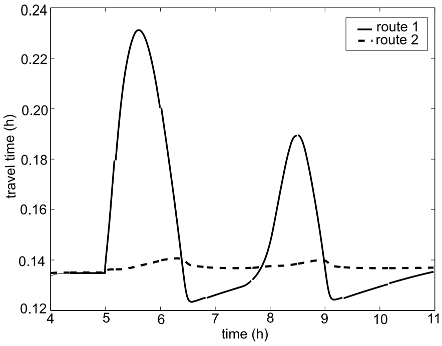
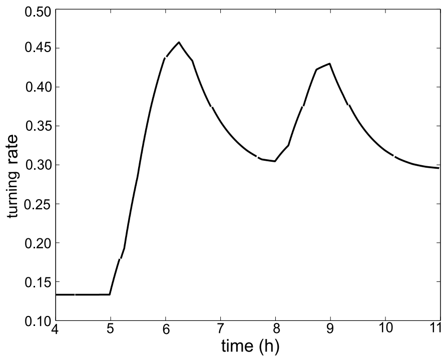


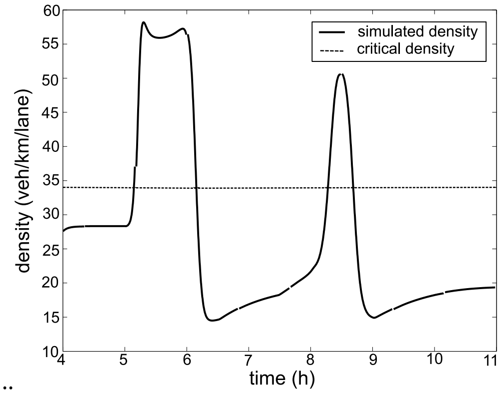
Figure 3.11: Traffic demand on the on-ramp.



(a) Travel time on the routes.



(b) Turning rate toward route 2.



(c) Density on segment downstream of the on-ramp

Figure 3.12: Simulation results without ramp metering.

ramp. The density is increasing when the maintenance works start. Due to the re-routing the density decreases until 6 a.m., when the travel time on route 1 becomes shorter than the travel time on route 2. When the travel time of route 2 becomes longer, the density starts to increase slowly, until 8 a.m. At 8 a.m. the traffic demand on the on-ramp increases, resulting in an increase of the density. At 8.30 a.m. the re-routing process results in a lower density. After 9.00 a.m. the density slowly increases towards its equilibrium value. The TTS in this situation is 6250 veh·h.

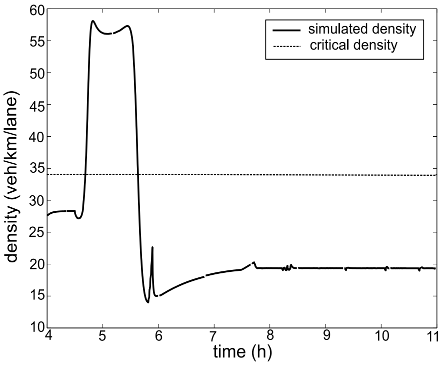
The results for the simulation with ramp metering are shown in Figure 3.13. The simulation with control starts with the same equilibrium traffic assignment as the simulation without control. At 5.00 a.m. the maintenance works start, resulting in a difference in travel times between the two routes, see Figure 3.13(e). The density on route 1 becomes higher than the critical value (Figure 3.13(a)), which results in the activation of the ramp metering (Figure 3.13(b)). When the density on route 1 becomes lower due to the re-routing, the ramp metering rate becomes higher so the queue, shown in Figure 3.13(f), can empty. At 7.30 a.m. the demand on the on-ramp increases. The ramp metering is activated so a direct increase of the density on the freeway can be prevented. This results in a queue on the on-ramp. This queue is emptied by increasing the metering rates slowly, resulting in a slow increase of the density on the freeway. This causes the turning rates to change slowly toward an equilibrium, see Figure 3.13(c).

Figure 3.13(d) shows the traffic demand on the on-ramp and the flow allowed to enter the freeway in one plot. The difference between the two leads to the queue length on the on-ramp, shown in Figure 3.13(c). When the maintenance works begin, the ramp metering limits the flow toward the freeway. At 6 a.m. the travel time on route 1 becomes short enough to empty the queue, resulting in a high flow leaving the on-ramp. When the demand on the on-ramp increases, a queue is formed. This queue decreases afterwards due to the increasing ramp metering rate. The TTS in this simulation is 5551 veh·h. This means that the use of the MPC-based anticipative control method results in a 11.2 % increase of the performance.

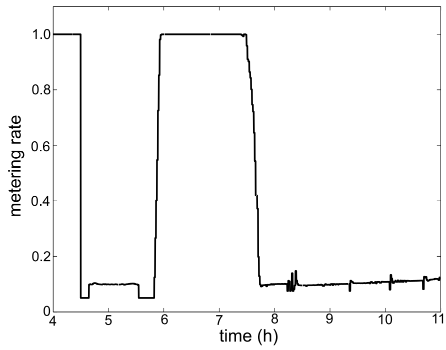
3.4 Anticipative control using route-choice-based DTA

The second control strategy, which is developed in this section, uses a route-choice-based DTA model which determines the traffic assignment implicitly. In this section we first develop the route choice model that is the basis of the DTA algorithm. This model describes within-day as well as day-to-day route choice. However, for the second control strategy we use only the within-day part of this en-route route-choice-based model. Within the control strategy, the DTA model is used to anticipate on changes in the assignment. As control measure, a ramp metering installation is used.

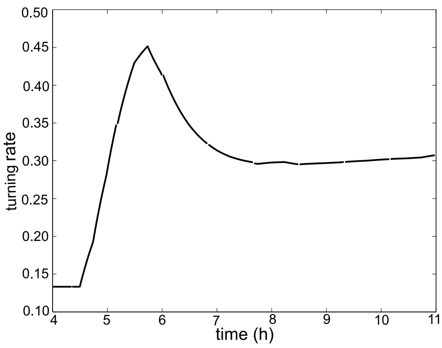
The performance of the controller is illustrated with a case study, in which three different ramp metering control strategies are compared. Finally we present the results that can be obtained when off-ramp metering is applied instead of on-ramp metering.



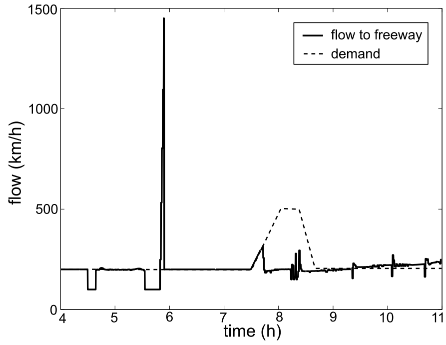
(a) Density on the segment downstream of the on-ramp.



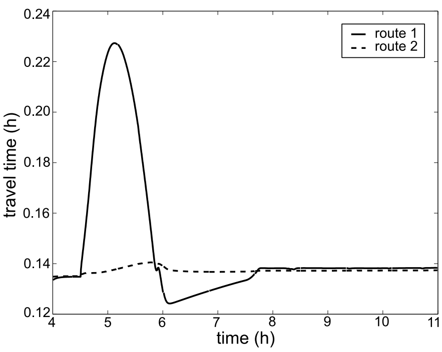
(b) Metering rate.



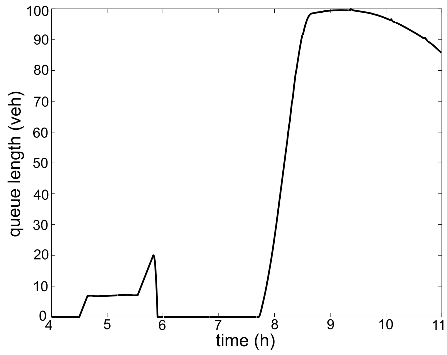
(c) Turning rate toward route 2.



(d) Traffic demand on the on-ramp, and flow allowed to enter the freeway.



(e) Travel time on the routes.



(f) Queue on the on-ramp.

Figure 3.13: Simulation results with MPC-based ramp metering.

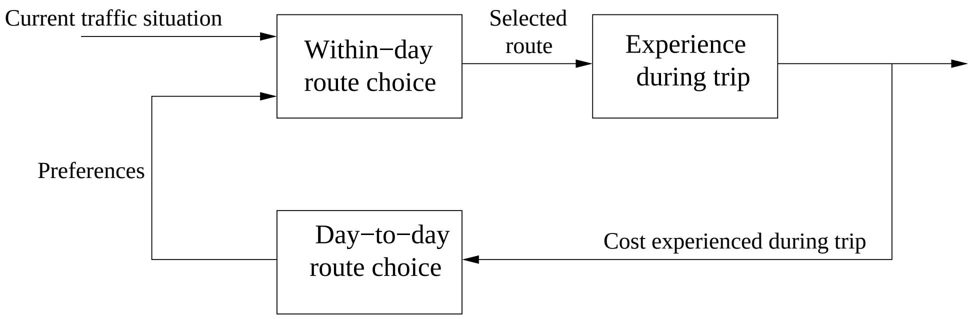


Figure 3.14: Overview of the route choice process.

3.4.1 Route choice model

We now develop the route choice model that is used for the route-choice-based DTA algorithm. A general route choice model describes the route choice behavior of drivers based on the current state of the network as experienced by the drivers, and does not have to result in an equilibrium traffic assignment. When we want to include a route choice model in a controller, we could select one of the models described in [18, 103]. However, these models are complex and detailed, and as a result they require too much computational effort to be used as a prediction model in an on-line controller. For the controller that we develop in this section, we formulate a route choice model based on statistical learning, which is also done in [32, 36, 80]. The model that we develop includes the day-to-day route choice as well as within-day route choice, as illustrated in Figure 3.14.

Within-day route choice process

We assume that the within-day route choice process of a driver is divided in three steps:

1. First the driver analyzes the current traffic situation on the road upstream of the splitting node. For the sake of simplicity of the exposition we will from now on assume that the driver makes his decisions based on one important variable only, e.g., the density. However, the approach can easily be generalized to the case where several variables determine the decision, such as the flow, speed, weather, time, or the news on the radio. We divide all possible densities in, say, three groups: low, medium, and high density. The driver selects to which group the current density belongs.

2. In the second step the driver estimates which route will result in the lowest costs, based on the current density. For sake of simplicity we assume that the only factor that influences these costs is the travel time, but note that the extension to more factors is straightforward. This means that the driver will select the route that according to his beliefs has the shortest travel time, given the current density.

3. During the last step, the driver decides whether he will indeed take the route with the lowest cost, or, e.g., when two routes have approximately equal costs, which route is

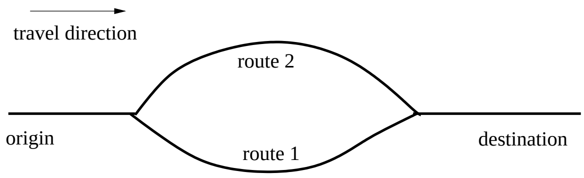


Figure 3.15: Network with one splitting node and two routes.

the best one to select.

Day-to-day route choice process

The main decisions in the within-day route choice process, as formulated above, are based on the knowledge of the driver. This knowledge is described by the day-to-day route choice model. The model that we will develop is based on Bayesian learning, and it is suitable for on-line control due to its simplicity and thus low computation times, due to the fact that it allows for the modeling of drivers habits, and since it correctly maintains the turning rates in equilibrium situations. The model that we develop contains estimated travel times for the different routes, and the probabilities that drivers select each route. After each day the variables are updated, using the experiences during this day. The days are counted with index k_d.

We assume that drivers determine the current state of the network based on the density on that they experience while they drive at the segment before the splitting node. We divide the possible densities in three groups with values ρ^low, ρ^medium, and ρ^high, with boundaries η_low and η_high between the groups. For each density group, the estimated travel time of each route is determined. These estimated travel times are computed by taking the average of earlier experienced travel times experienced under circumstances with the same density, using a forgetting factor because the last experiences are seen as more important according to [18]:

$$\bar{\tau}_r^{\mathrm{est},\Delta}(k_\mathrm{d}+1) = \omega\bar{\tau}_r^{\mathrm{experienced},\Delta}(k_\mathrm{d}) + (1-\omega)\bar{\tau}_r^{\mathrm{est},\Delta}(k_\mathrm{d}) \tag{3.14}$$

where $\bar{\tau}_r^{\mathrm{est},\Delta}(k_\mathrm{d}+1)$ is the new computed average travel time for density group Δ on day $k_\mathrm{d}+1$, with $\Delta \in [\rho^\mathrm{low}, \rho^\mathrm{medium}, \rho^\mathrm{high}]$, $\bar{\tau}_r^{\mathrm{est},\Delta}(k_\mathrm{d})$ is the average travel time at day k_d, $\omega \in [0,1]$ is a multiplication factor, and $\bar{\tau}_r^{\mathrm{experienced}}(k_\mathrm{d})$ is the last experienced travel time in a situation with a density in group Δ.

The probability that a driver selects a specified route given a density is based on earlier experiences. To compute this probability statistical information of previous trips is used. Note that we aggregate the knowledge of the drivers assuming that historical experiences can be accumulated. We will illustrate the procedure with a situation with one splitting node, at which two routes are available, see Figure 3.15. The following notation is used:

$S(r)$	route r is selected
$R(r)$	route r has the lowest cost
EQ	the routes have approximately equal costs
$P(S(r)\|\Delta)(k_\mathrm{d})$	probability that route r is selected given that the density is in density group Δ at day k_d
$P(S(r)\|R(r))(k_\mathrm{d})$	probability that route r is selected given that route r has the lowest cost at day k_d
$P(R(r)\|\Delta)(k_\mathrm{d})$	probability that route r has the lowest cost given that the density is in density group Δ at day k_d

We compute the probability of selecting route 1 when the density is in density group Δ at day k_d, $P(S(1)|\Delta)(k_\mathrm{d}+1)$. This probability is computed as:

$$\begin{aligned} P(S(1)|\Delta)(k_\mathrm{d}+1) = & P(S(1)|R(1))(k_\mathrm{d}) \cdot P(R(1)|\Delta)(k_\mathrm{d}) \\ & + P(S(1)|R(2))(k_\mathrm{d}) \cdot P(R(2)|\Delta)(k_\mathrm{d}) \\ & + P(S(1)|\mathrm{EQ})(k_\mathrm{d}) \cdot P(\mathrm{EQ}|\Delta)(k_\mathrm{d}) \end{aligned} \tag{3.15}$$

The first term describes the probability that the travel time on route 1 is the shortest under density Δ times the probability that route 1 is selected when route 1 is the shortest. The second term describes the probability that route 1 selected, while route 2 is the shortest times the probability that route 2 is the shortest given density Δ. The last term expresses the probability that route 1 is selected when the routes are equally long times the possibility that the routes are equally long given density Δ. To determine whether the travel times on the routes are equally long a tolerance is used. When the absolute difference between the two travel times is smaller than the tolerance, the routes are assumed to be equally long. Because this probability that both routes are equally long is also included, the model does not tend to a route choice of fifty-fifty when the travel times are equal, but maintains the route choice ratio that has resulted in the equal travel times.

The probabilities are updated after each day, based on measurements during the last day. The number of times $\phi_{A,B}^{\mathrm{experienced}}(k_\mathrm{d})$ that each combination of A and B (with A and B $\in [S(r), R(r), Delta]$), appears during day k_d. Based on these counters, the number of appearances $\phi_{A,B}^{\mathrm{est}}(k_\mathrm{d}+1)$ during the next day is computed, using a forgetting factor to describe the effect that last experiences are more important, e.g.:

$$\phi_{R(1),\Delta}^{\mathrm{est}}(k_\mathrm{d}+1) = \omega\phi_{R(1),\Delta}^{\mathrm{experienced}}(k_\mathrm{d}) + (1-\omega)\phi_{R(1),\Delta}^{\mathrm{est}}(k_\mathrm{d})$$

where $\phi_{R(1),\Delta}^{\mathrm{experienced}}(k_\mathrm{d})$ is the actual number of times that route 1 was the shortest during day k_d.

Based on this estimation the probabilities for the next day are adapted:

$$P(R(1)|\Delta)(k_\mathrm{d}) = \frac{\phi_{R(1),\Delta}^{\mathrm{est}}(k_\mathrm{d})}{\phi_{R(1),\Delta}^{\mathrm{est}}(k_\mathrm{d}) + \phi_{R(2),\Delta}^{\mathrm{est}}(k_\mathrm{d}) + \phi_{\mathrm{EQ},\Delta}^{\mathrm{est}}(k_\mathrm{d})}$$

$$P(S(1)|\mathrm{EQ})(k_\mathrm{d}) = \frac{\phi_{S(1),\mathrm{EQ}}^{\mathrm{est}}(k_\mathrm{d})}{\phi_{S(1),\mathrm{EQ}}^{\mathrm{est}}(k_\mathrm{d}) + \phi_{S(2),\mathrm{EQ}}^{\mathrm{est}}(k_\mathrm{d})}$$

Other probabilities are computed similar.

We use the probability $P(S(1)|\Delta)(k_\mathrm{d})$ computed with (3.15) as the fraction of the traffic that selects a route. This gives for the turning rates toward link 1 on node 1 on the network with two routes (as used in the case study below):

$$\beta_{1,1}^{\mathrm{routechoice},\Delta}(k_\mathrm{d}) = P(S(1)|\Delta)(k_\mathrm{d}) \ . \tag{3.16}$$

For a general network the density dependent turning rates can be computed according to:

$$\beta_{n,m}^{\mathrm{routechoice},\Delta}(k_\mathrm{d}+1) = \frac{\sum\limits_{r\in R_{n,m}^{\mathrm{link}}} P(\mathrm{S}(r)|\Delta)(k_\mathrm{d})}{\sum\limits_{\zeta\in R_{n}^{\mathrm{node}}} P(\mathrm{S}(\zeta)|\Delta)(k_\mathrm{d})}$$

where $R_{n,m}^{\mathrm{link}}$ is the set of routes that passes through node n toward freeway link m, and R_{n}^{node} is the set of routes that passes through node n. This leads to the following formulation for the turning rates at day k_d:

$$\beta_{n,m}^{\mathrm{routechoice}}(k_\mathrm{f}) = \begin{cases} \beta_{n,m}^{\mathrm{routechoice},\rho^{\mathrm{low}}}(k_\mathrm{d}) & \rho_{m,0}(k_\mathrm{f}) \le \eta_{\mathrm{low}} \\ \beta_{n,m}^{\mathrm{routechoice},\rho^{\mathrm{medium}}}(k_\mathrm{d}) & \eta_{\mathrm{low}} \le \rho_{m,0}(k_\mathrm{f}) \le \eta_{\mathrm{high}} \\ \beta_{n,m}^{\mathrm{routechoice},\rho^{\mathrm{high}}}(k_\mathrm{d}) & \eta_{\mathrm{high}} \le \rho_{m,0}(k_\mathrm{f}) \end{cases} ,$$

where $\rho_{m,0}(k_\mathrm{f})$ is the virtual upstream density of node n, computed with:

$$\rho_{m,0}(k_\mathrm{f}) = \frac{\sum\limits_{\mu\in L_{n}^{\mathrm{enter}}} \rho_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f}) q_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f})}{\sum\limits_{\mu\in L_{n}^{\mathrm{enter}}} q_{\mu,n_{\mathrm{last},\mu}}(k_\mathrm{f})} \ .$$

3.4.2 Ramp metering with route-choice-based DTA

We will now develop the control method for ramp metering that uses the within-day part of the route choice model. We use an MPC-based control structure, see Section 3.2.2. To obtain the prediction model for the MPC-based controller we combine the within-day part of the route choice model of Section 3.4 with the METANET model as described in Section 3.2.2.

As performance indicator for the freeways we use the total time spent (TTS). The TTS is the total time all vehicles spent in the network, and is computed according to (3.13). For local roads, we use the mean density (MD) as performance indicator. The MD is an indicator for the undesired effects of a queue forming on the local road, e.g., pollution and noise. The MD is determined as follows:

$$\mathrm{MD}(k_\mathrm{c}) = \frac{\sum\limits_{(m,i)\in M^{\mathrm{local}}} \sum\limits_{k\in\mathscr{K}_\mathrm{f}(k_\mathrm{c},k_\mathrm{c}+N_\mathrm{p})} \rho_{m,i}(k)}{E_{M^{\mathrm{local}}} + E_{\mathscr{K}_\mathrm{f}(k_\mathrm{c},k_\mathrm{c}+N_\mathrm{p})}}$$

where M^{local} is the set of pairs of indices (m,i) of all links and segments in the local road, and $E_{M^{\mathrm{local}}}$ and $E_{\mathscr{K}_\mathrm{f}(k_\mathrm{c},k_\mathrm{c}+N_\mathrm{p})}$ give the number of elements of respectively M^{local} and $\mathscr{K}_\mathrm{f}(k_\mathrm{c},k_\mathrm{c}+N_\mathrm{p})$.

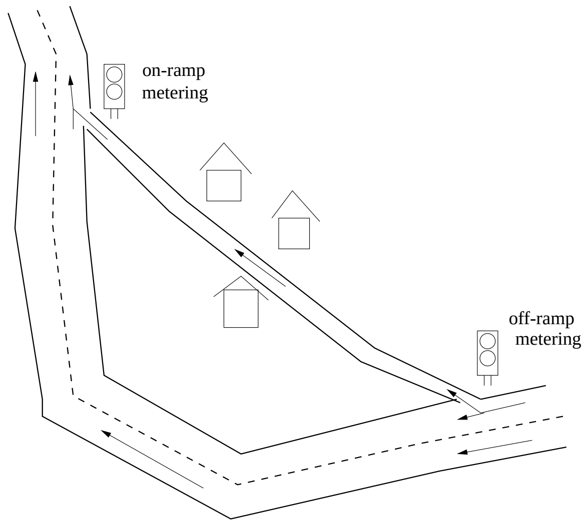


Figure 3.16: Network for the case study with route-choice-based ramp metering control.

We selected these performance criteria because they allow for a good trade-off between the traffic situation on the local road and in the freeway network. Since we compare the systems based on the TTS and the MD, the total performance indicator for the MPC controller is selected as a combination of them:

$$J(k_{\mathrm{c}}) = \alpha_1 \mathrm{TTS} + \alpha_2 \mathrm{MD}$$

where α_1 and α_2 are weighting factors.

3.4.3 Case study

We now illustrate the route-choice-based within-day ramp metering controller with a case study. The network used for the case study consists of two roads: a long two lane freeway of 9 km, and a short-cut through over a local road with a length of 7 km, see Figure 3.16. The beginning and the end of the routes overlap. This means that until segment 4 the routes are equal. After this segment the off-ramp is located. The on-ramp is located at segment 13 of the freeway, and segment 10 of the local road. The last four segments of the two routes overlap again, see Figure 3.17. The demand starts at 4000 veh/h, increases to 8000 veh/h after 45 minutes, and then decreases again to 4000 veh/h after 60 minutes, see Figure 3.18. A period of 100 minutes is simulated. We select the density dependent turning rates as $\beta_{n,m}^{\mathrm{routechoice},\rho^{\mathrm{low}}}(k_{\mathrm{d}}) = 0.3361$, $\beta_{n,m}^{\mathrm{routechoice},\rho^{\mathrm{medium}}}(k_{\mathrm{d}}) = 0.2891$, and $\beta_{n,m}^{\mathrm{routechoice},\rho^{\mathrm{high}}}(k_{\mathrm{d}}) = 0.2080$, which corresponds to an equilibrium traffic assignment when no control is applied. For the cost function we select weighting factors $\alpha_1 = 0.01$ and $\alpha_2 = 1$. The parameters of the METANET model are selected according to [87]: $\rho_{\mathrm{crit},m} = 35$ veh/km/lane, $\rho_{\mathrm{max},m} = 180$ veh/km/lane, $\tau = 18$ s, $\nu = 65$ km^2/h, $\kappa = 40$ veh/km/lane, $a_m = 1.867$, and $T_{\mathrm{f}} = 10$ s. For the freeway link m we use $Q_{\mathrm{cap},m} = 6000$ veh/h, and $v_{\mathrm{free},m} = 120$ km/h, while for the local link μ we have selected $Q_{\mathrm{cap},\mu} = 1000$ veh/h, and $v_{\mathrm{free},\mu} = 50$ km/h.

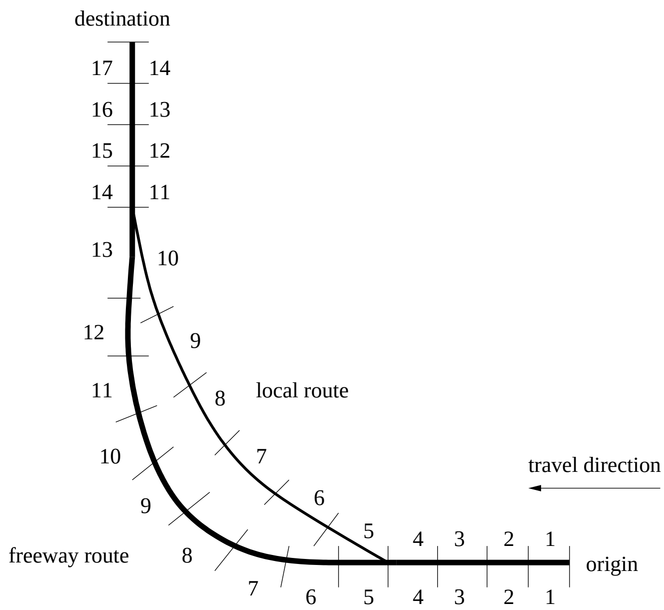


Figure 3.17: Network for route-choice-based ramp metering, including segment numbers for both routes. Note that for both routes the segments are numbered in ascending order. This however means that the freeway segment counters 14 to 17 refer to the same segments as the local segment counters 11 to 14.

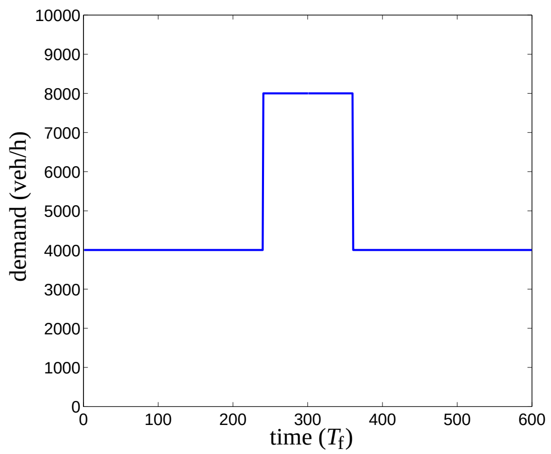
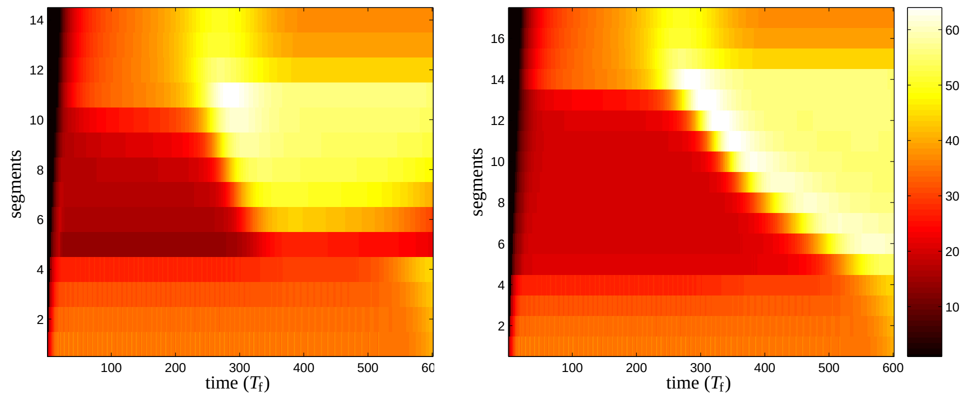


Figure 3.18: Demand for the case study with route-choice-based ramp metering.



(a) Density on the local route (veh/km/lane).

(b) Density on the freeway route (veh/km/lane).

Figure 3.19: Simulation results when no control is applied.

We first simulate the network without control. Then we introduce on-ramp metering, and we compare three different control methods: fixed-time control, ALINEA, and the MPC-based method. Next, we use off-ramp metering, and also compare the three different control methods.

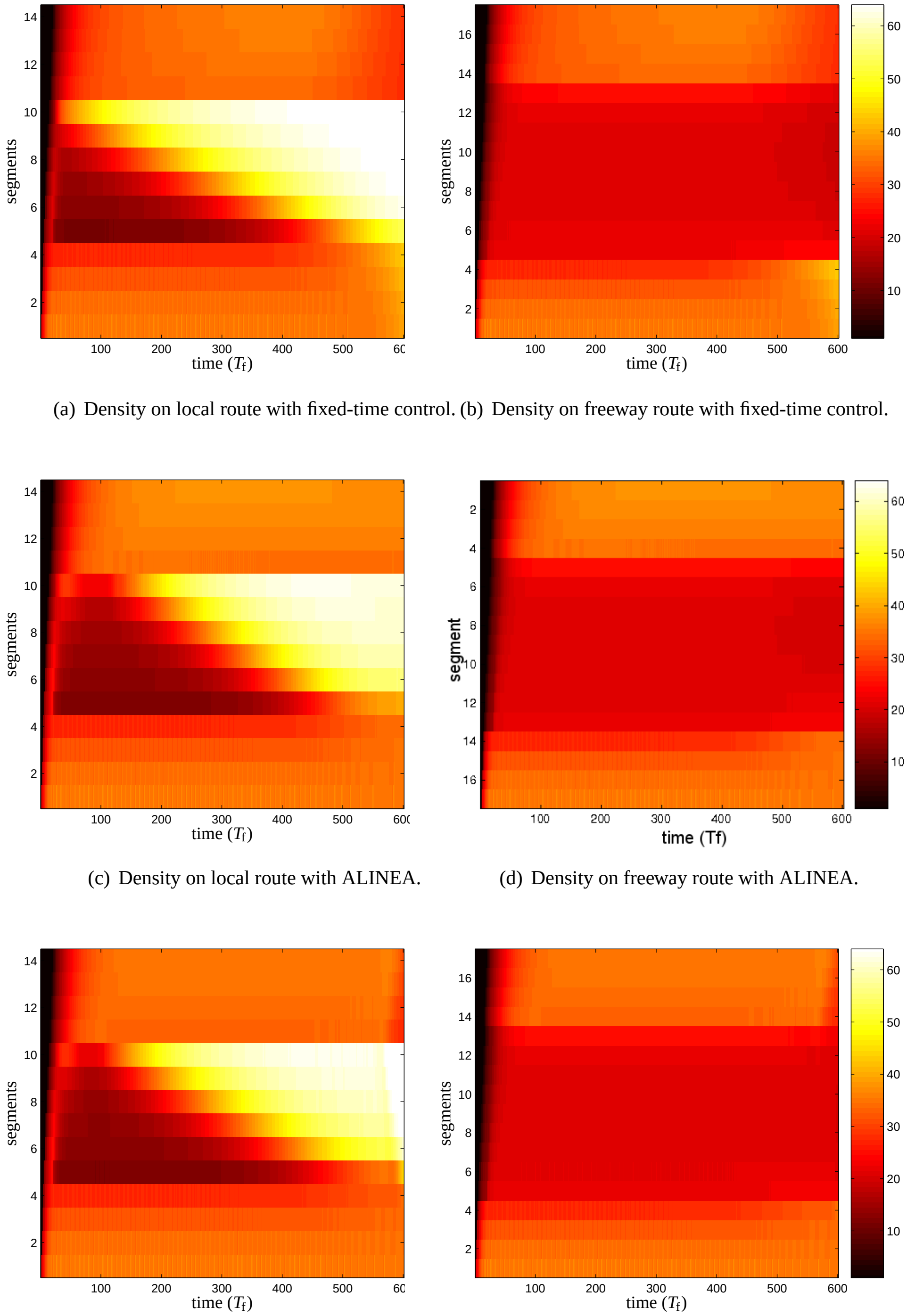
Simulations without control

The first experiment considers no control at all. The results can be seen in Figures 3.19(a) and 3.19(b). Figure 3.19(a) shows the density on the route over the local road, and Figure 3.19(b) shows the density on the freeway route. The time is given at the x-axis, and the color represents the density. The y-axis represents the segments; the vehicles travel from the bottom to the top of the figures. Recall that off-ramp is located after segment 4, and the on-ramp after freeway segment 13 and local segment 10. The congestion starts to appear at the location downstream of the on-ramp, and spills back in the local as well as in the freeway network. For the no control case, the TTS in the network is 11205 veh·h, and the MD is 33.6 veh/km/lane.

Simulations with on-ramp metering

The second experiment is performed using on-ramp metering. We compare fixed-time control, ALINEA, and MPC-based control, which are described in Section 3.2.2.

When fixed-time control is used, the metering rate is set to 0.78, which only limits the flow during the peak in the demand. This value has been determined off-line via an optimization algorithm that optimized the cost function for the given scenario with respect to the selected ramp metering rate. The results of the simulation with fixed-time control are shown in Figures 3.20(a) and 3.20(b). The on-ramp metering prevents the congestion on the freeway, but results in a queue on the local road. The TTS is 10987 veh·h, and the MD is 41.1 veh/km/lane.



(a) Density on local route with fixed-time control.

(b) Density on freeway route with fixed-time control.

(c) Density on local route with ALINEA.

(d) Density on freeway route with ALINEA.

(e) Density on local route with MPC.

(f) Density on freeway route with MPC.

Figure 3.20: Simulation results for on-ramp metering, densities in veh/km/lane.

Figures 3.20(c) and 3.20(d) show the results obtained with the existing on-ramp metering method ALINEA. For the gain K_o 0.015 is selected, after testing different gains between 0.01 and 0.1 for the given scenario. The TTS is 10966 veh·h, and the MD 35.7 veh/km/lane. The ALINEA controller performs better than the fixed-time controller when looking at the MD. The TTS is nearly the same as with fixed-time control.

The results of MPC on-ramp metering are shown in Figures 3.20(e) and 3.20(f). The TTS is 10955 veh·h, and the MD 35.7 veh/km/lane. The MPC based controller improved the TTS in the uncontrolled case with 2.2%, which is slightly better than the results of the ALINEA controller. With respect to the MD the MPC controller performs better than the fixed-time controller, and equal to the ALINEA controller.

An overview of the results is shown in Table 3.1.

Simulations with off-ramp metering

The third experiment uses off-ramp metering, again comparing fixed-time control, ALINEA, and MPC-based control.

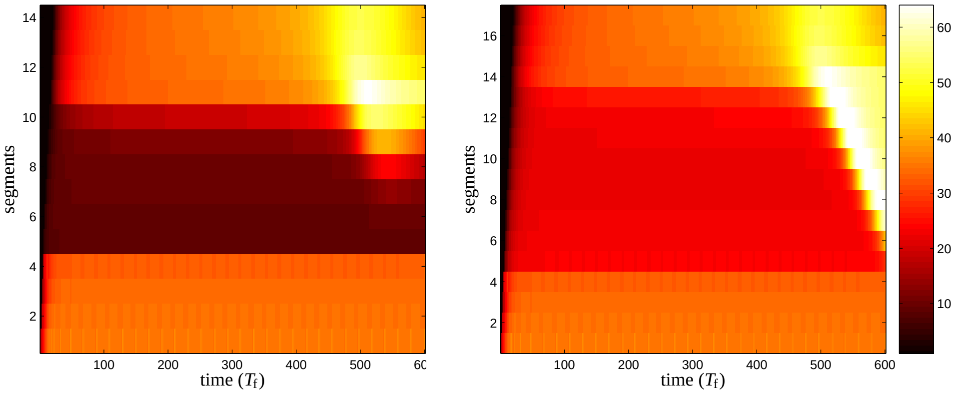
The results of the simulation with fixed-time off-ramp metering are shown in Figures 3.21(a) and 3.21(b). The metering rate is set on 0.45, which is determined by optimizing the performance of the network for a whole day. During the simulation, the congestion is prevented for some time, but appears at the end of the considered period. The TTS is 11065 veh·h, and the MD is 13.8 veh/km/lane. This low value is mainly due to the fact that nearly all traffic is kept on the freeway during the beginning of the simulation, which compensates for the congestion at the end.

Remark 3.4 In this section we use the ALINEA strategy for off-ramp metering, but note that this is not the original purpose of the strategy. We decided to use the density downstream of the on-ramp also as measurement input for the off-ramp, because this is the location where the problems start. A disadvantage of this is the long delay between the control action at the off-ramp and the measurable effects of this action downstream of the on-ramp. This can lead to oscillations in the control signal. These oscillations can be decreased by selecting a smaller K_o, but this makes the controller less effective. A second problem with the distance between the measurement and the ramp metering installation is that the measured density can be influenced by traffic that does not pass the ramp metering installation. When this effect is large, the influence of the ramp metering installation decreases. □

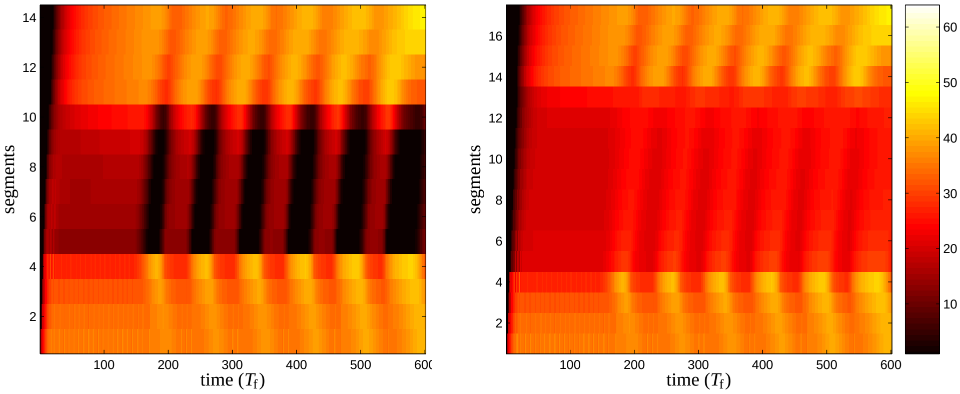
ALINEA for off-ramp metering does not lead to congestion on the local road, see Figures 3.21(c) and 3.21(d). The oscillations due to the delay between action and measuring the effect can be seen clearly. The TTS is 10982 veh·h, which almost equals the result that is obtained with on-ramp metering in the previous experiment. The MD however is lower: 11.0 veh/km/lane.

At last, Figures 3.21(e) and 3.21(f) show the results with MPC-based off-ramp metering. The TTS is 10956 veh·h, and the MD 9.8 veh/km/lane. The TTS for off-ramp metering is nearly the same as for on-ramp metering. The off-ramp metering performs better with respect to the MD. The MD for off-ramp metering is more than a factor three lower than the MD for on-ramp metering.

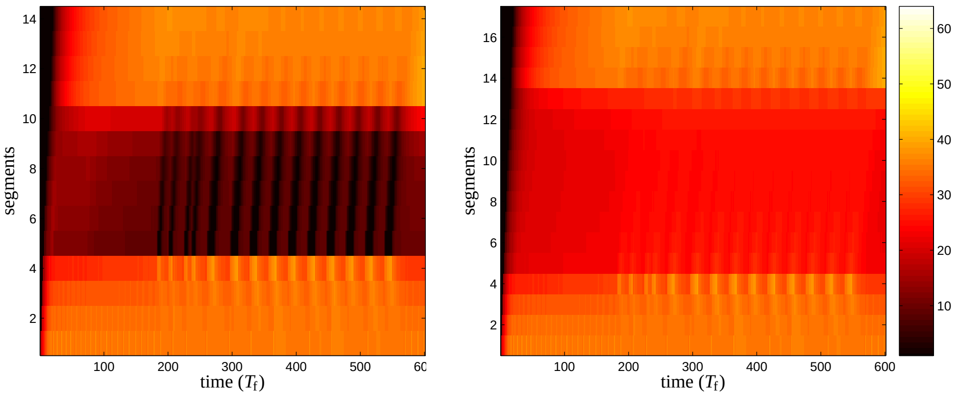
An overview of all results of the case study is presented in Table 3.1. The improvements are specified with respect to the uncontrolled case. The differences between the control



(a) Density on local route with fixed-time control. (b) Density on freeway route with fixed-time control.



(c) Density on local route with ALINEA. (d) Density on freeway route with ALINEA.



(e) Density on the local route with MPC. (f) Density on the freeway route with MPC (veh/km/lane).

Figure 3.21: Simulation results for off-ramp metering, densities in veh/km/lane.

Table 3.1: TTS (veh·h), MD (veh/km/lane), and improvement with the respect to the uncontrolled case (%) for on-ramp and off-ramp metering using various control methods.

	On-ramp				Off-ramp			
	TTS	impr.	MD	impr.	TTS	impr.	MD	impr.
no control	11205	-	33.0	-	11205	-	33.0	-
fixed time	10987	-2.0	41.1	-22.3	11065	1.2	13.8	58.9
ALINEA	10966	-2.1	35.7	-6.2	10982	2.0	11.0	67.2
MPC	10955	-2.2	35.7	-6.2	10956	2.2	9.8	70.8

approaches are small, only the MD for fixed-time control for on-ramp metering differs from the MDs obtained with the other methods. For all off-ramp control methods the MD is a factor three lower than for on-ramp control. The MPC controller results in the lowest total time spent, without causing high densities on the local road.

3.5 Integrated control of information providing and speed limits

The third controller that we develop uses the within-day as well as the day-to-day part of the route choice model developed in Section 3.4. Further, whereas the previously developed first and second control strategies only take changes in the route choice into account, the strategy that we develop in this section tries to actively steer the route choice. This allows for a more efficient use of the available roads, and can improve the safety by reducing flows on dangerous roads. The possible effects of steering the route choice are illustrated in Appendix A, where we show that influencing the route choice with traffic control measures using basic control methods already can improve the network performance, which encourages the development of more advanced control methods for steering the route choice.

To influence the route choice, different measures are available. There are ‘hard measures’ like traffic signals [58], speed limits [2, 64], and ramp metering signals [123], which the drivers have to obey. These hard measures however have only an indirect influence on the route choice, via the travel time. On the other hand, ‘soft measures’ are available, to which the drivers can comply or not. Providing information is such a ‘soft measure’. Although drivers are not forced to react on the information, providing this information can nevertheless be an effective measure to improve the network performance. The information is often provided via dynamic route information panels (DRIPs). The effect of displaying information on drivers is described in, e.g., [47, 90]. The information that is displayed on the DRIPs can consist of, e.g., queue lengths, travel times, or route advises. Which information should be presented is a subject of ongoing research and is discussed in, e.g., [18].

Providing the information can have two goals: informing the drivers about what they can expect, and trying to influence the route choice of the drivers. We target the second goal, i.e., we want to influence the route choice of the drivers. Research in which DRIPs are used to control the traffic is described in [47, 81]. In these papers methods are described in which DRIPs are combined with other control measures to influence the route choice of the

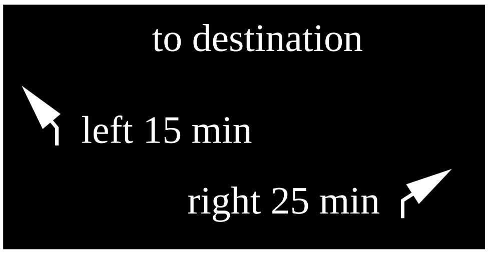


Figure 3.22: Dynamic route information panel displaying travel times.

drivers, based on predictions made with METANET and a DTA algorithm. In this section, we develop a control approach that integrates DRIPs and variable speed limits to influence the route choice. In this way it is possible to approximate a *desired traffic assignment*.

To determine which desired traffic assignments are useful and can be obtained in practice, traffic management bodies have started to explore the possibilities of changing the traffic assignment, and formulated a procedure to select the most desired assignment, as described by [110]. The procedure goes as follows. First preferred routes towards special destinations like, e.g., the city center, a main business building, or recreational areas are selected. Then a traffic assignment with high flows on these preferred routes and low flows on other routes is defined. In this way drivers with those specific destinations are encouraged to stay out of residential and/or industrial areas. The defined traffic assignment is then called the desired traffic assignment according to the traffic management bodies.

In the remainder of this section we first model the reaction of the drivers on the information that is provided on DRIPs. Then we describe the integrated control method, and illustrate its performance with a case study.

3.5.1 Reaction on information

The controller that we develop influences the route choice of the drivers by providing information. The provided information consists of travel times for different routes in the network, and it is displayed on a DRIP, as shown in Figure 3.22.

The provided information influences the within-day route choice of the drivers. How many drivers change their route based on the provided information depends on the difference in the expected travel time of their preferred route r and the displayed travel time on the other route p, and on the number of drivers that can be influenced to change their route. Below we first formulate the likelihood that drivers will change their route, and then we use this likelihood to compute the resulting turning rates.

The model developed by [50] is used as a model to describe the likelihood that drivers will change their route based on the difference in travel times. The model computes the likelihood ($l_{n,r,p}$) that drivers at node n, with a preference for route r, will change their preferred route into route p, according to:

$$l_{n,r,p}(k_\mathrm{f}) = \begin{cases} 1-\exp\left(\theta_{n,r}(k_\mathrm{d})(J_p(k_\mathrm{f})-J_r(k_\mathrm{f}))\right) & \text{if } J_p(k_\mathrm{f}) < J_r(k_\mathrm{f}) \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

where $J_p(k_\mathrm{f})$ is the cost of route p and $J_r(k_\mathrm{f})$ the cost of route r, as displayed on the DRIP, in this case the travel times as computed with (3.11), and $\theta_{n,r}(k_\mathrm{d}) \in [0,1]$ represents the fraction of traffic on route r that can be influenced by the provided information on node n at day k_d. We assume that this fraction of traffic depends on the correctness $\xi_{n,r}(k_\mathrm{d})$ of the displayed travel times, and on the fraction of the drivers $\theta^0_{n,r}$ that can be influenced by the provided information. The fraction $\theta^0_{n,r}$ is included since not all traffic will be able to react on the information. This can be caused by the fact that the drivers do not want to go to the specified destination, by a lack of knowledge about the available routes, or by the need for an intermediate stop at, e.g., a service station, or car pool location. The final fraction of the drivers that can be influenced is then given by:

$$\theta_{n,r}(k_\mathrm{d}) = \theta^0_{n,r}\xi_{n,r}(k_\mathrm{d})$$

where $\xi_{n,r}(k_\mathrm{d})$ characterizes the correctness of the information as experienced by the drivers. It is determined by the percentage of trips that the displayed travel times were sufficiently close to the travel times that were experienced. To determine the value of $\xi_{n,r}(k_\mathrm{d})$ the travel times that were experienced are compared with those displayed on the DRIP, using a margin η_info depending on the length of the routes, which allows for small differences that do not influence the perception of the drivers. So the travel times are assumed to be correct if:

$$\tau_{n,r}(k_\mathrm{f}) - \eta_\mathrm{info} < \tau^\mathrm{info}_{n,r}(k_\mathrm{f}) < \tau_{n,r}(k_\mathrm{f}) + \eta_\mathrm{info} \tag{3.18}$$

where $\tau_{n,r}(k_\mathrm{f})$ is the experienced travel time on node n on route r for drivers that reach the splitting node at time k_f, $\tau^\mathrm{info}_{n,r}(k_\mathrm{f})$ is the displayed travel time at node n at this time, and η_info is the margin used. The value $\xi^\mathrm{experienced}_{n,r}(k_\mathrm{d})$ is computed as follows:

$$\xi^\mathrm{experienced}_{n,r}(k_\mathrm{d}) = \phi^\mathrm{true}_{n,r} / (\phi^\mathrm{true}_{n,r} + \phi^\mathrm{false}_{n,r})$$

where the counters $\phi^\mathrm{true}_{n,r}$ and $\phi^\mathrm{false}_{n,r}$ respectively count the times that the information is correct in the sense of (3.18). The value for $\xi_{n,r}(k_\mathrm{d})$ is then updated similar as the estimated travel time in (3.14):

$$\xi^\mathrm{est}_{n,r}(k_\mathrm{d}+1) = \omega \xi^\mathrm{experienced}_{n,r}(k_\mathrm{d}) + (1-\omega)\xi^\mathrm{est}_{n,r}(k_\mathrm{d}) \tag{3.19}$$

where $\xi^\mathrm{est}_{n,r}(k_\mathrm{d}+1)$ is the estimated correctness value for the next day.

Now we again consider a network with one node and two possible routes, as in Section 3.4.1, and determine the factors $\theta_{1,1}(k_\mathrm{d})$ and $\theta_{1,2}(k_\mathrm{d})$, and the corresponding likelihoods $l_{1,1,2}(k_\mathrm{f})$ and $l_{1,2,1}(k_\mathrm{f})$. The likelihoods are used to adapt the turning rates toward route 1 according to:

$$\beta^\mathrm{information}_{1,1}(k_\mathrm{f}) = \beta^\mathrm{routechoice}_{1,1}(k_\mathrm{f}) - l_{1,1,2}(k_\mathrm{f})\beta^\mathrm{routechoice}_{1,1}(k_\mathrm{f}) + l_{1,2,1}(k_\mathrm{f})\beta^\mathrm{routechoice}_{1,2}(k_\mathrm{f}) \tag{3.20}$$

where $\beta^\mathrm{information}_{1,1}(k_\mathrm{f})$ is the turning rate toward route 1 based on information and route choice, $\beta^\mathrm{routechoice}_{1,1}(k_\mathrm{f})$ and $\beta^\mathrm{routechoice}_{1,2}(k_\mathrm{f})$ the turning rates toward route 1 and 2 resulting from (3.16).

Remark 3.5 For a network where the routes do not overlap upstream of the last internal node, (3.20) can be rewritten as:

$$\beta_{n,m}^{\text{information}}(k_{\text{f}}) = \beta_{n,m}^{\text{routechoice}}(k_{\text{f}}) - \sum_{\psi \in R_{n,m}^{\text{link}}} l_{n,m,\psi}(k_{\text{f}}) \beta_{n,m}^{\text{routechoice}}(k_{\text{f}}) + \sum_{\psi \in \{\psi | m \in R_{n,\psi}^{\text{link}}\}} l_{n,\psi,m}(k_{\text{f}}) \beta_{n,\psi}^{\text{routechoice}}(k_{\text{f}})$$

For a general network with overlapping routes it is however not possible to compute $\beta_{n,m}^{\text{information}}(k_{\text{f}})$, since when $\beta_{n,m}(k_{\text{f}})$ for the METANET model is determined, the information about the route-dependent turning rates gets lost due to the fact that $\beta_{n,m}(k_{\text{f}})$ is valid for all the traffic arriving at node n, independent of the route r. To solve this problem the destination-oriented version of the METANET model should be used. □

3.5.2 Integrated route choice controller

We will now describe the route choice controller that integrates information providing on DRIPs with variable speed limits. The controller uses MPC with as prediction models the METANET model, the route choice model of Section 3.4, and the model for the reaction on information of Section 3.5.1.

As performance indicator we consider the total time spent (TTS) by all vehicles in network, but note that the proposed approach also works for other performance indicators. The TTS is computed according to (3.13). Since the reliability of the provided travel time information influences the effect of the control actions via the parameter $\theta(k_{\text{d}})$, see (3.17), a penalty is added to prevent displaying incorrect travel times. This penalty consists of the difference between the displayed travel times and the experienced travel times. The penalty term is included in the performance indicator as follows:

$$J(k_{\text{c}}) = \text{TTS}(k_{\text{c}}) + \chi_1 \sum_{n \in N} \sum_{r \in R} \sum_{k_{\text{f}} \in \mathscr{K}_{\text{f}}(k_{\text{c}}, k_{\text{c}}+N_{\text{p}})} \left(\bar{\tau}_{n,r}(k_{\text{f}}) - \bar{\tau}_{n,r}^{\text{info}}(k_{\text{f}}) \right)_2^2$$

where N is the set of all nodes in the network, R is the set of all routes in the network, $\bar{\tau}_{n,r}$ the average real experienced travel time on route r from node n on during control period k_{c}, and $\bar{\tau}_{n,r}^{\text{info}}(k_{\text{c}})$ the average travel time displayed on the DRIP at node n for route r.

Since the MPC method can handle hard constraints minimum and maximum values the control signal can be taken into account. Also a maximum length for, e.g., the queues at the origins can be guaranteed.

3.5.3 Case study

We will use a simple case study to illustrate the effects of the integrated route choice control method. In the remainder of this section we will first describe the network and traffic scenario that are used for the case study, and next we describe the simulation results for simulations without control, for simulations with the systems that are currently used in The Netherlands, and for simulations with the MPC-based method developed in this section.

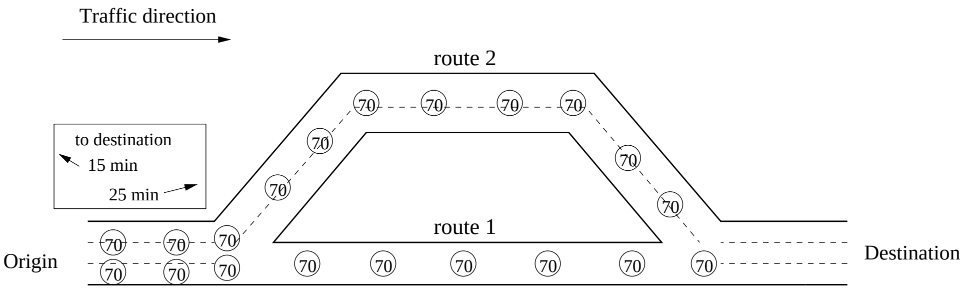


Figure 3.23: Layout of the network.

Network and traffic scenario

For the case study we have selected a network with two possible routes, see Figure 3.23. A three-lane freeway of 2 km splits into a one-lane freeway of 3 km and a two-lane freeway of 5 km, so the two-lane freeway is longer than the one-lane freeway. Later the two freeways join each other again in a three-lane freeway of 2 km. The shortest route with one lane is route 1, the longest route with two lanes is route 2. Since the traveled distance on route 1 is shorter, this route is preferred by the drivers. A disadvantage of route 1 is that its capacity is low due to the fact that there is only one lane. This means that the route will get congested fast when the demand is increasing. This makes route 2 attractive when the demand (and thus the density before the splitting node) is high. The DRIP is located upstream of the splitting node, and shows travel times for both routes. Speed limits are applied at every 500 m on both routes.

The traffic scenario is chosen such that all three density groups appear, and that the highest density results in a congestion at the shortest route. The traffic demand varies in discrete steps starting at 2000 veh/h, increasing to 4000 veh/h after 20 minutes, to 8500 veh/h after 40 minutes, and decreases back to 4000 veh/h after 1 hour, and to 2000 veh/h after 80 minutes, see Figure 3.24.

The parameters of the METANET model are selected according to [87]: $v_{\text{free},m} = 120$ km/h, $\rho_{\text{crit},m} = 35$ veh/km/lane, $\rho_{\text{max},m} = 180$ veh/km/lane, $Q_{\text{cap},} = 6000$ veh/h, $\tau = 18$ s, $\nu = 65$ km^2/h, $\kappa = 40$ veh/km/lane, $a_m = 1.867$, and $T_\text{f} = 10$ s. We simplify the route choice algorithm in such a way that it reduces the number of counters that should be updated, and thus saves available memory and computation time. We do this by making the assumption that the drivers in our simulation do not become habitual drivers that select always the same route independent of the cost of this route. For the case study this can be justified by the fact that the simulated costs of a route are solely based on the experienced travel time, and thus not includes the development of other preferences of the drivers. This means that the counters $\phi^{\text{est}}_{S(1),R(2)}(k_\text{d})$ and $\phi^{\text{est}}_{S(2),R(1)}(k_\text{d})$ will stay zero during the simulation. Since in this case the probability that the shortest route is also the route that is selected is very large, we set this probability to one: $P(S(1)|R(1)) = 1$. In the same way we assume that selecting the longest route is not very likely, so we select $P(S(2)|R(1)) = 0$. With the assumed values, (3.15) can be simplified to:

$$P(S(1)|\Delta) = P(R(1)|\Delta) + P(S(1)|\text{EQ})P(\text{EQ}|\Delta)$$

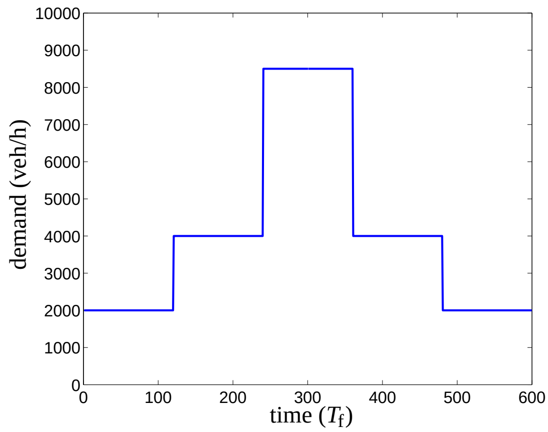


Figure 3.24: Demand for the case study on integrated control with travel time information.

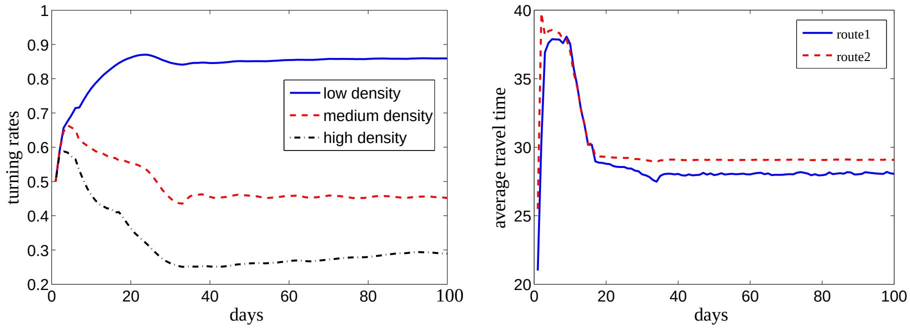
Note however that in reality a significant part of the drivers has developed a habit [17], and thus in reality this simplification is not valid.

The control signal that we use for the case study consists of the values for the speed limits and the values for the travel times displayed on the DRIP. For the MPC controller we have selected $N_p = N_c = 10$ and $T_c = 1$ minute. The within-day route choice is considered in the model, but since the day-to-day learning is an important aspect of the model, we investigate the day-to-day behavior and repeat each of the simulations for 100 successive days.

During the case study we will compare three different situations. In the first simulation no control is applied, in the second simulation the current situation in The Netherlands is investigated, and during the last simulation the developed MPC-based controller is applied. The system that is currently used in the Netherlands id not designed to influence the route choice of the drivers, but to increase the safety on the road and to inform the drivers. This means that we do not perform a comparison between to control methods, but that we illustrate improvements that can be obtained by introducing a control method.

No control

The first simulation examines the performance of the network without control. Figure 3.25(a) shows the evolution of the preferred turning rate toward route 1, for each of the three density groups, from day 1 to day 100. As starting value for all turning rates we have selected 0.5. The solid line represents the turning rates for low-density situations. At low densities it is logical that the shorter but smaller route is the fastest, and indeed, the turning rate toward route 1 increases over the days. The dashed line represents the turning rate for medium-density conditions. As the density on the shortest route is a little higher, the travel time on this route increases. Both routes have nearly the same travel time, resulting in an equilibrium turning rate of 0.45. Finally, the dash-dotted line shows the turning rates for high-density conditions. Route 1 is congested, and so more drivers select the second route



(a) Evolution of the turning rates toward route 1.

(b) Evolution of the average travel times.

Figure 3.25: Simulation results when no control is applied.

and the turning rate towards route 1 decreases.

Figure 3.25(b) shows the average travel time for each day. Initially, while the drivers are still learning, the travel times on both routes are high. After 17 days the drivers have selected the best routes, and the travel times reach a more or less stable value. The two average travel times are not exactly equal; this is because the drivers do not change their route when the difference in travel times is small. The TTS in the network is 761.1 veh·h on the first, and 546.7 veh·h on the last day.

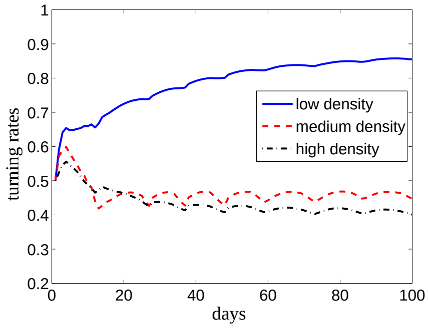
Instantaneous travel times and incident detection

The second simulation shows the situation that is currently present in The Netherlands. Here, the available measures are not used to control the traffic, but to inform the drivers, and to prevent head-tail collisions when congestion is present. This means that the comparison with the control method developed in this chapter is not a totally fair comparison, since the purposes of the systems differ. However, an indication of the improvements that are possible by applying integrated control can be given.

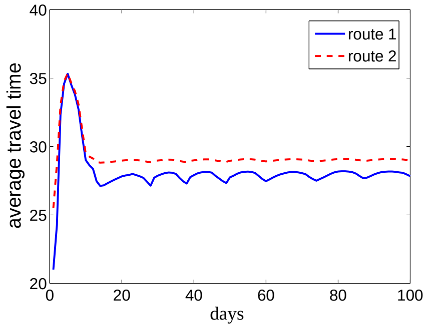
In The Netherlands, the DRIPs display instantaneous travel times, mainly to inform the drivers. The variable speed limits are used as incident detection system, which works as follows. When the speed in a segment drops below 40 km/h the speed limit in this segment is set to 50 km/h and the speed limit in the upstream segment to 70 km/h. When the speed increases above 50 km/h, the speed limits are deactivated.

In Figure 3.26(b) it can be seen that with this method the constant values of the travel times are reached within 12 days, and the peak in the average travel time stays lower than in the no control case.

The TTS on the first day is 512.3 veh·h, and on the last day it is 542.1 veh·h. The increase in the total time spent is caused by the fact that many drivers react on the provided information on the first day, but that they do not experience the same travel time as presented on the DRIP, which lowers the value $\xi(k_\mathrm{d})$ for the correctness of the provided information. As a result more drivers start to neglect the information during later days, which finally leads to nearly the same traffic assignment as without control.

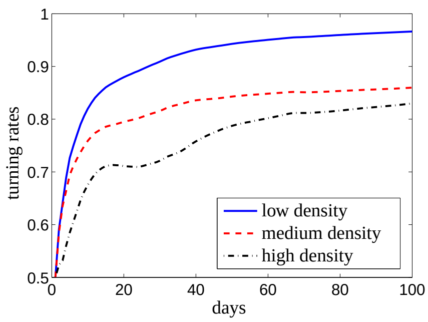


(a) Turning rates for the current situation in the Netherlands.

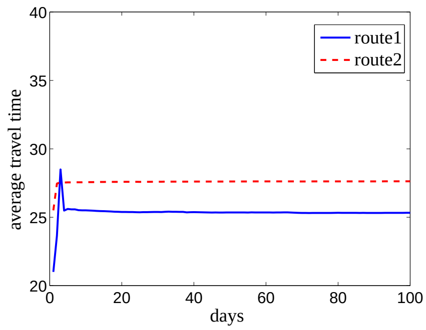


(b) Travel times for the current situation in the Netherlands.

Figure 3.26: Simulation results with instantaneous travel times and incident detection.



(a) Turning rates using MPC.



(b) Average travel times using MPC.

Figure 3.27: Simulation results when MPC is applied.

Integrated route choice control

The last simulation presents the effect of the MPC method developed in this section. Figures 3.27(b) and 3.27(a) show the results.

The TTS on the last day is 533.5 veh·h, which is only an improvement of 3% compared to the no control case, and of 2% to the current situation in The Netherlands. A contribution of the developed MPC based method can be seen in the first 10 days. The high peak in average travel times is prevented, showing that the drivers learn faster what the best route is. It can also be seen that during the next days a more stable situation occurs, and that there are less fluctuations in travel times, making the routes more reliable. A side effect is that more drivers use the first route, which is shorter in distance and thus leads to less vehicle kilometers. The performance of the MPC-based controller can still be improved by, e.g., using a longer prediction horizon, performing more optimization runs with different initial values, and selecting a more suitable optimization algorithm.

3.6 Conclusions

We have investigated three traffic flow control strategies that take the change in the traffic assignment caused by the control actions into account. All control strategies are based on model predictive control (MPC). The model METANET is used as prediction model to predict the evolution of the traffic flows. The control strategies use different routing models and control measures.

The first control strategy considered within-day route choice, and used equilibrium-based dynamic traffic assignment (DTA) to describe how the flows are divided over the network. Within the control method, the DTA algorithm has been used to anticipate on the change in route choice induced be the actions of the controller. The performance of this controller has been illustrated with a case study on ramp metering.

The second control strategy used an en-route route-choice-based assignment model for within-day route choice to determine the traffic assignment, and also used the DTA algorithm to anticipate on changes in the route choice. In a case study we considered on-ramp as well as off-ramp metering. With on-ramp metering the total time spent has been reduced compared to the no control case, while the mean density stayed nearly equal. With off-ramp metering the mean density has been reduced without increasing the total time spent.

Third, we have developed an integrated control method that uses an en-route route-choice-based DTA algorithm, and that on purpose influences the route choice of the drivers. The control method considered within-day as well as day-to-day route choice. In a case study we have illustrated the performance of the route choice control method.

Topics for future research are: investigation of the trade-off between accuracy and computational complexity for the different methods, investigating other DTA algorithms, investigating the use of destination-dependent models, calibration and validation of the developed models, and robustness tests of the controller. Further the concept of off-ramp metering should be investigated, and some attention should be paid to the legal aspects of displaying not-yet-realized travel times. Finally, more extensive case studies should be performed to further investigate and compare the performance of the controllers.

3.A List of symbols

Timing

k_{f}	simulation time step counter
k_{c}	control time step counter
k_{d}	day counter
k_{a}	assignment update time step counter
T_{f}	simulation time step (h)
T_{c}	controller time step (h)
T_{a}	assignment update time step for the DTA model (h)
$\mathscr{K}_{\mathrm{f}}(k_{\mathrm{c}}^{a}, k_{\mathrm{c}}^{b})$	set of simulation steps k_{f} that correspond to the time interval $[k_{\mathrm{c}}^{a} T_{\mathrm{c}}, k_{\mathrm{c}}^{b} T_{\mathrm{c}}]$

Sets

O	set of all origins in the network
L_n^{enter}	set of all freeway links entering node n
L_n^{leave}	set of all freeway links leaving node n
N	set of all nodes in the network
R	set of all routes in the network
R_n^{node}	set of routes that pass through node n
$R_{n,m}^{\text{link}}$	set of routes that passes through node n toward freeway link m
M	set of pairs of indices (m,i) of all links of the network
M_r^{link}	set of pairs of indices (m,i) of all links and segments belonging to route r
M_r^{node}	set of pairs of indices (n,m) belonging to route r
M^{urban}	set of pairs of indices (m,i) of all links and segments belonging to the local roads in the network

Control

$r_o(k_\text{f})$	ramp metering rate at ramp o at simulation time step k_f
$v_{m,\text{control}}(k_\text{f})$	value of the speed limit at freeway link m at simulation time step k_f (km/h)
$\text{MD}(k_\text{f})$	mean density on the local road at simulation time step k_f (veh/km/lane)
K_o	gain of ALINEA ramp metering installation on ramp o
$J(k_\text{c})$	performance indicator at control time step k_c for the period $[k_\text{c}T_\text{c},(k_\text{c}+N_\text{p})T_\text{c})$
N_p	prediction horizon (control time steps)
$\text{TTS}(k_\text{c})$	total time spent in the network during simulation period $[k_\text{c}T_\text{c},(k_\text{c}+N_\text{p})T_\text{c})$ (veh·h)

Metanet

$q_{m,i}(k_\text{f})$	outflow of segment i of freeway link m during simulation time step k_f (veh/h)
$\rho_{m,i}(k_\text{f})$	density on segment i of freeway link m at simulation step k_f (veh/km/lane)
$v_{m,i}(k_\text{f})$	mean speed on segment i of freeway link m at simulation time step k_f (km/h)
$V(\rho_{m,i}(k_\text{f}))$	desired speed at segment i of freeway link m at simulation time step k_f
$w_o(k_\text{f})$	queue length at on-ramp o at simulation time step k_f (veh)
$\tau_r(k_\text{f})$	travel time on route r for a vehicle starting at simulation time step k_f (h)
$\beta_{n,m}(k_\text{f})$	turning rates toward freeway link m on node n

Equilibrium based model

$q_{r,j}^{\text{MSA}}(k_\text{f})$	flow on route r during iteration step j of the MSA algorithm performed at simulation time step k_f (veh/h)

$q^{\mathrm{AON}}_{r,j}(k_{\mathrm{f}})$	flow on route r determined by the AON assignment after iteration step j of the MSA algorithm performed at simulation time step k_{f} (veh/h)
$\bar{\tau}^{\mathrm{est}}_{r}(k_{\mathrm{d}})$	estimated average travel time at simulation day k_{d} (h)
$\bar{\tau}^{\mathrm{experienced}}_{r}(k_{\mathrm{d}})$	average experienced travel time during day k_{d} (h)
$\beta^{\mathrm{MSA}^*}_{n,m}(k_{\mathrm{f}})$	turning rates toward freeway link m at node n computed with the MSA algorithm at simulation time step k_{f}

Route-choice-based model

$R(r)$	route r has the lowest travel time
EQ	the travel times on both route is approximately equal
$S(r)$	statement: route r is selected
$P(S(r)\|\Delta)(k_{\mathrm{d}})$	probability that route r is selected when the density is in density group Δ at simulation day k_{d}
$\phi^{\mathrm{experienced}}_{A,B}(k_{\mathrm{d}})$	counter of the number of times that the combination of A and B is experienced during day k_{d}
$\phi^{\mathrm{est}}_{A,B}(k_{\mathrm{d}})$	estimation of the number of times that the combination of A and B will occur during day k_{d}
$\beta^{\mathrm{routechoice}}_{n,m}(k_{\mathrm{f}})$	turning rates toward freeway link m at node n computed with the route choice model

Reaction to information

$\tau^{\mathrm{info}}_{n,r}(k_{\mathrm{f}})$	displayed travel time for route r on node n at simulation time step k_{f} (h)
$l_{n,r,p}(k_{\mathrm{f}})$	likelihood of traffic on route r selecting route p at node n at simulation time step k_{f}
$\theta_{n,r}(k_{\mathrm{d}})$	fraction of the vehicles on route r at node n that can be influenced by the provided information at day k_{d}
$\xi_{n,r}(k_{\mathrm{d}})$	correctness of the displayed trave times for route r on node n as experienced by the drivers at day k_{d}
$\beta^{\mathrm{information}}_{n,r}(k_{\mathrm{f}})$	turning rates toward route r at node n including the reaction on information at day k_{d}

Chapter 4

Model-based control of day-to-day route choice in traffic networks

In this chapter we develop a model-based control approach for day-to-day route choice control. Therefore we formulate a simplified route choice model that can be used to obtain fast predictions of the route choice behavior and that is suitable for obtaining a first impression of the traffic assignment, for use in on-line optimization algorithms, or as initial value for more complex optimization algorithms. We use this model in a model-based control setting, which uses prediction and an optimization procedure to obtain optimal values for control measures. In particular, we investigate speed limit control and outflow control. The objective of the controller is to influence the route choice of the drivers such that a predefined cost function is optimized. We illustrate the possibilities of the control approach with an example based on the Braess paradox.

4.1 Introduction

As the number of vehicles and the need for transportation grow, cities around the world face serious traffic congestion problems: almost every weekday morning and evening during rush hours the saturation point of the main roads is attained. This often causes drivers to divert to minor roads, causing large flows in residential areas, near primary schools, or near shopping centers. These large flows can lead to undesired or unsafe situations. They also cause pollution and noise, which does not only affect humans living near the roads, but can also have a negative impact on humans living further away, and on nature reserves due to dispersion.

The location of the large traffic flows and the corresponding congestion is the result of the route choices of the drivers. Road administrators can try to change the route choices, in order to prevent or to relocate unwanted large traffic flows or to reduce the travel times. In this paper we develop a control method that supports the administrators by influencing the route choices of drivers, using existing traffic control measures, such as variable speed limits, or mainstream metering. These measures influence the travel times on different routes, and hence induce a change in route choice of the drivers.

Let us now consider the route choice process of drivers. With respect to the time scale of the route choice, there exists within-day route choice and day-to-day route choice. Within-day route choice [18, 27] describes the route choice of drivers during their trip. This means that drivers change their route during their trip based on the current state of the traffic network. Day-to-day route choice [28, 101, 179] focuses on the changes in route choice from one day to the next. This means that the route choice of the drivers for the current day is based on experiences of the previous days. The drivers consider these experiences and weigh them, leading to a preferred route for the next day. In this chapter we only consider day-to-day route choice.

Day-to-day route choices of drivers are modeled with (dynamic) traffic assignment (DTA) models [8, 13, 15, 42, 130]. These models describe how the traffic flows divide themselves over the network. An overview of DTA models is presented in [131], where different classes of formulations of the DTA problem are given, and where some directions of future developments are presented. With respect to the use of DTA models in traffic controllers, we can divide the models into two main categories: equilibrium-based models and route-choice-based models, see Figure 4.1. Equilibrium-based models assume that an equilibrium traffic assignment[1] will appear in which no driver can change his route without increasing his costs [182]. The use of this assumption means that for every simulation run within the controller the equilibrium should be computed, which leads to large computation times and thus makes the models less suitable for on-line use. En-route route-choice-based models [18, 37, 103] do not explicitly assume an equilibrium, but use en-route route-choice models which determine the route choice based on the current state of the network and from the experiences of the drivers during previous days. Using these models the reasoning of drivers with respect to the route choice can be captured in a natural way, and it is not necessary to determine the expected travel times on the whole network to determine the turning rates at one intersection. Advantages of the en-route route-choice-based models[2] are the low computation time, and the applicability for the situations where no equilibrium assignment might appear. This makes the models suitable for the use in the route choice controller that we will develop in this chapter.

For our controller, we develop a basic model that is suitable to obtain a first impression of the evolution of the turning rates for planning purposes and especially for the use within an on-line traffic control framework where the model can be used to obtain fast predictions or good initial solutions for non-linear optimization procedures which use more accurate models.

We develop the model in three steps with a gradually increasing complexity:

Case A: constant demand, separate routes, single origin and destination,

Case B: piecewise constant demand, separate routes, single origin and destination,

[1] Note that for the considered categorization, it is not important whether the considered equilibrium can be deterministic, stochastic, static, or dynamic.

[2] In the remainder of this chapter we will use the shorter term ’route-choice-based models’ instead of ’en-route route-choice-based models’ for ease of notation.

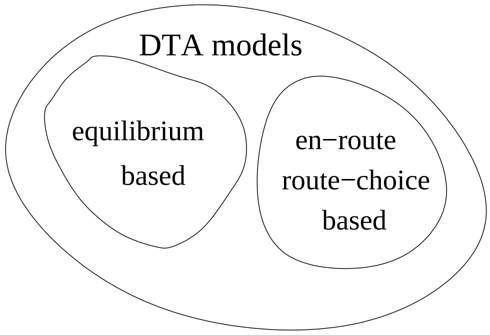


Figure 4.1: Different types of DTA models.

Case C: piecewise constant demand, overlapping routes, multiple origins and destinations, and restricted link inflow capacities.

The model of Case **A** allows for analytical descriptions and for intuitive explanations due to its simplicity. The model requires only a small amount of computation time to simulate a network and can thus be used to obtain a first impression of the evolution of the turning rates. Case **B** uses time-varying piecewise constant demands to get a better approximation of the flows in the network. At last, Case **C** introduces overlapping routes, restricted link inflow capacities, and networks with multiple origins and destinations, and thus allows for the modeling of general traffic networks.

Within the route choice model, we assume that the main factor in the route choice of the drivers is the travel time [18]. Note that in reality the route choice of drivers depends on many factors, such as road type, travel time reliability, surroundings, travel distance, and personal preferences [18, 28, 147]. Different factors can be easily implemented for Case **C**, while for Cases **A** and **B** only factors that can be formulated as linear functions of the model variables can be used, e.g., the travel distance or average queue length on the route that is used. In this chapter we will use the travel times, which can be computed with detailed models as described in, e.g., [7, 40, 106, 137], or with less detailed models presented in, e.g., [15, 26, 33, 54]. These models still lead to large computation times. Therefore, in this chapter we will use a somewhat less detailed model, with low accuracy, which results in low computational effort. We use the mean travel times over the whole day to determine the route choice. We assume that the mean experienced travel times are known by all drivers, which means that all the drivers are completely informed (e.g., via travel information services), so they know the travel times on both routes, independently of the route they have selected.

To control the route choice of the drivers existing traffic control measures can be used. In [63, 150] it has already been shown that traffic control measures that do not directly influence route choice but that do have an impact on the travel time (such as traffic signals, variable speed limits, and ramp metering) can be used for this purpose. This has led to the theoretical development of methods to incorporate the effect of existing traffic control measures on route choice, see [11, 81, 177]. The control methods described in these papers use a prediction model that describes the evolution of the traffic flows as well as the evolution of the turning rates, and then use existing traffic control measures to influence the turning rates. The methods differ with respect to the control algorithm that is applied, and with respect to the models that are used. A specific control method that is suitable for route choice control is Model Predictive Control (MPC) [25, 100]. This is a model-based control approach that uses a prediction model in combination with an optimization algorithm to determine optimal values for the traffic control measures. The optimal values are then applied using a rolling horizon approach. MPC has been used earlier to influence the route choice *within a day* in [10, 64, 89, 130]. Within this chapter we use MPC for *day-to-day* route choice.

We develop a high-level controller that can influence the speed and outflow of links. Lower-level controllers should be used to translate the values for maximum speeds and outflows into settings for the traffic control measures. The state of the system includes the turning rates. With the prediction model we can determine the influence of the control actions on the turning rates, and the control actions that can improve the performance of the network.

The MPC-based control approach includes solving an optimization problem. We address the three different cases using different optimization methods. The model of Case **A** is a piecewise affine model. When this model is combined with piecewise affine control objectives and piecewise affine constraints, the optimization problem can be reformulated as a mixed integer linear programming (MILP) problem. For MILP problems there exist efficient solvers, that guarantee to find the global optimum. In Case **B** the problem is not completely an MILP problem but it can be *approximated* with an MILP formulation. The obtained MILP solution can then be used as initial value for a general non-linear non-convex optimization method. This largely reduces the computation time. Case **C**, which uses the most general model, allows the control of complicated networks, but results in a non-linear non-convex optimization problem.

The remainder of this chapter is organized as follows. In Section 4.2 we first formulate a general control approach including control measures, possible control objectives, and constraints. Then we describe model predictive control, and present an overview of optimization algorithms that can be used, with a focus on mixed integer linear programming. Then we consider the three different cases **A**, **B**, and **C** in Sections 4.3, 4.4, and 4.5. For each case we develop the route choice model and a corresponding controller. Section 4.6 illustrates the developed control approach with two examples. A simple network is controlled using the approach of Case **A**, while the performance of the general approach, Case **C**, is illustrated by a more complex network in which the Braess paradox appears. Finally, conclusions are presented in Section 4.7.

4.2 Control approach

Recall that the objective of this chapter is to develop a control method for model-based day-to-day route choice control. In this section we first formulate the overall optimal control problem including control measures, control objectives, and constraints. Since this optimal control problem is not tractable in an on-line setting (except when small horizons are used), we propose an on-line control method based on model predictive control (MPC). This control method uses an optimization algorithm. We will shortly discuss some global optimization algorithms, and then focus on mixed integer linear programming, which is suitable for some of the controllers that we will develop later in this chapter.

Note that an overview of the main symbols used in this chapter is given in Appendix 4.A.

4.2.1 Control signal

In this chapter we consider two control measures to influence the route choice of the drivers: outflow control and speed limit control. Outflow control can be done via, e.g., traffic signals, off-ramp metering installations, or mainstream metering installations. The outflow limit $Q_l(d)$ of link l represents the maximum flow (veh/h) that is allowed to leave link l at day d. Speed limit control uses dynamic speed limits $v_l(d)$ for all links l. Speed limits are displayed using variable message signs, and show the maximum allowed speed on the links. In general we represent the values of the control signal with $c(d)$. We assume that these values are constant during a day d. However, for Case **C** the approach can be extended to piecewise constant control signals.

4.2.2 Control objective

The objective of the controllers is to optimize the performance of the network, which can be formalized by selecting a cost function. Typical examples of cost functions in the context of route choice are the total time the vehicles spend in the network, the total queue length, or the norm of the difference between the realized flows and the desired flows on the routes. These cost functions serve either to handle as much traffic as possible in a short time, or to keep vehicles away from protected routes (e.g., routes through residential areas or nature reserves).

We will give three examples of possible cost functions. The total travel time can be computed as follows:

$$J^{\mathrm{TT}} = \sum_{d=1}^{N} \sum_{r \in R} \omega_r \beta_r(d) \tau_r^{\mathrm{route}}(d) \tag{4.1}$$

with N the number of considered days[3], R the set of all routes, $\beta_r(d)$ the turning rate toward route r at day d, $\tau_r^{\mathrm{route}}(d)$ the average travel time on route r at day d, and $\omega_r > 0$ the weight for route r. The second possible control objective is to approximate desired travel times $\tau_l^{\mathrm{desired}}(d)$:

$$J^{\mathrm{DTT}} = \sum_{d=1}^{N} \sum_{r \in R} \omega_r |\tau_r^{\mathrm{route}}(d) - \tau_r^{\mathrm{desired}}(d)| \tag{4.2}$$

with weights $\omega_r > 0$.

Remark 4.1 Note that the cost function described in (4.2) corresponds to the (weighted) 1-norm of the difference between the experienced and the desired travel time vectors. However, other norms can also be used, e.g., the 2-norm or the ∞-norm. Which norm should

[3]Note that when the demand varies between days, it might be useful to multiply the experienced travel times on each day with the corresponding demand.

be used depends on the goal and structure of the controller. The 1-norm and ∞-norm will result in linear problems, while the 2-norm yields a smooth function that allows for the use of efficient, gradient-based or Hessian-based optimization algorithms. □

Another option is to approximate a desired flow $Q_l^{\text{desired}}(d,\cdot)$ on link l. Note that for the cases with piecewise constant demands the periods that the demand is constant are counted with index i, where each period starts at event time i. The set of all event times i is only available at the end of the simulation. The appropriate desired piecewise constant flow pattern $Q_l^{\text{desired}}(d,i)$ can be obtained by fitting the continuous pattern $Q_l^{\text{desired,continuous}}(d,\cdot)$ on the event times set available at the end of the simulation. The difference between the realized flow and this desired flow is then given by

$$J^{\text{DF}} = \sum_{d=1}^{N} \sum_{l\in L} \sum_{i\in I_{v_l^{\text{d}}}} \left| \sum_{r\in R_l} Q_{l,r}^{\text{in}}(d,i) - Q_l^{\text{desired}}(d,i) \right| \tag{4.3}$$

where R_l denotes the set of all routes using link l, $I_{v_l^{\text{d}}}(d)$ denotes the set of all time period indices for the downstream vertex v_l^{d} of link l on day d, $Q_{l,r}^{\text{in}}(d,i)$ is the inflow of link l for route r during event time period i, and $Q_l^{\text{desired}}(d,i)$ the desired flow on link l during period i. With respect to the norm, Remark 4.1 is also valid for (4.3).

The state of the network at the end of the simulation for each day d should also be considered. Therefore a penalty on, e.g., the final queue lengths can be added:

$$J^{\text{final}} = \sum_{d=1}^{N} \sum_{l\in L} \sum_{r\in R_l} N_{l,r}^{\text{veh}}(d, n_{v_l^{\text{d}}}) \ .$$

To prevent large variations in the control input c which could lead to unstable, or even dangerous, traffic conditions, a penalty is formulated for these variations:

$$J^{\text{var}} = \sum_{d=1}^{N} ||c(d) - c(d-1)|| \ , \tag{4.4}$$

where also other norms could be used, see Remark 4.1.

The final cost function is usually a weighted combination of the different costs, combined with the penalty on variations in the control signal:

$$J = w_1 J^{\text{TT}} + w_2 J^{\text{DF}} + w_3 J^{\text{DTT}} + \ldots + w_{m-1} J^{\text{var}} + w_m J^{\text{final}}$$

with weights $w_j > 0$. In fact, the weights consist of two factors: a scaling of each partial cost function by its nominal value (e.g., the historical average), and a factor that represents the importance of the partial cost.

4.2.3 Constraints

Minimizing the cost function can have negative side effects. Reducing the flow on one route could, e.g., lead to an increased flow some on other routes, with congestion and longer travel times as results. To solve this problem, maximum or minimum values can be selected for flows, travel times, and queue lengths:

$$\sum_{r\in R_l} Q_{l,r}^{\text{eff}}(d,i) \le Q_l^{\text{eff,max}}(d,i) \ ,$$

$$\tau_r^{\text{route}}(d) \le \tau_r^{\text{route,max}}(d) \ ,$$

and

$$N_{l,r}^{\text{veh}}(d,i) \le N_{l,r}^{\text{veh,max}}(d,i)$$

where $Q_l^{\text{eff,max}}(d,i)$, $\tau_r^{\text{route,max}}(d)$, and $N_{l,r}^{\text{veh,max}}(d,i)$ are the maximum allowed values for the outflow, travel time and queue length on link l and route r at day d during period i.

Technical hardware possibilities and safety restrictions can lead to bounds on the values for the outflow limits and speed limits:

$$Q_l^{\min}(d) \le Q_l(d) \le Q_l^{\max}(d) \ ,$$

and

$$v_l^{\min}(d) \le v_l(d) \le v_l^{\max}(d)$$

with $Q_l^{\min}(d)$ and $Q_l^{\max}(d)$ the minimum and maximum values for the flow on link l at day d, and $v_l^{\min}(d)$ and $v_l^{\max}(d)$ the minimum and maximum values for the speed limits.

4.2.4 Overall optimal control problem

The overall control problem can then be formulated as follows:

$$\begin{aligned} \min_{(c(1),\ldots,c(N))} \quad & J(c(1),\ldots,c(N)) \\ \text{s.t.} \quad & \text{model equations} \\ & \text{constraints} \end{aligned}$$

where c is the control signal, and J are the total costs.

This optimal control problem is intractable, and the closed loop system that will appear when the controller is applied to the network will not be robust. To be able to handle errors in the model, changes in the demand, and other disturbances, we propose to perform the control with a method based on MPC as explained below.

4.2.5 Model predictive control

In MPC [25, 100] the objective is to determine at day d the control inputs $c(d)\ldots c(d+N_\text{p}-1)$ that optimize a cost function $J(d)$ over a given prediction period of N_p days ahead, given the current state of the network, the future demand, and a model of the system, and subject to operational and other constraints.

An overview of the process is given in Figure 4.2.5. The real traffic network is measured or estimated based on measurements. The measurements or estimates are fed into the MPC controller every time step. Then the controller uses the prediction model in combination with an optimization algorithm to determine the optimal control inputs based on the

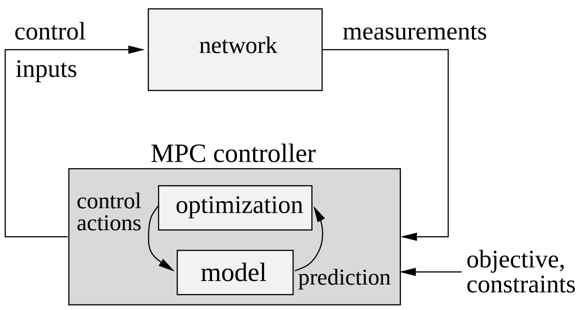


Figure 4.2: Model predictive control scheme.

selected control objective and constraints. The optimal control inputs are then applied in the real network in a receding horizon manner i.e. of the optimal control signal sequence only the first sample $c^*(d)$ is applied to the system. Next, at day $d+1$, the procedure is repeated given the new state of the system, and a new optimization is performed for days $d+1$ up to $d+N_\mathrm{p}$. Of the resulting control signal again only the first sample is applied, and so on.

The cost function that is selected for the optimal control problem in Section 4.2.2 can be adapted for the MPC controller. Each day, the model is used to predict the traffic and to compute the cost for the prediction period covering days d up to $d+N_\mathrm{p}$[4]. For the total travel time this results in:

$$J^{\mathrm{TT}}(d) = \sum_{j=1}^{N_\mathrm{p}} \sum_{r \in R} \omega_r \beta_r(d+j) \tau_r^{\mathrm{route}}(d+j) \quad .$$

The MPC cost functions corresponding to J^{DTT} and J^{DF} (see (4.2), (4.3)) can be defined in a similar way.

The prediction horizon N_p should be selected long enough to capture all the effects of the control actions. So when considering route choice, it should at least be equal to the travel time of the longest route, to incorporate the effect of the travel time on drivers that take this route.

The optimization results in a sequence of optimal control inputs $c^*(d), c^*(d+1), \ldots,$ $c^*(d+N_\mathrm{p}-1)$. To reduce the computational complexity a control horizon N_c ($N_\mathrm{c} < N_\mathrm{p}$) is usually introduced and the control sequence is constrained to vary only for the first N_c days, after which the control inputs are set to stay constant, i.e. $c(d+j) = c(d+N_\mathrm{c}-1)$ for $j = N_\mathrm{c}, \ldots, N_\mathrm{p}-1$.

Note that the proposed approach is generic and modular. Many models, cost functions, and optimization algorithms can be used. In this chapter we will develop a basic route choice model, and use speed limit control and outflow control within the MPC controller. However, if it is required, other, more complex models and control measures could be used instead.

[4]Note that the controller cannot influence the performance at day d anymore, since this performance depends on the turning rates, which are the state of the network and have already been determined during day $d-1$.

4.2.6 Optimization

MPC uses an optimization algorithm to determine the optimal values for the control variables. In general, the optimal solution of the route choice control problem cannot be computed analytically. The given problem is a non-convex nonlinear optimization problem with possibly multiple local minima. This requires the use of advanced optimization algorithms, which in general cannot guarantee that the optimal solution is found, and often reach a local optimum, which can lead to suboptimal behavior. Hence, a global optimization method is required such as genetic algorithms, simulated annealing, pattern search, or multi start local optimization [19, 44, 53, 61, 126]. This, however, increases the computation time, which is undesired for on-line computations. The selection of an optimization algorithm is thus based on the trade-off between the accuracy of the solution and the required computational effort.

When the control problem can be reformulated as a linear or piecewise affine problem, mixed integer linear programming (MILP) can be used for the optimization. Within a MILP problem the optimization of real and integer variables is combined, leading to a general formulation of the problem for a given matrices A, A^{eq}, and vectors b, b^{eq}, and c:

$$\begin{aligned} \min_x \ & c^{\mathrm{T}} x \\ \text{s.t.} \ & Ax \leq b \\ & A^{\text{eq}} x = b^{\text{eq}} \end{aligned}$$

where

$$x = \begin{bmatrix} x_i \\ x_r \end{bmatrix}, \ \ x_i \in \mathbb{Z}, \text{ and } x_r \in \mathbb{R} \ \ .$$

Several efficient branch-and-bound MILP methods [55] are available. Moreover, there exist several commercial and free solvers for MILP problems such as, e.g., CPLEX, Xpress-MP, GLPK, or lp_solve (see [5, 98] for an overview). In principle, and when the algorithm is not terminated prematurely, these algorithms guarantee to find the global optimum of the MILP problem efficiently. This makes them suitable for online optimization.

To reformulate optimization problems as MILP problems we will have to remove the nonlinearities of the models. This is done by recasting the nonlinear equations into linear ones, and by introducing additional auxiliary variables. To perform these transformations we use the following equivalences [12], where δ represents a binary-valued scalar variable, y a real-valued scalar variable, and f a scalar function defined on a bounded set X with upper and lower bounds U_f and L_f for the function values:

$$U_f = \max_{x \in X} f(x) \ \ ,$$

$$L_f = \min_{x \in X} f(x) \ \ .$$

We have

P1: $[f \leqslant 0] \leftrightarrow [\delta = 1]$ is true if and only if

$$\begin{cases} f \leqslant U_f (1 - \delta) \\ f \geqslant \epsilon + (L_f - \epsilon)\delta \ \ , \end{cases}$$

where ϵ is a small positive number (typically the machine precision[5]),

P2: $y = \delta f$ is equivalent to

$$\begin{cases} y \leqslant U_f \delta \\ y \geqslant L_f \delta \\ y \leqslant f - L_f(1-\delta) \\ y \geqslant f - U_f(1-\delta) \end{cases} .$$

P3: $\delta = \delta_1 \delta_2$ is equivalent to

$$\begin{cases} -\delta_1 + \delta \leqslant 0 \\ -\delta_2 + \delta \leqslant 0 \\ \delta_1 + \delta_2 - \delta \leqslant 1 \end{cases} .$$

4.3 Case A: Constant demand, separate routes

We now develop a route choice control method for a network with a constant demand and separate routes. First we formulate a basic route choice model, and then we develop the route choice controller based on this model.

4.3.1 Model development

The route choice model describes how the travel time experienced on a given route on a particular day affects the route choice on the next day. Within this subsection, we will first present a basic route choice network, then formulate the travel time model, and next describe the resulting route choice model.

Basic route choice network

To explain the modeling approach, we consider a network with one origin and one destination that are connected via multiple routes, see Figure 4.3. Such a network contains all features that are required for route choice, but it is small enough to make intuitive understanding possible. We assume that drivers enter this network at the origin and make their route choice immediately. Then they experience a travel time during their trip through the network and leave the network at the destination.

We will look at the day-to-day evolution of the traffic flows in particular part of the day, e.g., the morning peak. The considered period is denoted by the time interval $[0, T]$. For a given day d we assume that the traffic demand $D(d)$ (veh/h) in the network is constant. The demand is distributed over the routes according to the turning fraction $\beta_r(d)$, which gives the fraction of the vehicles that select route r on day d. The turning fraction is computed with the route choice model that will be described below. The sum of the turning rates should be equal to 1:

$$\sum_{r \in R} \beta_r(d) = 1$$

[5] The reason for introducing ε is that an equation like $Ax - b > 0$ does not fit the MILP framework, in which only non-strict inequalities are allowed. Therefore, $Ax - b > 0$ will be replaced by the equation $Ax - b \geq \varepsilon$ with ε a small tolerance, typically the machine precision, where we assume that in practice the case $0 < Ax - b < \varepsilon$ cannot occur due to the finite number of bits used for representing real numbers on a computer.

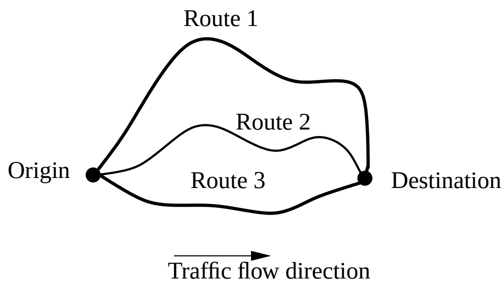


Figure 4.3: Network with three routes.

where R is the set of all routes in the network.

Since in this case the model considers separate routes, there is no difference between route parameters and link parameters. In this section we therefore indicate both of them with the index l, so

$$\beta_r(d) := \beta_l(d) \quad .$$

Each route/link $l \in R$ in the network can be described by the following parameters. The length of link l is denoted by ℓ_l (km). Recall that the speed limit $v_l(d)$ (km/h) gives the maximum speed that is allowed on link l on day d. The outflow limit $Q_l(d)$ (veh/h) gives the number of vehicles per hour that are allowed to leave the link l on day d.

Travel time model

Since we assume that the route choice of the drivers is based on the experienced travel times, we first formulate a travel time model. In our approach, the queues are assumed to be vertical[6]. This means that the vehicles drive the whole route without delay, experiencing the free-flow travel time. At the end of the route, the vehicles enter the vertical queue and wait in this queue until they can leave the route. As a consequence, the travel time τ_l^{route} on a given link l consists of two components: the free-flow travel time τ_l^{free} and the average (over all vehicles) time spent in the queue τ_l^{queue} (which is taken to be 0 if no queue is present):

$$\tau_l^{\text{route}}(d) = \tau_l^{\text{free}}(d) + \tau_l^{\text{queue}}(d) \quad . \tag{4.5}$$

The free-flow travel time $\tau_l^{\text{free}}(d)$ on link l is given by:

$$\tau_l^{\text{free}}(d) = \frac{\ell_l}{v_l(d)} \quad . \tag{4.6}$$

The time spent in the queue depends on the number of vehicles in the queue. Let us consider link l. During one peak period, the queue at the end of link l grows as shown in Figure 4.4. Recall that the length of the considered period is denoted by T and that $\beta_l(d)D(d)$ gives the flow on route l. Let $N_l^{\text{veh}}(d,t)$ be the number of vehicles in the queue at time t on day d. When the free-flow travel time has passed, the first vehicles reach the end of the link and the queue starts to build up if the demand exceeds the outflow limit of the link.

[6] A vertical queue is a queue that has no physical length but stores the vehicles just in front of the bottleneck.

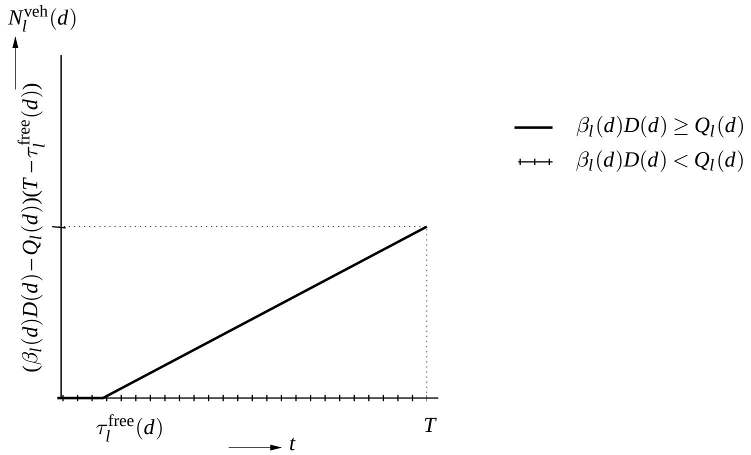


Figure 4.4: Queue length on a link during one day.

In Figure 4.4 the queue length is plotted for different demands. When the demand on link l is less than the outflow limit of link l, i.e. $\beta_l(d)D(d) \leq Q_l(d)$, no queue appears. When the demand is larger than the outflow limit a queue starts to grow, with rate $\beta_l(d)D(d) - Q_l(d)$.

In order to compute the average time in the queue, we first compute the area $A_l(d)$ under the queue length curve of Figure 4.4, and divide this by the total number of vehicles that have exited the link. We consider two cases:

- If $\beta_l(d)D(d) \geq Q_l(d)$ then a queue will be created, and the maximum queue length is given by

$$N_l^{\text{veh,max}}(d) = (\beta_l(d)D(d) - Q_l(d))(T - \tau_l^{\text{free}}(d)) \ .$$

  Moreover, in this case the outflow of a link l is equal to $Q_l(d)$. This outflow is present during the a period with length $T - \tau_l^{\text{free}}(d)$. Hence, the area below the queue length graph is

$$A_l(d) = \frac{1}{2} N_l^{\text{veh,max}}(d)(T - \tau_l^{\text{free}}(d)) = \frac{1}{2}(\beta_l(d)D(d) - Q_l(d))(T - \tau_l^{\text{free}}(d))^2 \ ,$$

  and thus

$$\tau_l^{\text{queue}}(d) = \frac{A_l(d)}{Q_l(T - \tau_l^{\text{free}}(d))} = \frac{1}{2}\frac{(\beta_l(d)D(d) - Q_l(d))(T - \tau_l^{\text{free}}(d))}{Q_l(d)} \ . \tag{4.7}$$

- If $\beta_l(d)D(d) < Q_l(d)$ then no queue will arise, and thus $\tau_l^{\text{queue}}(d) = 0$.

Since we may assume without loss of generality that $\tau_l^{\text{free}}(d) < T$, we can combine both situations:

$$\tau_l^{\text{queue}}(d) = \max\left(0, \frac{1}{2}\frac{(\beta_l(d)D(d) - Q_l(d))(T - \tau_l^{\text{free}}(d))}{Q_l(d)}\right) \ . \tag{4.8}$$

Remark 4.2 We can also take into account the vehicles that are still in the queue at the end of the simulation period as follows. Note that such a queue can only be created if $\beta_l(d)D(d) \geq Q_l(d)$. The outflow rate of the queue is then $Q_l(d)$ and the number of vehicles in the queue at time T is equal to $N_l^{\text{veh,max}}(d)$. Hence, it takes on the average $\frac{1}{2}\frac{N_l^{\text{veh,max}}(d)}{Q_l(d)}$ time units for the vehicles to leave the queue. This quantity has to be added to (4.7). So in this case (4.7) is still valid but the factor $\frac{1}{2}$ disappears. □

Route choice model

The travel times are used as input for the route choice model. Based on a difference in travel time between the routes, the turning rates will change as follows. The turning rate toward route 1 is given by

$$\beta_1(d+1) = \min\Big(\max\big(0, \beta_1(d) + \sum_{\rho \in R, \rho \neq 1} \kappa_{\rho,1}(\tau_\rho^{\text{route}}(d) - \tau_1^{\text{route}}(d))\big), 1\Big) \quad , \tag{4.9}$$

where $\kappa_{\rho,1}$ expresses the fraction of the drivers on route ρ that change their route toward route 1 from one day to the next based on the travel time difference.

The turning rates $\beta_2 \ldots \beta_{l^{\text{last}-1}}$ can be determined with

$$\beta_l(d+1) = \min\Big(\max\big(0, \beta_l(d) + \sum_{\rho \in R, \rho \neq l} \kappa_{\rho,l}(\tau_\rho^{\text{route}}(d) - \tau_l^{\text{route}}(d))\big), 1 - \sum_{\ell=1}^{l-1} \beta_\ell(d+1)\Big) \quad , \tag{4.10}$$

where $\kappa_{\rho,l}$ expresses the fraction of the drivers on route ρ that change their route toward route l from one day to the next based on the travel time difference, and l^{last} is the turning rate toward the last route.

Finally, the turning rate toward the last route l^{last} is given by

$$\beta_{l^{\text{last}}}(d+1) = 1 - \sum_{l=1}^{l^{\text{last}-1}} \beta_l(d+1) \tag{4.11}$$

The maximum and minimum functions in (4.10) keep the value of $\beta_l(d+1)$ between 0 and 1, while (4.11) guarantees that the sum of all β's is equal to 1.

4.3.2 Controller development

We will develop the route choice controller based on the model described above. As control method we use MPC, see Section 4.2.5. We consider variable speed limit control, but note that for outflow control the same reasoning can be followed. We assume that the speed limits can only have two values v_a and v_b.

Since the model is linear we can formulate the optimization problem as a MILP problem, using the Properties **P1** and **P2**. We first reformulate the route choice model, and next we define the cost function. Finally we present the overall MILP optimization problem.

Reformulating the route choice model

We now translate the route choice model in such a way that the optimization problem will have the form of an MILP.

Since we assume that there are two values for the speed limits, the corresponding free-flow travel times can be represented by one binary variable δ_l as follows. Define (cf. (4.6))

$$\tau_{l,a}^{\text{free}} = \frac{\ell_l}{v_a}\ , \quad \tau_{l,b}^{\text{free}} = \frac{\ell_l}{v_b}\ , \text{ and } \Delta_l = \tau_{l,b}^{\text{free}} - \tau_{l,a}^{\text{free}} \quad .$$

Then we can select v_a or v_b on link l for day d by introducing a binary variable $\delta_{1,l}(d)$ and setting

$$\tau_l^{\text{free}}(d) = \tau_{l,a}^{\text{free}} + \Delta_l \delta_{1,l}(d) \quad .$$

Remark 4.3 The implementation of more values for the speed limits is straightforward. As example, we consider four values for the speed limits that correspond to four equidistant free-flow travel times, formulated as follows:

$$\tau_1^{\text{free}}(d) = \tau_{1,\min}^{\text{free}} + \Delta_1(\delta_{1a,1}(d) + \delta_{1b,1}(d) + \delta_{1c,1}(d)) \quad ,$$

where $\tau_{1,\min}^{\text{free}} = \frac{\ell_1}{v^{\text{high}}}$, $\tau_{1,\max}^{\text{free}} = \frac{\ell_1}{v^{\text{low}}}$, and $\Delta_1 = \frac{\tau_{1,\max}^{\text{free}} - \tau_{1,\min}^{\text{free}}}{3}$. For $\tau_2^{\text{free}}(d)$ a similar construction can be used. □

Recall that we consider the case of speed control with no outflow control; so $Q_l(d) = Q_l^{\max}$ for all d. If we substitute the above expression for $\tau_l^{\text{free}}(d)$ in (4.5) and (4.8) we get

$$\tau_l^{\text{route}}(d) = \max(0, y_{2,l}(d)) + \tau_{l,a}^{\text{free}} + \Delta_l \delta_{1,l}(d) \tag{4.12}$$

with

$$y_{2,l}(d) = a_{1,l}(d)\beta_l(d) + a_{2,l}(d)\delta_{1,l}(d)\beta_l(d) + a_{3,l}\delta_{1,l}(d) + a_{4,l} \tag{4.13}$$

with $a_{1,l}(d) = \frac{1}{2Q_l}D(d)(T - \tau_{l,a}^{\text{free}})$, $a_{2,l}(d) = -\frac{1}{2Q_l}D(d)\Delta_l$, $a_{3,l} = \frac{1}{2}\Delta_l$, and $a_{4,l} = -\frac{1}{2}(T - \tau_{l,a}^{\text{free}})$. By introducing an extra variable $y_{1,l}(d) = \delta_{1,l}(d)\beta_l(d)$ and using Property **P2** with $f_l(x) = \beta_l(d)$, $L_{f_l} = 0$, and $U_{f_l} = 1$, (4.13) can be transformed into a system of linear inequalities.

Now define the auxiliary variables $\eta_l(d)$ and $\sigma_l(d)$ such that (cf. (4.27), (4.10))

$$\sigma_l(d) = \beta_l(d) + \sum_{\rho \in R} \kappa_{r,\rho}(\tau_\rho^{\text{route}}(d) - \tau_l^{\text{route}}(d)) \tag{4.14}$$

$$\eta_l(d) = \max(0, \sigma_l(d)) \quad . \tag{4.15}$$

Then we have

$$\beta_l(d+1) = \min(\eta_l(d), 1) \quad . \tag{4.16}$$

Consider (4.16) and define the binary variable $\delta_{4,l}(d)$ such that

$$\delta_{4,l}(d) = 1 \text{ if and only if } \eta_l(d) \le 1 \quad .$$

Note that this equivalence can be recast into a system of linear inequalities via Property **P2**. It is easy to verify that now we have

$$\beta_l(d+1) = \min(\eta_l(d), 1) = \delta_{4,l}(d)\eta_l(d) + 1 - \delta_{4,l}(d) \ ,$$

which after introducing the auxiliary variable $z_l(d) = \delta_{4,l}(d)\eta_l(d)$ (this equivalence can also be recast as a system of linear inequalities via Property **P1**), results in the linear equation

$$\beta_l(d+1) = z_l(d) + 1 - \delta_{4,l}(d) \ .$$

If we now collect all variables for day d in one vector

$$\begin{aligned} w(d) = \big[&\beta_1(d) \ \dots \ \beta_{n_l}(d) \ \delta_{1,l}(d) \ \dots \ \delta_{4,n_l} \ y_{1,l}(d) \ \dots \ y_{3,n_l}(d) \ \tau_1^{\text{route}}(d) \ \dots \ \tau_{n_l}^{\text{route}}(d) \\ &\sigma_1(d) \ \dots \ \sigma_{n_l}(d) \ \eta_1(d) \ \dots \ \eta_{n_l}(d) \ z_1(d) \ \dots \ z_{n_l}(d) \big]^T \ , \end{aligned}$$

we can express $\beta_l(d+1)$ as an affine function of $w(d)$: $\beta_l(d+1) = aw(d) + b$ for a properly defined vector a and scalar b, where $w(d)$ satisfies a system of linear equations $Cw(d) = e$, $Fw(d) \leqslant g$, which corresponds to the various linear equations and constraints introduced above.

MILP cost function

To be able to transform the route choice control problem into an MILP problem, the cost function should be linear or piecewise affine. Possible objectives of the controller of Section 4.2.2 that allow reformulation into linear or piecewise affine form are minimizing the flow on a route, reaching a desired flow on one of the routes, or reaching a desired travel time. The MPC cost function for a minimum flow on route 1 is given by:

$$J(d) = \min \sum_{j=1}^{N_\text{p}} \beta_l(d+j)D(d+j) \ .$$

Let us define

$$\tilde{F}_l(d) = \begin{bmatrix} \beta_l(d+1)D(d+1) \\ \vdots \\ \beta_l(d+N_\text{p})D(d+N_\text{p}) \end{bmatrix}, \qquad \tilde{F}_l^{\text{desired}}(d) = \begin{bmatrix} Q_l^{\text{desired}}(d+1) \\ \vdots \\ Q_l^{\text{desired}}(d+N_\text{p}) \end{bmatrix} \ ,$$

where $Q_l^{\text{desired}}(d+j)$ denotes the desired flow on route l at day $d+j$.

The MPC cost function corresponding to reaching a desired flow on route l is then given by:

$$J(d) = \min \| Q_l^{\text{desired}}(d) - F_l(d) \| \ .$$

When either the 1-norm or the ∞-norm are used, this cost function will be linear and can be reformulated for the MILP problem. When a 1-norm is used, the problem can be transformed into a linear one as follows:

$$\min \|\tilde{F}_l^{\text{desired}}(d) - \tilde{F}_l(d)\|_1 = \min \sum_{j=1}^{N_\text{p}} |Q_l^{\text{desired}}(d+j) - \beta_l(d+j)D(d+j)| \tag{4.17}$$

$$= \min \sum_{j=1}^{N_\text{p}} q(d+j)$$

$$\text{s.t. } q(d+j) \geqslant Q_l^{\text{desired}}(d+j) - \beta_l(d+j)D(d+j)$$

$$q(d+j) \geqslant -Q_l^{\text{desired}}(d+j) + \beta_l(d+j)D(d+j)$$

$$\text{for } j = 1, \dots, N_\text{p}.$$

It is easy to verify that for the optimal solution of the latter problem we have

$$q^*(d+j) = \max\left(Q_l^{\text{desired}}(d+j) - \beta_l^*(d+j)D(d+j), -Q_l^{\text{desired}}(d+j) + \beta_l^*(d+j)D(d+j)\right)$$

$$= |Q_l^{\text{desired}}(d+j) - \beta_l^*(d+j)D(d+j)|$$

for all j.

Similarly, for the ∞-norm we have

$$\min \|\tilde{F}^{\text{desired}}(d) - \tilde{F}(d)\|_\infty = \min \max_{j=1,\dots,N_\text{p}} |Q_l^{\text{desired}}(d+j) - \beta_l(d+j)D(d+j)|$$

$$= \min q$$

$$\text{s.t. } q \geqslant Q_l^{\text{desired}}(d+j) - \beta_l(d+j)D(d+j)$$

$$q \geqslant -Q_l^{\text{desired}}(d+j) + \beta_l(d+j)D(d+j)$$

$$\text{for } j = 1, \dots, N_\text{p},$$

which is also a linear problem.

Another possibility is to strive for desired travel times on the routes. Let $\tau_1^{\text{desired}}(d+j)$ and $\tau_2^{\text{desired}}(d+j)$ denote the desired travel times on respectively route 1 and route 2 at day $d+j$. The problem of reaching desired travel times on each of the routes is then given by

$$\min \sum_{j=1}^{N_\text{p}} \omega_1 |\tau_1^{\text{route}}(d+j) - \tau_1^{\text{desired}}(d+j)| + \omega_2 |\tau_2^{\text{route}}(d+j) - \tau_2^{\text{desired}}(d+j)| \tag{4.18}$$

with $\omega_1, \omega_2 > 0$. This cost function is piecewise affine, but it can be rewritten as

$$\min \sum_{j=1}^{N_\text{p}} \omega_1 \phi_1(d+j) + \omega_2 \phi_2(d+j)$$

$$\text{s.t. } \phi_r(d+j) \geqslant \tau_r^{\text{route}}(d+j) - \tau_r^{\text{desired}}(d+j)$$

$$\phi_r(d+j) \geqslant -\tau_r^{\text{route}}(d+j) + \tau_r^{\text{desired}}(d+j)$$

$$\text{for } j = 1, \dots, N_\text{p} \text{ and for } r = 1, 2\,.$$

which is a linear programming problem[7]. For the optimal solution we have

$$\begin{aligned}\phi_r^*(d+j) &= \max\left(\tau_r^*(d+j)-\tau_r^{\text{desired}}(d+j), -\tau_r^*(d+j)+\tau_r^{\text{desired}}(d+j)\right)\\ &= |\tau_r^*(d+j)-\tau_r^{\text{desired}}(d+j)|\end{aligned}$$

for all j and for $r=1,2$. Hence, (4.18) also leads to a linear programming problem.

Constraints

It might be useful to add a constraint on the travel time of a certain route (see Section 4.2.3), because minimizing, e.g., the flow on other routes can result in a higher flow and thus a longer travel time on this route:

$$\tau_l(d+j) \leqslant \tau_l^{\max}(d+j) \quad \text{for } j=0,\ldots,N_{\text{p}}-1 \ , \tag{4.19}$$

where $\tau_l^{\max}(d+j)$ denotes the maximal travel time on route l on day $d+j$. Using (4.12) and (4.13) we can easily eliminate $\tau_l(d+j)$ from the constraint (4.19). This yields the equivalent system of constraints

$$\begin{aligned}&\tau_{l,a}^{\text{free}}+\Delta_l\delta_l(d+j) \leqslant \tau_l^{\max}(d+j)\\ &y_{2,l}(d+j)+\tau_{l,a}^{\text{free}}+\Delta_l\delta_{1,l}(d+j) \leq \tau_l^{\max}(d+j)\end{aligned}$$

for $j=0,\ldots,N_{\text{p}}-1$. Note that these constraints are also linear.

An alternative constraint is to have a minimal or maximal flow on a given route. For route l this would result in

$$Q_l^{\text{in,min}}(d+j) \leqslant (1-\beta_l(d+j))D(d+j) \leqslant Q_l^{\text{in,max}}(d+j) \ ,$$

for $j=1,\ldots,N_{\text{p}}$, where $Q_l^{\text{in,min}}(d+j)$ and $Q_l^{\text{in,max}}(d+j)$ denote respectively the minimal and maximal allowed flow on route l on day $d+j$. This constraint is also linear.

Overall MILP problem for constant demand

If we collect the linear objective function and all the linear constraints introduced above into one large problem, we get an MILP problem in the variables $w(d), w(d+1),\ldots,w(d+N_{\text{p}}-1), \beta_1(d+N_{\text{p}})$ and $q(d+1), q(d+2),\ldots,q(d+N_{\text{p}})$ (when the 1-norm is used for the cost function) or q (when the ∞-norm is used). This means that within the controller a MILP solver can be used, which reduces the computation time, and guarantees the detection of the global optimum. This will improve the performance of the controller, and due to the lower computation times it allows for the control of larger networks.

4.4 Case B: Time-varying demand, separate routes

In this section, we extend the previous results to time-varying demand profiles. We first formulate the model, and next develop the controller.

[7] Note that if instead of the absolute value the square of the difference is minimized, the problem can be formulated as a mixed integer quadratic programming problem (MIQP).

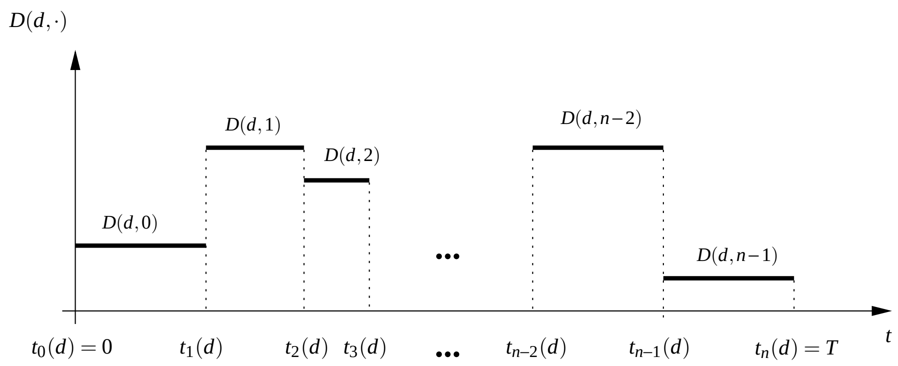


Figure 4.5: Time-varying demand profile $D(d,\cdot)$ on day d.

4.4.1 Model development

Time-varying demand profiles can be determined using historical data since often the same demand patterns occur every day with some variations depending on the type of day (week day or weekend), the weather, and the season (summer or winter). Since here we allow a different demand profile for each day, such variations can be taken into account in our approach. In particular, we consider piecewise constant demand profiles, which allows for a good representation/approximation of reality, while preserving linearity properties. We denote the piecewise constant demand function at the origin on day d as $D(d,\cdot)$. More specifically, we have

$$D(d,t) = D(d,i) \quad \text{for } t \in [t_i(d), t_{i+1}(d))$$

for $i = 0,\ldots,n-1$ where $t_0(d) = 0$, $t_n(d) = T$, and $t_i(d) < t_{i+1}(d)$ for $i = 0,\ldots,n-1$ (see Figure 4.5).

The introduction of the piecewise constant demand does not change the equations for the free flow travel time (4.6) and the turning rates (4.10). However, the formula for the time spent in the queue should be adapted.

Just as before, the average time in the queue τ_l^{queue} on link l depends on the number of vehicles in the queue. We still assume that the queues are vertical queues that build up at the end of each route. So during the period $[0, T+\tau_l^{\text{free}}(d))$ the queue on link l grows as shown in Figure 4.6. Note that the time is divided into periods $[t_i(d)+\tau_l^{\text{free}}(d), t_{i+1}(d)+\tau_l^{\text{free}}(d))$ corresponding to the periods $[t_i(d), t_{i+1}(d))$ in the demand. Here, the term $\tau_l^{\text{free}}(d)$ is due to the fact that vehicles entering link l at time t will reach the queue at time $t+\tau_l^{\text{free}}(d)$. Next, the queue will grow or shrink depending on the value of the net growth of the queue, which for link l is given by $\beta_l(d)D(d,i) - Q_l(d)$ for $t \in [t_i(d)+\tau_l^{\text{free}}(d), t_{i+1}(d)+\tau_l^{\text{free}}(d))$ and $i = 0,\ldots,n-1$.

If we denote the number of vehicles in the queue on link l on day d and during time period $[t_i(d)+\tau_l^{\text{free}}(d), t_{i+1}(d)+\tau_l^{\text{free}}(d))$ by $N_l^{\text{veh}}(d,i)$, we have

$$N_l^{\text{veh}}(d,i) = 0 \tag{4.20}$$

$$N_l^{\text{veh}}(d,i+1) = \max\big(0, N_l^{\text{veh}}(d,i) +$$

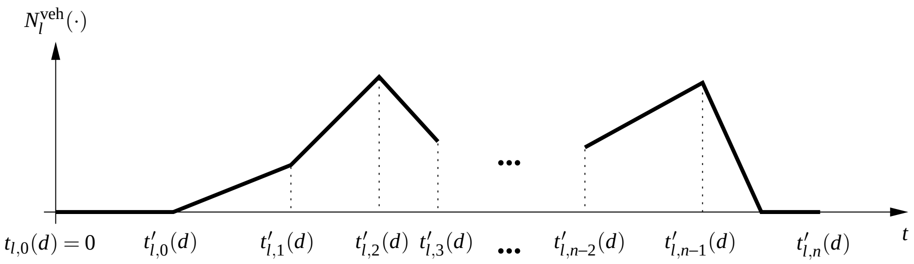


Figure 4.6: Evolution of the queue length $N_l^{\text{veh}}(d,\cdot)$ on link l and day d during the period $[0, T + \tau_l^{\text{free}}(d)]$. For the sake of simplicity of notation we have used $t'_{l,i}(d) = t_i(d) + \tau_l^{\text{free}}(d)$ in the figure.

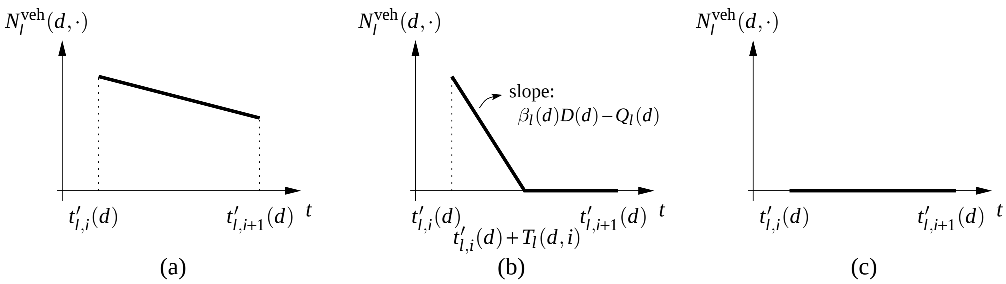


Figure 4.7: Three possible cases for the evolution of the queue length $N_l^{\text{veh}}(d,\cdot)$ on link l and day d during the time interval $[t_i(d) + \tau_l^{\text{free}}(d), t_{i+1}(d) + \tau_l^{\text{free}}(d))$ with $t'_{l,i}(d) = t_i(d) + \tau_l^{\text{free}}(d)$.

$$
\begin{aligned}
&(\beta_l(d)D(d,i) - Q_l(d))\big((t_{i+1}(d) + \tau_l^{\text{free}}(d)) - (t_i(d) + \tau_l^{\text{free}}(d))\big)\\
&= \max\big(0, N_l^{\text{veh}}(d,i) + (\beta_l(d)D(d) - Q_l(d))(t_{i+1}(d) - t_i(d))\big) \quad .
\end{aligned}
$$

Note that if $N_l^{\text{veh}}(d,i) + (\beta_l(d)D(d) - Q_l(d))(t_{i+1}(d) - t_i(d)) < 0$, the queue length already becomes 0 at some time $t_i(d) + T_l(d,i)$ with

$$
T_l(d,i) = \frac{N_l^{\text{veh}}(d,i)}{(\beta_l(d)D(d) - Q_l(d))} \quad .
$$

At this moment the queue on link l becomes empty (see Figure 4.7).

In order to compute the average time the vehicles spend in the queue on link l, we first compute the total area under the $N_l^{\text{veh}}(d,\cdot)$ curve. If we denote the area under the $N_l^{\text{veh}}(d,\cdot)$ curve between $t_i(d) + \tau_l^{\text{free}}(d)$ and $t_{i+1}(d) + \tau_l^{\text{free}}(d)$ by $A_{l,i}(d)$, there are three possible cases (see Figure 4.7(b)):

- If $N_l^{\text{veh}}(d,i+1) > 0$ then we have

$$
\begin{aligned}
A_{l,i}(d) =& \frac{1}{2}\Big(N_l^{\text{veh}}(d,i) + N_l^{\text{veh}}(d,i+1)\Big)\big((t_{i+1}(d) + \tau_l^{\text{free}}(d)) - (t_i(d) + \tau_l^{\text{free}}(d))\big) \qquad (4.21)\\
=& \frac{1}{2}\Big(N_l^{\text{veh}}(d,i) + N_l^{\text{veh}}(d,i+1)\Big)\big(t_{i+1}(d) - t_i(d)\big)\,.
\end{aligned}
$$

We then define $T_l^{\text{flow}}(d,i) = t_{i+1}(d) - t_i(d)$.

- If $N_l^{\text{veh}}(d,i+1) = 0$ and $N_l^{\text{veh}}(d,i) > 0$, then $Q_l(d) \neq \beta_l(d)D(d,i)$ and the queue length already becomes 0 at time $t_i(d) + T_l(d,i)$. Then we have

$$A_{l,i}(d) = \frac{1}{2} N_l^{\text{veh}}(d,i) T_l(d,i) = \frac{(N_l^{\text{veh}}(d,i))^2}{Q_l(d) - \beta_l(d)D(d,i)} .$$

  Now we have to make a distinction whether or not there is an outflow of link l after the time $t_i(d) + T_l(d,i)$. If $\beta_l(d)D(d,i) \neq 0$ then there will be an outflow and then we define $T_l^{\text{flow}}(d,i) = t_{i+1}(d) - t_i(d)$. Otherwise, we set $T_l^{\text{flow}}(d,i) = T_l(d,i)$.

- If $N_l^{\text{veh}}(d,i+1) = 0$ and $N_l^{\text{veh}}(d,i) = 0$, then $Q_l(d) = \beta_l(d)D(d,i)$ and the total area is zero: $A_{l,i}(d) = 0$, since in this case the N_l^{veh} curve is horizontal. Just as in the previous case we define $T_l^{\text{flow}}(d,i) = t_{i+1}(d) - t_i(d)$ if $\beta_l(d)D(d,i) \neq 0$ and $T_l^{\text{flow}}(d,i) = 0$ otherwise.

The total area under the N_l^{veh} curve is then equal to $A_l(d) = \sum_{i=0}^{n-1} A_{l,i}(d)$. Since there is an outflow of link l (with value $Q_l(d)$) during $T_l^{\text{tot}}(d) = \sum_{i=0}^{n-1} T_l^{\text{flow}}(d,i)$ time units, the total number of vehicles leaving the link is equal to $Q_l(d)T_l^{\text{tot}}(d)$. So the average time that the vehicles spend in the queue at the end of route l is given by:

$$\tau_l^{\text{queue}}(d) = \begin{cases} \dfrac{A_l(d)}{Q_l(d)T_l^{\text{tot}}(d)} & \text{if } T_l^{\text{tot}}(d) > 0\,, \\ 0 & \text{if } T_l^{\text{tot}}(d) = 0\,. \end{cases}$$

The total travel time for a route can then be computed with (4.5), and the turning rates with (4.10).

4.4.2 Controller development

In this section we show that for linear or piecewise affine cost functions the previously formulated MPC route choice optimization problem for Case **B** can be approximated by an MILP problem. In particular, we will consider the case of outflow control only (so there is no speed control). Further, we will consider the case where the network consists of only two routes ($R = \{1,2\}$), which leads to $\beta_2(d) = 1 - \beta_1(d)$. Note however that an extension to a network with more routes is straightforward.

Transformation of the model equations

We assume that the outflow limits can only have two non-zero values $Q_{r,a}$ and $Q_{r,b}$ and for simplicity we consider control for route 1 only.

Later on we will see that in the model equations the factor $\frac{1}{Q_r(d)}$ will appear. This factor can be represented by introducing binary variable as follows. If we define

$$\Delta_r = \frac{1}{Q_{r,b}} - \frac{1}{Q_{r,a}} \quad ,$$

then we can select $Q_{r,a}$ or $Q_{r,b}$ on route r for day d by introducing a binary variable $\delta_r(d)$ and setting

$$\frac{1}{Q_r(d)} = \frac{1}{Q_{r,a}} + \Delta_r \delta_r(d) \ . \tag{4.22}$$

Let us now first rewrite the equations for the evolution of $N_1^{\mathrm{veh}}(d)$ (for $N_2^{\mathrm{veh}}(d)$ a similar reasoning holds, see Appendix 4.B). If we define

$$m_1(d,i) = \frac{N_1^{\mathrm{veh}}(d,i)}{Q_1(d)} \ \text{ for all } i,$$

then it follows from (4.20) that

$$m_1(d,i+1) = \max\left(0, m_1(d,i) + \left(\frac{\beta_1(d)D(d,i)}{Q_1(d)} - 1\right)(t_{i+1}(d) - t_i(d))\right) \tag{4.23}$$

with $m_1(d,0) = 0$ (cf. (4.20)).

We will now transform (4.23) into mixed-integer linear equations. If we substitute (4.22) into (4.23) we get an expression of the form

$$m_1(d,i+1) = \max\big(0, m_1(d,i) + a_{1,1,i}(d)\beta_1(d) + a_{2,1,i}(d)\delta_{1,1}(d)\beta_1(d) + a_{3,1,i}(d)\big) \tag{4.24}$$

with $a_{1,1,i}(d) = \frac{D(d,i)}{Q_{1,a}}(t_{i+1}(d) - t_i(d))$, $a_{2,1,i}(d) = D(d,i)\Delta_1(t_{i+1}(d) - t_i(d))$, and $a_{3,1,i}(d) = -t_{i+1}(d) + t_i(d)$. By introducing an extra variable $y_{1,1}(d) = \delta_{1,1}(d)\beta_1(d)$ and using Property **P2** with $f_1(x) = \beta_1(d)$, $L_{f_1} = 0$, and $U_{f_1} = 1$, (4.24) can be transformed into a system of linear inequalities together with the nonlinear equation

$$m_1(d,i+1) = \max\big(0, m_1(d,i) + a_{1,1,i}\beta_1(d) + a_{2,1,i}y_{1,1}(d) + a_{3,1,i}\big) \, .$$

Now we define binary variables $\delta_{2,1,i}(d)$ such that $\delta_{2,1,i}(d) = 1$ if and only if $m_1(d,i) + a_{1,1,i}\beta_1(d) + a_{2,1,i}y_{1,1}(d) + a_{3,1,i} \geqslant 0$. Using Property **P1** this equivalence can be recast as a system of linear inequalities. Then we get

$$m_1(d,i+1) = \delta_{2,1,i}(d)(m_1(d,i) + a_{1,1,i}\beta_1(d) + a_{2,1,i}y_{1,1}(d) + a_{3,1,i}) \, .$$

By introducing extra variables $y_{2,1,i}(d) = \delta_{2,1,i}(d)m_1(d,i)$, $y_{3,1,i}(d) = \delta_{2,1,i}(d)\beta_1(d)$, and $y_{4,1,i}(d) = \delta_{2,1,i}(d)y_{1,1}(d)$, and using Property **P2** we obtain again a system of linear inequalities together with the equation

$$m_1(d,i+1) = y_{2,1,i}(d) + a_{1,1,i}y_{3,1,i}(d) + a_{2,1,i}y_{4,1,i}(d) + a_{3,1,i}\delta_{2,1,i}(d) \ ,$$

which is a linear equation.

Now we make the following approximation (see Figure 4.8): We always take expression (4.21) for $A_1(d,i)$. Moreover, we always take $T_1(d,i) = t_{1,i+1}(d) - t_{1,i}(d)$ even if there is no outflow of the link. It is important to note that for $n = 1$ this will still result in an exact value for $\tau_1^{\mathrm{queue}}(d)$ since $N_1^{\mathrm{veh,max}}(d,i)$ will first be multiplied by $\frac{1}{2}T_1(d,i)$ to get the area, and next be divided again by $Q_1(d)T_1(d,i)$ to obtain the average time spent in the queue. This implies that for $n = 1$ we will get the (exact) results of Section 4.3. However, for $n > 1$ we will only get an approximation for $\tau_1^{\mathrm{queue}}(d)$.

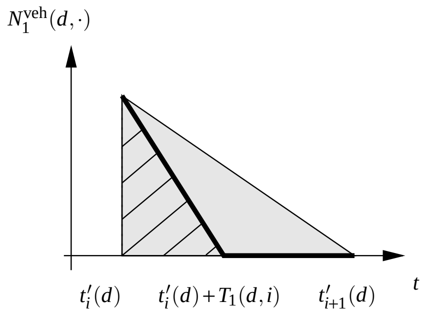


Figure 4.8: The area of the hashed triangle will be approximated by the area of the shaded triangle.

Since now $T_1^{\text{tot}}(d) = \sum_{i=0}^{n-1} T_1(d,i) = T$, the approximation results in

$$\tau_1^{\text{queue}}(d) = \frac{1}{2}\frac{\sum_{i=0}^{n-1}\big(N_1^{\text{veh}}(d,i)+N_1^{\text{veh}}(d,i+1)\big)\big(t_{1,i+1}(d)-t_{1,i}(d)\big)}{Q_1(d)T}$$
$$= \frac{1}{2T}\sum_{i=0}^{n-1}\big(m_1(d,i)+m_1(d,i+1)\big)\big(t_{1,i+1}(d)-t_{1,i}(d)\big) \quad .$$

Note that this expression is linear in m_1. Hence, it follows from (4.5) that $\tau_1^{\text{route}}(d)$ is also linear in m_1. Similarly (see also Appendix 4.B), $\tau_2^{\text{route}}(d)$ can be written as a linear expression in m_2 by introducing the additional real-valued auxiliary variables $y_{1,2,i}(d)$, $y_{2,2,i}(d)$, $y_{3,2,i}(d)$, $y_{4,2,i}(d)$, and binary auxiliary variables $\delta_{1,2}(d)$, $\delta_{2,2,i}(d)$ and $\delta_{3,2,i}(d) = \delta_{2,2,i}(d)\delta_{1,2}(d)$ (cf. Property **P3**) for $i = 0,\ldots,n-1$).

Now the remainder of the procedure corresponds to the procedure of Case **A**, starting with (4.14).

Overall MILP problem

The overall control problem is similar to the control problem of case **A** described in Section 4.3.2, and allows for the use of the cost functions formulated in this Section.

However, since the models used to generate the MILP problem are in this case only an approximation of the non-linear models, it is not recommended to directly apply the obtained optimum. Nevertheless this optimum could be used as initial starting point for the original nonlinear route choice MPC optimization problem, which significantly reduces the computation time.

4.5 Case C: Overlapping routes, multiple origins, and restricted link inflow capacities

In this section, we extend the approach above to include networks with overlapping routes, multiple origins and destinations, and restricted link capacities. We first develop the extended model, and next design a controller for this network.

4.5.1 Model development

First, we introduce the more complicated network with overlapping routes, and multiple origins and destinations. Then we adapt the timing that is used in the model, and present the link variables. Further, equations for the travel times, average queue lengths, and turning rates are presented, and the origins and destinations are modeled.

Network set-up

The network that we consider consists of multiple origins $o \in O$ and destinations[8] $e \in E$, where O and E are the sets of all origins and destinations in the network. The routes in the network can overlap, meaning that one link l can be used by multiple routes $r \in R$. The restricted inflow capacity of the links is denoted by Q_l^{cap} (veh/h).

Travel time model

The total travel time again consists of the sum of the free flow travel time and the time spent in the queue:

$$\tau_r^{\text{route}}(d) = \sum_{l \in L_r} \left(\tau_l^{\text{free}}(d) + \tau_{l,r}^{\text{queue}}(d) \right)$$

where L_r is the set of all links l in route r. The free flow travel time can be computed as before, see (4.6).

The computation of the time spent in the vertical queue is more involved. We assume that the queue on a link is divided into several independent partial queues, one for each route that uses the link. Let us now compute the time spent in each of these partial queues.

Timing

To be able to consider overlapping routes, we have to adapt the timing that is used in the model. In the previous cases **A** and **B** the event timings were route based. Now, we change the timing to be vertex-based as follows. For each vertex v in the network we introduce event times $t_{v,i}(d)$. Such an event time can correspond to two types of changes in the output flows of the vertex:

- a change in the input flow of one of the upstream links connected to the vertex, which is delayed by the free flow travel time on the given link,
- a partial queue becoming empty on one of the upstream links of a vertex.

Since these changes are not known beforehand, the traffic is simulated from the current event time $t_{v,i}(d)$ until the next known event time $t_{v,i+1}(d)$. When during the simulation of this period one of the two changes appears, a new event time $t_{v,\text{new}}(d)$ is created. The computations then have to be (re-)done for the period $[t_{v,i}(d), t_{v,\text{new}}(d))$, which leads to re-definition of the next event time instant with $t_{v,i+1}(d) := t_{v,\text{new}}(d)$.

[8] Since the index d is already used for the days, we denote destinations with e, coming from the word 'endpoint'.

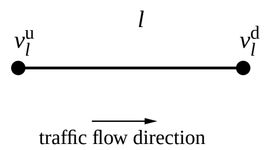


Figure 4.9: A link l with upstream vertex v_l^{u} and downstream vertex v_l^{d}.

Link variables

Consider link l, with its upstream vertex v_l^{u} and its downstream vertex v_l^{d}, as in Figure 4.9. The inflow $Q_{l,r}^{\mathrm{in}}(d,i')$ of each link l is given in the timing of the upstream vertex v_l^{u}, and present during the period $[t_{v_l^{\mathrm{u}},i'}(d), t_{v_l^{\mathrm{u}},i'+1}(d))$, which is the i'th period for vertex v_l^{u}. This flow experiences a delay equal to the free flow travel time, and then becomes the flow $Q_{l,r}^{\mathrm{in,queue}}(d,i)$ that enters the queue in the link during $[t_{v_l^{\mathrm{d}},i}(d), t_{v_l^{\mathrm{d}},i+1}(d))$, which is period i in the timing of the downstream vertex:

$$Q_{l,r}^{\mathrm{in,queue}}(d,i) = Q_{l,r}^{\mathrm{in}}(d,i')$$

with

$$[t_{v_l^{\mathrm{d}},i'}(d), t_{v_l^{\mathrm{d}},i'+1}(d)) = [t_{v_l^{\mathrm{u}},i}(d) + \tau_l^{\mathrm{free}}(d), t_{v_l^{\mathrm{u}},i+1}(d) + \tau_l^{\mathrm{free}}(d))$$

Recall that we assume that in each link there can be a partial queue for each route. The number of vehicles during time period i in the partial queue at the end of link l belonging to route r is given by $N_{l,r}^{\mathrm{veh}}(d,i)$. As indicated above, the inflow of the queue is denoted by $Q_{l,r}^{\mathrm{in,queue}}(d,i)$. The amount of traffic that can leave the queue depends on different factors:

- the outflow limit of the link,
- the number of queues on the link and their length,
- the capacity of the downstream links,
- the size of the flows that want to enter the downstream links.

We first introduce factors $\gamma_{l,r}(d,i)$ which divide the outflow limit $Q_l(d)$ proportionally over the different queues (see Figure 4.10):

$$\gamma_{l,r}(d,i) = \frac{\dfrac{N_{l,r}^{\mathrm{veh}}(d,i)}{\tau_{l,r}} + Q_{l,r}^{\mathrm{in,queue}}(d,i)}{\displaystyle\sum_{\rho \in R_l} \left(\dfrac{N_{l,\rho}^{\mathrm{veh}}(d,i)}{\tau_{l,\rho}} + Q_{l,\rho}^{\mathrm{in,queue}}(d,i) \right)} \tag{4.25}$$

where $\tau_{l,r}$ is a delay factor[9] representing the time that vehicles require to leave the queue on link l for route r. In this equation we assume that the flow that wants to leave the link consists of the vehicles that are in the link, and of the vehicles that enter the link during the current period.

[9]Note that the delay factor $\tau_{l,r}$ should be larger than $t_{v_l^{\mathrm{d}},i+1}(d) - t_{v_l^{\mathrm{d}},i}(d)$ to prevent the model from generating vehicles. If $\tau_{l,r} \leq t_{v_l^{\mathrm{d}},i+1}(d) - t_{v_l^{\mathrm{d}},i}(d)$ then an extra time step $t_{v_l^{\mathrm{d}},i'}(d)$ should be introduced between $t_{v_l^{\mathrm{d}},i}(d)$ and $t_{v_l^{\mathrm{d}},i+1}(d)$.

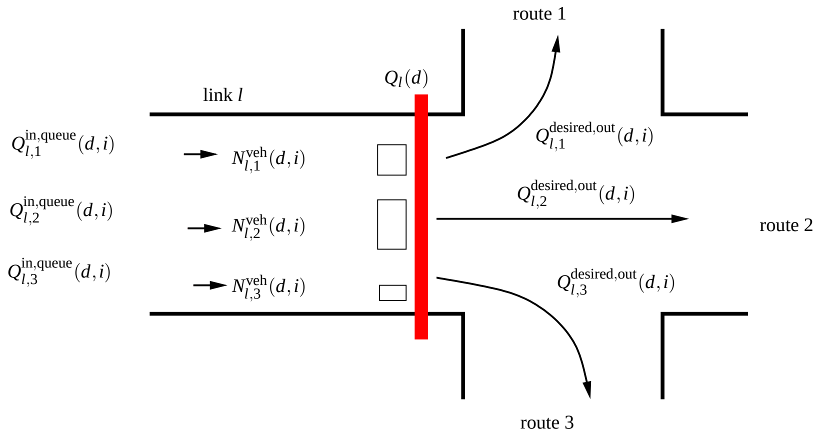


Figure 4.10: Routes on a link share the available outflow according to the proportions given by (4.25).

The flow that wants to leave a queue in the link during the period $[t_{v_l^{\mathrm{d}},i}(d), t_{v_l^{\mathrm{d}},i+1}(d))$ provided that the downstream capacity is large enough, is then given by

$$Q^{\mathrm{desired,out}}_{l,r}(d,i) = \min(\gamma_{l,r}(d,i)Q_l(d), \frac{N^{\mathrm{veh}}_{l,r}(d,i)}{\tau} + Q^{\mathrm{in,queue}}_{l,r}(d,i)) \quad .$$

Now we introduce the effect of the restricted inflow capacity of the links, see Figure 4.11. Due to this restricted capacity, the desired outflow $Q^{\mathrm{desired,out}}_{l,r}(d,i)$ is reduced to the effective outflow $Q^{\mathrm{eff}}_{l,r}(d,i)$ as follows. The inflow capacity of a downstream link l^{d} of link l is divided proportionally over the flows that want to enter the link, using a factor $\alpha_{l^{\mathrm{d}}}(d,i)$. This factor should be computed for each downstream link $l^{\mathrm{d}} \in D_l$:

$$\alpha_{l^{\mathrm{d}}}(d,i) = \min\left(1, \frac{Q^{\mathrm{cap}}_{l^{\mathrm{d}}}}{\sum\limits_{\xi \in U_{l^{\mathrm{d}}}} \sum\limits_{\rho \in R_{l^{\mathrm{d}}}} Q^{\mathrm{desired,out}}_{\xi,\rho}(d,i)}\right) \tag{4.26}$$

where $U_{l^{\mathrm{d}}}$ is the set of upstream links of link l^{d} and D_l the set of all downstream links of link l.

The flow that effectively leaves link l on route r toward link $l^{\mathrm{d}}_{l,r}$ is then given by:

$$Q^{\mathrm{eff}}_{l,r}(d,i) = \alpha_{l^{\mathrm{d}}_{l,r}}(d,i) Q^{\mathrm{desired,out}}_{l,r}(d,i)$$

where $l^{\mathrm{d}}_{l,r}$ is de link downstream of link l on route r. This outflow equals the inflow of the downstream link

$$Q^{\mathrm{in}}_{l^{\mathrm{d}}_{l,r},r}(d,i) = Q^{\mathrm{eff}}_{l,r}(d,i) \quad .$$

The number of vehicles in the queue on link l of route r can now be computed:

$$N^{\mathrm{veh}}_{l,r}(d,i+1)) = \max\left(0, N^{\mathrm{veh}}_{l,r}(d,i) + (Q^{\mathrm{in,queue}}_{l,r}(d,i) - Q^{\mathrm{eff}}_{l,r}(d,i))(t_{v_l^{\mathrm{d}},i+1}(d) - t_{v_l^{\mathrm{d}},i}(d))\right) \quad .$$

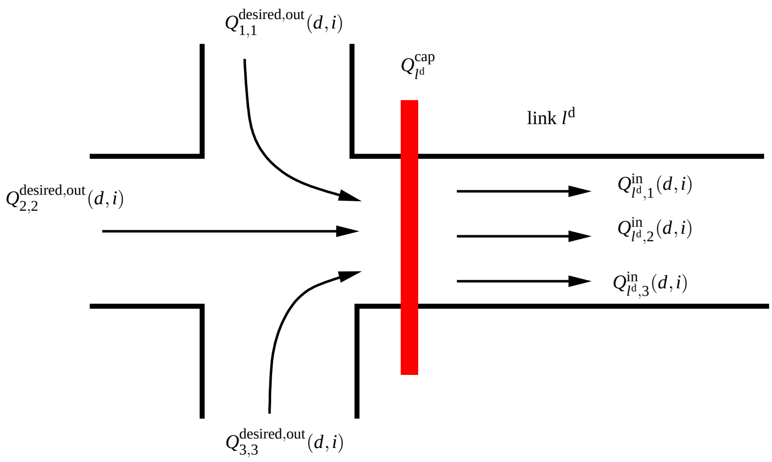


Figure 4.11: Inflow capacity shared by entering flows.

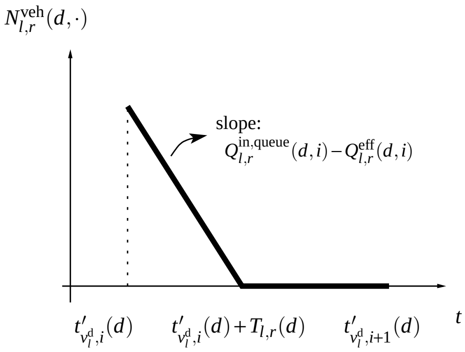


Figure 4.12: Evolution of the queue length $N_{l,r}^{\text{veh}}(d,\cdot)$ on link l of route r and day d during the interval $[t_{v_l^\text{d},i}(d),t_{v_l^\text{d},i+1}(d))$ when a queue becomes zero during this period.

If $N_{l,r}^{\text{veh}}(d,i)+(Q_{l,r}^{\text{in,queue}}(d,i)-Q_{l,r}^{\text{eff}}(d,i))(t_{v_l^\text{d},i+1}(d)-t_{v_l^\text{d},i}(d))<0$, the queue length already becomes 0 at some time $t_{v_l^\text{d},i}(d)+T_{l,r}(d,i)$ with

$$T_{l,r}(d,i)=\frac{N_{l,r}^{\text{veh}}(d,i)}{Q_{l,r}^{\text{eff}}(d,i)-Q_{l,r}^{\text{in,queue}}(d,i)} \quad .$$

At this moment the partial queue for the traffic on link l going via route r becomes empty (see Figure 4.12). This means that a new time instant $t_{v_l^\text{d},i+1,\text{new}}(d)=t_{v_l^\text{d},i}(d)+T_{l,r}(d,i)$ should be added to the timing of the downstream vertex v_l^d, and the computations for the current period should be re-done.

After the computations for the whole period are performed, the total number of vehicles in a link can be plotted as in Figure 4.13.

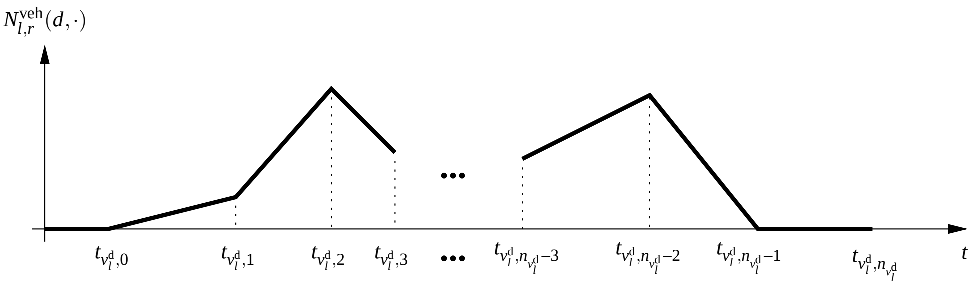


Figure 4.13: Evolution of the queue length $N^{\text{veh}}_{l,r}(d,\cdot)$ on link 1 and day d during the period $[0, T+\tau^{\text{free}}_l(d)]$.

Average time in the queue

In order to compute the average time the vehicles spend in the queue, we first compute the area under the $N^{\text{veh}}_{l,r}(d,\cdot)$ curve. This is done at the end of the simulation, so that the event timing is completely fixed. If we denote the area under the $N^{\text{veh}}_{l,r}(d,\cdot)$ curve between $t_{v^{\text{d}}_l,i}(d)$ and $t_{v^{\text{d}}_l,i+1}(d)$ by $A_{l,r}(d,i)$, there are two possible cases:

- If $N^{\text{veh}}_{l,r}(d,i) > 0$ or $N^{\text{veh}}_{l,r}(d,i+1) > 0$ then we have
$$A_{l,r}(d,i) = \frac{1}{2}\Big(N^{\text{veh}}_{l,r}(d,i)+N^{\text{veh}}_{l,r}(d,i+1)\Big)\big(t_{v^{\text{d}}_l,i+1}(d)-t_{v^{\text{d}}_l,i}(d)\big) \enspace .$$
- If $N^{\text{veh}}_{l,r}(d,i) = 0$ and $N^{\text{veh}}_{l,r}(d,i+1) = 0$ we have $A_{l,r}(d,i) = 0$ since the $N^{\text{veh}}_{l,r}(d,\cdot)$ curve is uniformly zero on the interval $[t_{v^{\text{d}}_l,i}(d), t_{v^{\text{d}}_l,i+1}(d))$.

Now we can compute the average time spent in the partial queue on link l for route r as

$$\tau^{\text{queue}}_{l,r}(d) = \frac{\sum_{i=0}^{n_{v^{\text{d}}_l}} A_{l,r}(d,i)}{\sum_{i=0}^{n_{v^{\text{d}}_l}} Q^{\text{eff}}_{l,r}(d,i)\big(t_{v^{\text{d}}_l,i+1}(d)-t_{v^{\text{d}}_l,i}(d)\big)}$$

where $n_{v^{\text{d}}_l}$ is the number of periods in the timing of vertex v^{d}_l.

Origin modeling

We model the origins as virtual links with length 0, see Figure 4.14. The demand at origin o with destination e during period $[t_{o,i}(d), t_{o,i+1}(d))$ is given by $D_{o,e}(d,i)$, for $i = 1,...,n_o$. This demand is divided over the routes via the turning rates

$$Q^{\text{in}}_{o,r}(d,i) = \beta_r(d) D_{o,e}(d,i) \quad \text{for all } \; r \in R_{o,e}.$$

where $R_{o,e}$ is the set of routes connecting origin o with destination e, and $Q^{\text{in}}_{o,r}(d,i)$ is the flow for route r that enters the virtual link connected to origin o during the period $[t_{o,i}(d), t_{o,i+1}(d))$. This flow enters the partial queues that can be present on this link. Furthermore, the origin is modeled in the same way as the links inside the network.

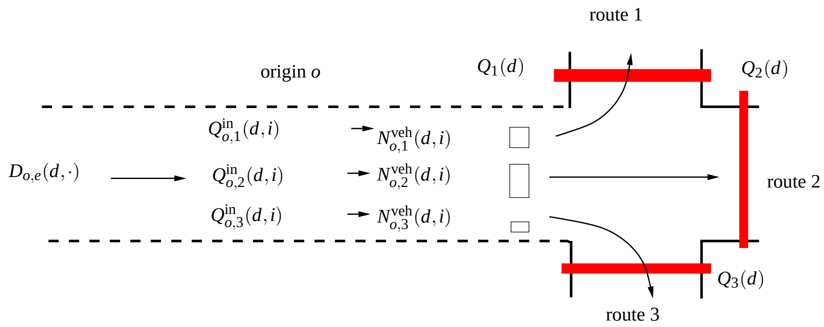


Figure 4.14: Virtual origin link.

Destination modeling

Just as the origins, we model the destinations as virtual links with length 0 and no outflow limit. Each virtual destination link is assumed to have an infinite inflow capacity. This implies that (cf. (4.26))

$$\alpha_e(d,i) = 1$$

for each destination e.

Turning rates

The route choice model used for the cases **A** and **B** is rather simple. We selected it since linear equations are necessary to form an MILP problem. Now we present an improved route choice model for case **C**, in which the differences in experienced travel times is used to determine non-normalized turning factors, which are later normalized between 0 and 1.

We still assume that the drivers will change their route when the travel time on another route is shorter[10]:

$$\zeta_r(d+1) = \max\left(0, \beta_r(d) + \sum_{\rho\in R_r^{\text{soe}},\rho\neq r} \kappa_{\rho,r}(\tau_r^{\text{route}}(d) - \tau_\rho^{\text{route}}(d))\right) \quad . \tag{4.27}$$

Here $\zeta_r(d)$ is the non-normalized turning factor of route r, and $\kappa_{\rho,r}$ includes the fraction of drivers on route ρ that change their route toward route r from one day to the next based on the travel time difference. Since the sum of the turning rates should be 1, they should be normalized to obtain the final turning rates:

$$\beta_r(d+1) = \frac{\zeta_r(d+1)}{\sum\limits_{\rho\in R_r}^{\text{oe}} \zeta_\rho(d+1)} \quad . \tag{4.28}$$

Remark 4.4 The formulation of the turning rates can be made even more exact by taking into account the number of vehicles that is currently using each link. This can be included

[10] Note that excluding the current route $\rho \neq l$ is not really necessary because the difference in travel times between the current route and the current route will always be 0.

as follows:

$$\zeta_r(d+1) = \max\Big(0, \beta_r(d) + \sum_{\rho \in R, \rho \neq r} \tilde{\kappa}_{\rho,r}\beta_\rho(d)\Delta_{\rho,r}(d) - \tilde{\kappa}_{r,\rho}\beta_r(d)\Delta_{r,\rho}(d)\Big) \quad , \tag{4.29}$$

where the two terms respectively express the number of vehicles that will divert from other routes ρ toward route r and the number of vehicles that will leave route r toward other routes ρ.

The differences in travel times $\Delta_{\rho,r}(d)$ and $\Delta_{r,\rho}(d)$ are given by

$$\Delta_{\rho,r}(d) = \begin{cases} 0 & \text{if } \tau_\rho^{\text{route}}(d) \leq \tau_r^{\text{route}}(d) \\ \tau_\rho^{\text{route}}(d) - \tau_r^{\text{route}}(d) & \text{if } \tau_\rho^{\text{route}}(d) > \tau_r^{\text{route}}(d) \end{cases} ,$$

$$\Delta_{r,\rho}(d) = \begin{cases} \tau_r^{\text{route}}(d) - \tau_\rho^{\text{route}}(d) & \text{if } \tau_\rho^{\text{route}}(d) < \tau_r^{\text{route}}(d) \\ 0 & \text{if } \tau_\rho^{\text{route}}(d) \geq \tau_r^{\text{route}}(d) \end{cases} .$$

The $\Delta_{\rho,r}(d)$ and $\Delta_{r,\rho}(d)$ are formulated in this way to express that drivers only change their route when the travel time on the other route shorter than on their current route.

In the approach that we use for case **C** we approximate (4.29) by including $\tilde{\kappa}_{\rho,r}$, $\tilde{\kappa}_{r,\rho}$, β_r, and β_ρ in the parameter $\kappa_{\rho,r}$ of (4.27). This introduces the property that the computed turning rates can exceed 1, and thus the resulting turning rates should be normalized, as in (4.28). For case **A** and **B** a linear model is required. Since the normalization (4.28) is a non-linear operation, we cannot use (4.27) and (4.28) for case **A** and **B**, but we have used (4.10) instead since that equation ultimately results in a mixed integer linear model. □

4.5.2 Controller development

When we use the model of Case **C** as prediction model, and when we assume real-valued control inputs, the optimization problem is a nonlinear non-convex real-valued problem. All cost functions described in Section 4.2.2 can be used, even when 2-norms are applied. To solve the resulting nonlinear type of problems multi-start local search methods (like SQP) and (semi-)global optimization methods (like genetic algorithms, pattern search, or simulated annealing) can be used, see [126]. The advantage of these methods is that they are suitable for complex optimization problems. However, these approaches in principle only yield a suboptimal solution since — in particular for larger networks or longer control horizons — it is in practice often not tractable to find the global optimum of the optimization problems that arise in MPC for route choice control. Moreover, the approaches lead to large computation times. Nevertheless, often these methods provide reasonably good solutions in a not too excessive computation time.

4.6 Worked example

We will illustrate the effects of the developed control approach with an example. The set-up is based on the Braess paradox [21, 114]. The Braess paradox states that adding a new link to a network could increase the total travel time in the network. We will first look at the network with two links, and apply control using MILP, as formulated for Case **A**. Next,

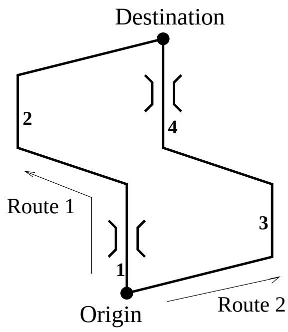


Figure 4.15: Network with two routes.

we will investigate the effect of adding a new link, using the model of Case **C**, which will indeed increase the total travel time when no control is applied. The controller that we have developed in Section 4.5.2 will be used to limit the outflow of the links in the network. We will show that the controller is able to find the well-known static optimum of removing the link. In general, the controller might be able to improve the situation even more by applying dynamic control. However, within this specific case study the dynamic and static optimum coincide.

4.6.1 Network with two routes

We will investigate a network with two routes, and apply a controller developed for case **A**. First the network is described, and then the settings for the control approach are given. At last, the simulation results are presented.

Set-up

We first consider a small network with two links, as in Figure 4.15. The parameters of the two links are: $Q_1 = Q_2 = 1200$ veh/h, $v_1^{\max} = v_2^{\max} = 120$ km/h, $v_1^{\min} = v_2^{\min} = 10$ km/h, and $\ell_1 = \ell_2 = 70$ km. For the route choice model we use $\kappa = 0.25$. We simulate a period of $T = 60$ minutes with a demand of $D = 3000$ veh/h, for 30 days. We will control this network with speed limit control.

In the uncontrolled case this should lead to an assignment where there is an inflow of 1500 veh/h on each link, due to the fact that the free flow travel times of both routes are then equal and a demand of 3000 veh/h enters the network. Now we apply outflow control, where we try to obtain a desired flow of 800 veh/h on route 1 which can, e.g., be useful when the route crosses a residential area.

Control approach

We formulate the control objective using the cost function given in (4.3), only using the 1-norm. A penalty on variations is added as formulated in (4.4). The prediction and control horizon of the MPC-based controllers are set to 8 days. For the sake of simplicity and to

Table 4.1: Computation time, cost, relative improvement (w.r.t. the uncontrolled case), and average turning rates for the entire closed-loop simulations with different optimization scenarios.

control method	computation time[11](s)	cost (h)	% improvement	average turning rate
no control	0	24444	0	0.54
`fmincon`, 1 init. point	10	13493	45	0.24
`fmincon`, 10 init. points	107	18211	25	0.18
`fmincon`, 20 init. points	380	15925	35	0.24
`fmincon`, MILP init. point	10	6975	72	0.34
MILP	0.4	14542	41	0.23

eliminate possible influences of model mismatches, we use the same model for the simulation and for the prediction by the MPC controller. For the speed limits we use a minimum value of $v^{\text{low}} = 60$ km/h, and a maximum of $v^{\text{high}} = v_1^{\max} = 120$ km/h.

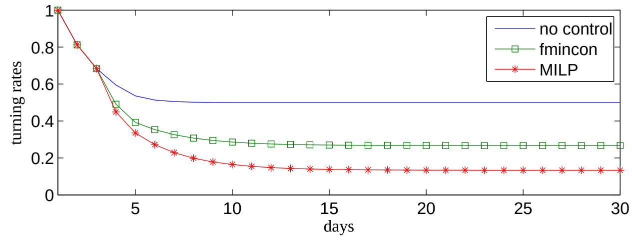
Simulation results

We simulate the traffic in the network in closed loop with different optimization strategies for the controller. Table 4.1 gives an overview of the results. The first simulation is performed without control, i.e. when the outflow is not limited and equal to its maximal value. Note that reducing the flow on route 1 is not in the drivers' interest, which explains the high cost in the no control case, and thus the relatively high improvements that can be obtained when control is applied.

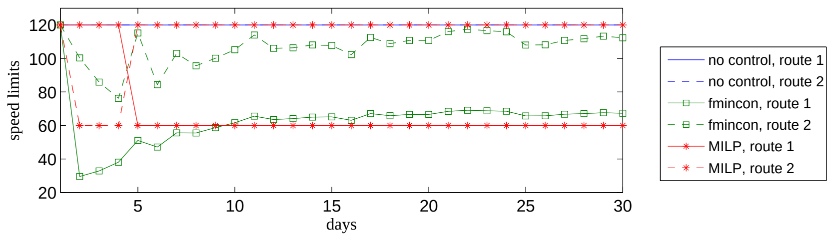
Next, we compare three approaches, all of which use the sequential quadratic programming (SQP) routine `fmincon` of the Matlab Optimization Toolbox [155], but with a different number of random initial points. When the number of randomly selected initial points increases, the performance increases, but the computation time also becomes larger. The computation time can be reduced by computing an initial point with MILP, which can then be used as initial point for one run of the `fmincon` algorithm. This reduces the computation time, and even improves the performance. To show that the additional `fmincon` run is really necessary, we also have performed a simulation with the MILP solution only. This simulation runs very fast, and gives already a good performance, but it can be improved significantly. The performance loss can be explained by the approximations that are required to obtain linear equations. The average turning rates obtained with the different controllers are given in the last column of Table 4.1. When MILP optimization is used as initial point for the `fmincon` optimization, the largest improvement is obtained.

Figure 4.16(a) shows the turning rates toward route 1 for the no control case, the MILP case, and the case with `fmincon` and an MILP initial point. When no control is applied, the turning rate converges to 0.5, meaning that the traffic divides equally over both routes. The MILP controller over-corrects this by steering a large part of the traffic to the second route. This problem is solved when `fmincon` is applied: the flow on the first route is lowered compared to the no control case, but higher than in the MILP case. Figure 4.16(b) shows the

[11] On a 1 GHz AMD Athlon 64x2 Dual Core processor.



(a) Turning rates.



(b) Speed limits.

Figure 4.16: Results for the two route network.

corresponding speed limits that are used to obtain the flows (note that in the uncontrolled case the speed limits are constant and equal to 120 km/h).

The simple example of this section illustrates that with a good initial point, which can be obtained via MILP, the optimization algorithm should only run once to obtain a good optimal value. This significantly reduces the computation time.

4.6.2 Network with the Braess paradox

Now we extend the network to a benchmark network in which the Braess paradox occurs. We consider the model and controller of case **C**.

Set-up

The network in which the Braess paradox occurs is shown in Figure 4.17, and is an extention of the network of Figure 4.15. The network consists of five links, and three routes. Route 1 consists of links 1 and 2, route 2 consists of links 3 and 4, and route 3 consists of links 1, 5, and 4. We select the link properties as follows. The inflow capacities are 1200 veh/h for all links. The maximum speeds on the network are $v_1^{\max} = v_4^{\max} = v_5^{\max} = 120$ km/h, $v_2^{\max} = v_3^{\max} = 60$ km/h, and the lengths of the links are $\ell_1 = \ell_4 = 30$ km, $\ell_2 = \ell_3 = 40$ km, and $\ell_5 = 10$ km. For the route choice model we use $\kappa = 0.25$. We simulate a period of $T = 100$ minutes with a demand of $D = 3000$ veh/h, for 10 days.

The network of Figure 4.17 illustrates the Braess paradox because two different equilibrium traffic assignments can be considered when investigating this network.

The first equilibrium appears in the network when all three routes are used. Drivers are attracted to route 3, which is the shortest route with respect to the number of kilometers. The

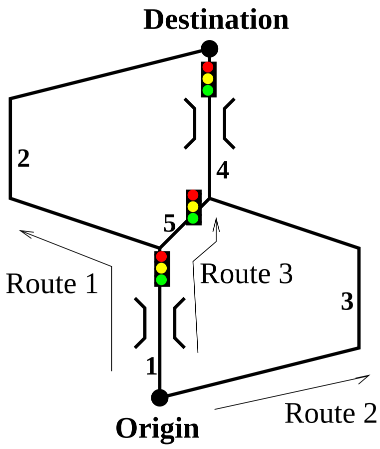


Figure 4.17: Network with overlapping routes.

capacity of this route however is lower than the demand and thus long queues are formed, resulting in a long travel time on this route. Due to this some drivers divert to the other two routes, and eventually an equilibrium assignment with relatively long travel times of 1.43 h appears.

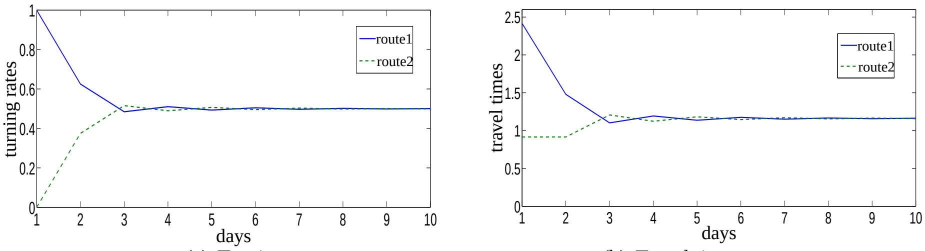
The second equilibrium uses only route 1 and 2. Assuming that route 3 is closed, the vehicles are divided over the two available routes. This results in an equilibrium assignment with travel times of 1.21 h, which is shorter than for the first equilibrium.

Control approach

We want to control the route choice in such a way that the travel time of 1.21 h appears. To reach this we apply outflow control on the network where all three routes are available. We expect that the control algorithm will lower the outflow limit for link 5, which corresponds to closing this link. In this case the situation with only two available routes appears, which will lead to the desired travel time.

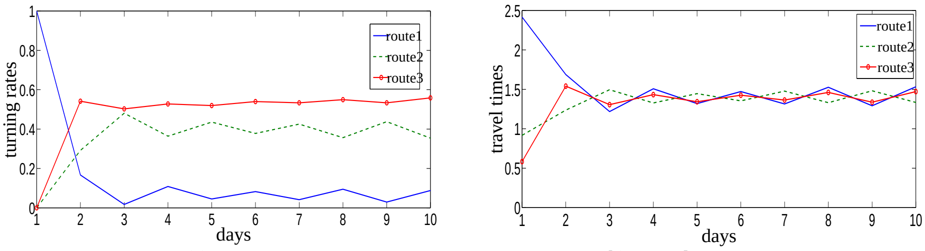
We use the developed control approach with the following parameters. Links 1, 4, and 5 are controlled using outflow control. The maximum outflow limits are $Q_1^{\max} = Q_4^{\max} =$ 600 veh/h, $Q_2^{\max} = Q_3^{\max} = Q_5^{\max} = 1200$ veh/h, and the minimum outflow limits are 0 veh/h for the controlled links. The prediction horizon is 10 days, and the control horizon 6 days. As cost function we select the total travel time in the network, just as for the network simulated earlier (see Section 4.6.1). As optimization algorithm we use SQP as implemented in `fmincon`.

We will perform three different simulations, two simulations of uncontrolled situations, and one of a controlled situation. The first simulation is a simulation of the network without link 5, and without control. The second simulation is a simulation of the whole network, still without control. These two simulations show that the Braess paradox is present in the network. The third simulation involves the whole network including the controller. For this specific case, where we consider a constant demand, a high learning rate, and the selected network layout, the dynamic and static equilibrium assignment coincide, and thus closing



(a) Turning rates. (b) Travel times.

Figure 4.18: Results of simulation 1: network without link 5, no control.



(a) Turning rates. (b) Travel times.

Figure 4.19: Results of simulation 2: whole network, no control.

link 5 is the optimal solution. We will illustrate that the developed controller is able to obtain this optimal solution.

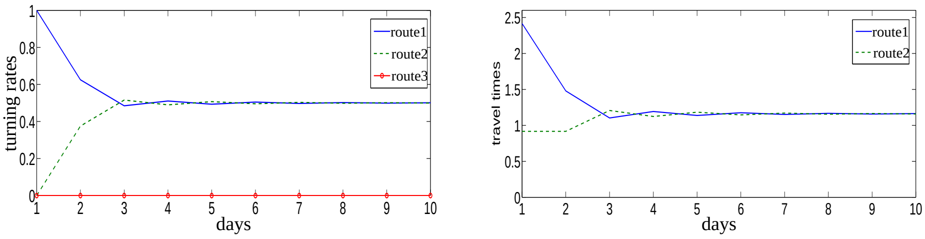
Simulation results

The two simulations without control show the effect that forms the Braess paradox, see Figures 4.18 and 4.19. Figures 4.18(a) and 4.19(a) show the turning rates for the different routes for Simulation 1 and 2. Figures 4.18(b) and 4.19(b) show the corresponding travel times. The simulation starts with initial turning rates $\beta_1(0) = 1$, $\beta_2(0) = \beta_3(0) = 0$. This means that the flows are not in equilibrium, and thus the turning rates change until an equilibrium assignment is reached. The total travel time J^{TT} for the first simulation is 12960 h, and for the second simulation 15130 h. The total travel time for the second simulation is indeed longer than the total travel time for the first simulation, illustrating that the use of the third route indeed increases the total travel time.

Figures 4.20(a) and 4.20(b) show the turning rates and travel times for Simulation 3, in which the controller is applied.The controller lowers the outflow of link 5 so that the flow toward the third route becomes 0 veh/h. As a result, the equilibrium assignment that uses only 2 routes is obtained, with the total travel time of 12960 h.

The advantage of the setup of this case study is that the optimal static solution of the control problem is known[12]. This solution is the removal of link 5, as shown with the first

[12]Note that in a general network the solution is not known, and thus no conclusion can be drawn about the



(a) Turning rates. (b) Travel times.

Figure 4.20: Results of simulation 3: whole network, control.

two simulations. The third simulation shows that the controller is indeed able to find the optimal solution, illustrating the performance of the controller.

4.7 Conclusions

We have developed a control approach based on model predictive control to influence the route choice of drivers, using control measures like outflow limits or variable speed limits. We have developed a route choice model for three cases. Case **A** included constant demand, separate routes, and one origin and destination. Case **B** extended the model to include piecewise constant demands, and Case **C** added overlapping routes, restricted link inflow capacities, and multiple origins and destinations.

We have used the developed models in a control approach based on model predictive control (MPC). For Case **A** and **B** we have respectively formulated and approximated the optimization problem of the MPC strategy as an mixed integer linear programming (MILP), which significantly reduces the computation time. For Case **C** multi-start local optimization or global optimization methods should be used. We have illustrated the control approach with two examples based on the Braess paradox. We showed that the obtained MILP solutions can be computed efficiently and can be used as good initial point for complex nonlinear non-convex optimization algorithms. For the Braess paradox, the optimal static solution is known and we have shown that our control approach indeed obtains this optimal solution.

Future research will include: calibration and validation of the model, development of robust controllers, developing faster optimization algorithms, and investigation of practical implementation issues, e.g., state estimators, required measurements, communication issues, and demand estimation.

absolute performance of the controller.

4.A List of symbols

General

L	set of all links in the network
L_r	set of all links in route r
U_l	set of all links immediately upstream of link l
D_l	set of all links immediately downstream of link l
ℓ_l	length of link l
$l^{\mathrm{d}}_{l,r}$	downstream link of link l on route r
v^{d}_l	downstream vertex of link l
v^{u}_l	upstream vertex of link l
R	set of all routes in the network
R_l	set of all routes that use link l
i	time period counter
$I_v(d)$	set of all time periods (only known after simulation) for vertex v on day d
$v_l(d)$	speed limit for link l on day d (km/h)
$v^{\min}_l$	minimum speed limit for link l (km/h)
$v^{\max}_l$	maximum speed limit for link l (km/h)
$Q_l(d)$	outflow limit for link l on day d (veh/h)
$Q^{\min}_l$	minimum outflow limit for link l (veh/h)
$Q^{\max}_l$	maximum outflow limit for link l (veh/h)
$\beta_r(d)$	turning rate for route r on day d
$\zeta_r(d)$	non-normalized turning rate for route r on day d
$\kappa_{r,l}$	parameter describing the part of drivers on route r that change their route toward route l
$\tau^{\mathrm{free}}_l(d)$	free flow travel time at link l on day d (h)
$\tau_l(d)$	travel time at link l during day d (h)
$\tau^{\mathrm{route}}_r(d)$	travel time for route r on day d (h)
T	length of the simulated period on a day (h)
N	total number of time steps in the simulation
N_{p}	prediction horizon (days)
N_{c}	control horizon (days)
c	general control signal
J^{var}	penalty in variations on the cost signal
J	total costs

Case A and B

$t_i(d)$	ith time event on day d (s)
$[t_i(d), t_{i+1}(d))$	ith time period on day d
$T_l(d,i)$	time after $t_i(d)$ that the queue on link l becomes 0 on day d in period $[t_i(d), t_{i+1}(d))$ (s)
$T^{\mathrm{tot}}_l(d)$	total time that the queue on link l is not empty at day d (h)
$D(d,i)$	demand at the origin on day d during time period i (veh/h)

$N_l^{\text{veh}}(d,i)$	number of vehicles in the queue on link l on day d during time period i (veh)
$A_l(d,i)$	area below the queue length graph of link l on day d during period i (veh·h)
$\tau_l^{\text{queue}}(d)$	travel time in the queue at link l during day d (h)

Case C

$t_{v,i}(d)$	ith event time for vertex v (s) on day d
$[t_{v,i}(d), t_{v,i+1}(d))$	ith time period for vertex v on day d
$R_{o,e}$	set of routes connecting origin o with destination e
R_r^{soe}	set of routes that connect the same origin destination pair as route r
$D_{o,e}(d,i)$	demand at origin o with destination e on day d during time period i (veh/h)
Q_l^{cap}	inflow capacity of link l (veh/h)
$Q_{l,r}^{\text{in}}(d,i)$	inflow of link l for route r at day d during time period i (veh/h)
$Q_{l,r}^{\text{in,queue}}(d,i)$	inflow of the queue on link l for route r at day d during time period i (veh/h)
$Q_{l,r}^{\text{desired,out}}(d,i)$	desired outflow of link l for route r on day d during time period i (veh/h)
$Q_{l,r}^{\text{eff}}(d,i)$	effective outflow of link l for route r on day d during time period i (veh/h)
$N_{l,r}^{\text{veh}}(d,i)$	number of vehicles in the queue on link l for route r on day d during time period i (veh)
$A_{l,r}(d,i)$	area below the queue length graph of link l for route r on day d during period i (veh·h)
$\gamma_{l,r}(d,i)$	factor that divides the available outflow limit over the queues in link l used by route r on day d during period i
$\alpha_l(d,i)$	factor that divides the available inflow capacity over the flows that enter link l on day d during period i
$\tau_{l,r}^{\text{queue}}(d)$	travel time in the queue at link l on route r during day d (h)

4.B Reformulation of $N_2^{\text{veh}}(d)$ for Case B

In Section 4.4.2 we have shown that the equations for N_1^{veh} and for τ_1 can be recast as a system of mixed integer linear equations and inequalities. In this appendix we explicitly derive the system of mixed integer linear equations and inequalities corresponding to N_2^{veh} and τ_2.

We first rewrite the equations for the evolution of N_2^{veh}:

$$N_2^{\text{veh}}(d,0) = 0 \tag{4.30}$$

$$N_2^{\text{veh}}(d,i+1) = \max\big(0, N_2^{\text{veh}}(d,i) + ((1-\beta_1(d))D(d,i) - Q_2(d))(t_{i+1}(d) - t_i(d))\big) \ . \tag{4.31}$$

If we define

$$m_2(d,i) = \frac{N_2^{\text{veh}}(d,i)}{Q_2(d)} \ ,$$

then it follows from (4.31) that

$$m_2(d,i+1) = \max\left(0, m_2(d,i) + \left(\frac{(1-\beta_1(d))D(d,i)}{Q_2(d)} - 1\right)(t_{i+1}(d) - t_i(d))\right) \tag{4.32}$$

with $m_2(d,0) = 0$ (cf. (4.30)).

Let us now transform (4.32) into mixed-integer linear equations. If we substitute (4.22) into (4.32) we get an expression of the form

$$m_2(d,i+1) = \max\left(0, m_2(d,i) + a_{1,2,i}\beta_1(d) + a_{4,2,i}\delta_{1,2}(d) + a_{2,2,i}\delta_{1,2}(d)\beta_1(d) + a_{3,2,i}\right) \tag{4.33}$$

with $a_{1,2,i} = -\frac{D(d,i)}{Q_{2,a}}(t_{i+1}(d) - t_i(d))$, $a_{4,2,i} = D(d,i)\Delta_2(t_{i+1}(d) - t_i(d))$, $a_{2,2,i} = -a_{4,2,i} = -D(d,i)\Delta_2(t_{i+1}(d) - t_i(d))$, and $a_{3,2,i} = \left(\frac{D(d,i)}{Q_{2,a}} - 1\right)(t_{i+1}(d) - t_i(d))$. By introducing an extra variable $y_{1,2}(d) = \delta_{1,2}(d)\beta_1(d)$ and using Property **P2**, (4.33) can be transformed into a system of linear inequalities together with the nonlinear equation

$$m_2(d,i+1) = \max\left(0, m_2(d,i) + a_{1,2,i}\beta_1(d) + a_{4,2,i}\delta_{1,2}(d) + a_{2,2,i}y_{1,2}(d) + a_{3,2,i}\right).$$

Now we define binary variables $\delta_{2,2,i}(d)$ such that $\delta_{2,2,i}(d) = 1$ if and only if $m_2(d,i) + a_{1,2,i}\beta_1(d) + a_{4,2,i}\delta_{1,2}(d) + a_{2,2,i}y_{1,2}(d) + a_{3,2,i} \geqslant 0$. Using Property **P1** this equivalence can be recast as a system of linear inequalities. Then we get

$$m_2(d,i+1) = \delta_{2,2,i}(d)(m_2(d,i) + a_{1,2,i}\beta_1(d) + a_{4,2,i}\delta_{1,2}(d) + a_{2,2,i}y_{1,2}(d) + a_{3,2,i}).$$

By introducing additional real-valued variables $y_{2,2,i}(d) = \delta_{2,2,i}(d)m_2(d,i)$, $y_{3,2,i}(d) = \delta_{2,2,i}(d)\beta_1(d)$, and $y_{4,2,i}(d) = \delta_{2,2,i}(d)y_{1,2}(d)$, and additional binary variables $\delta_{3,2,i}(d) = \delta_{2,2,i}(d)\delta_{1,2}(d)$ and using Properties **P2** and **P3** we obtain again a system of linear inequalities together with the linear equation

$$m_2(d,i+1) = y_{2,2,i}(d) + a_{1,2,i}y_{3,2,i}(d) + a_{4,2,i}\delta_{3,2,i}(d) + a_{2,2,i}y_{4,2,i}(d) + a_{3,2,i}\delta_{2,2,i}(d) \; .$$

Just as we did for $A_{1,i}(d)$ we now also always approximate $A_{2,i}(d)$. This results in

$$\tau_2^{\text{queue}}(d) = \frac{1}{2T}\sum_{i=0}^{n-1}\left(m_2(d,i) + m_2(d,i+1)\right)(t_{i+1}(d) - t_i(d))$$

which is a linear expression in the $m_2(d,i)$'s. Hence, it follows from (4.5) that $\tau_2^{\text{route}}(d)$ is also linear in m_2.

Chapter 5

Practical issues for model-based traffic control

Advanced traffic control systems can significantly improve the traffic flows on traffic networks. However, the implementation of such control systems is usually not straightforward. In this chapter we give an overview of practical issues related to the use of traffic controllers, and we pay attention to the steps that should be taken before they can be applied in practice. We look at issues related to the controller design, such as the network that should be controlled, the choice of the objectives and constraints for the controller, the selection of the model, and the selection of a control method. We also discuss implementation issues, in particular calibration and validation, state estimation, demand estimation, controller tuning, and performance evaluation.

In addition, we discuss the influence of the measurements on the performance of the controller, and we focus on different averaging methods for speed measurements.

In a case study we illustrate the steps of the controller development, and we investigate the influence of using the different speed averages on the performance of a dynamic speed limit controller. The results show that for the given case study the use of different averages results in a difference of a few percents in the controller performance.

5.1 Introduction

Current road networks often suffer from a lack of capacity, and/or an inefficient use of the available capacity. Advanced traffic control measures have been developed to reduce the corresponding problems, such as congestion and noise nuisance. The control measures influence the traffic in such a way that the existing road capacity is used more efficiently, and in this way the throughput of the network improves. However, the performance of these controllers largely depends on the choices made during the design process, which preceed the implementation process.

A specific type of advanced traffic controllers are model-based controllers, as used in many well-known traffic control systems, such as, e.g., UTOPIA [129], IN-TUC [49], and MITROP [59]. These controllers use a model of the traffic system to determine the settings for the traffic control measures. In this chapter we investigate this kind of controllers since they require the consideration of many practical issues, partly due to the inherent differences between the model and the real world. In this way we obtain a complete overview of all practical issues related to the implementation of general traffic controllers. We focus on a special form of model-based control: model predictive control (MPC). Traffic controllers that are based on MPC use a model to predict the future evolution of the traffic flows. Based on this prediction, the controller determines settings for the traffic control measures. Advantages of this control approach are that different control measures can be integrated into one control system and that the prediction allows for the investigation of the longer-term effects of the control actions and thus allows for the selection of control settings that are optimal for a longer period.

We divide the practical issues related to the deployment of traffic controllers into two main classes: design issues and implementation issues. The design issues that we consider are often related to the policy of the road authority: which traffic flows should be controlled, which network is considered, which measurements can be obtained, what is the objective of the controller, what are the constraints? The policy decisions influence more technical issues like the model selection, and the selection of the control method that is used.

When the design of the model-based controller is determined some implementation issues should be considered before the controller can be applied to a real traffic situation. The selected model should be calibrated and validated, a state estimation method should be selected, the expected demand must be estimated, and the controller must be tuned. After the implementation in the real network, the performance of the controller must be evaluated.

Many of the issues mentioned above strongly depend on measurements that are available. During the design a measurement structure should be designed, which includes selecting, e.g., detectors, communication networks, data polishing methods, and data handling methods. In this chapter we describe different detectors, and then focus on speed measurements obtained with loop detectors. The measured speeds should be averaged, for which six different avaraging methods are available. We investigate the influence of these different averaging methods on the performance of a model-based controller.

As a case study we develop a dynamic speed limit controller for the A12 freeway in the Netherlands. During the design of this controller we consider the design issues and implementation issues as far as they are usefull for a simulation study. The objective of the controller that is developed is to minimize the total time spent in the network by reducing shock waves. We use the developed controller to illustrate the influence of speed measurement averaging methods. We simulate the network with the controller using the six different averaging methods that are available, and we compare the resulting controller performances.

The remainder of this chapter is organized as follows. Section 5.2 explains model-based control and describes the general process of controller development. Section 5.3 considers issues that are related to the controller design, and Section 5.4 considers implementation related issues. The different methods for speed measurement averages are discussed in Section 5.5, and Section 5.6 presents the case study. Finally, conclusions are drawn in Section 5.7.

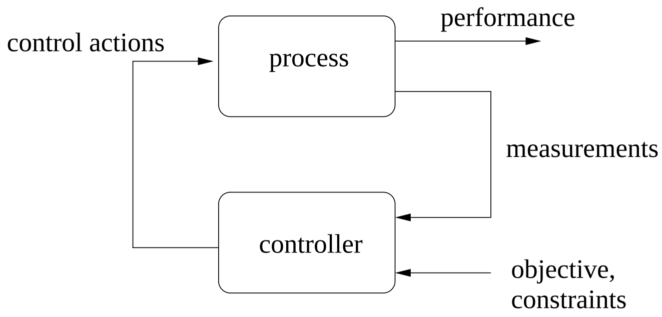


Figure 5.1: A general scheme for feedback control methods.

5.2 Model-based traffic control

This section gives a general description of model-based traffic control, and then explains model predictive control (MPC), which is a specific type of model-based control that is often used for traffic control. Next, an overview of the procedure of developing a controller is given.

5.2.1 General scheme for model-based traffic control

We first introduce the general setup of advanced traffic control methods. A feedback scheme of an advanced control method is shown in Figure 5.1. The process consists of the traffic network and the traffic flows that should be controlled. The traffic flows propagate through the network depending on the traffic scenario and the control actions, which leads to a specific performance of the network. To obtain information that could be used to determine the performance, the current state of the network should be measured or estimated based on measurements. These measurements can be performed using, e.g., radar detectors, loop detectors, and cameras. The measured quantities can be, e.g., flows, occupancies, and speeds. These measured values are fed into the controller, which determines the control signal based on these measurements and on desired performance of the network, which is described by the objectives of the controller. The control actions consist of the settings for the traffic measures, such as ramp metering rates, speed limit values, or timings for traffic signals. These measures then influence the process, and thus influence the performance. In this way, the controller is used to increase the network performance.

How the controller determines the control signal depends on the type of controller. Model-based controllers use an internal model of the traffic to determine the control signal. For urban areas the settings of the traffic signals can be determined using queue length models, which is investigated in Chapter 2 and in, e.g., [49, 129, 140, 185]. For freeways there are model-based controllers using dynamic speed limits, ramp metering installations, and peak lanes, described in, e.g., [64, 89]. Dynamic route guidance can be used for model-based route choice control, as in, e.g., [11, 46, 78, 81].

5.2.2 Model predictive control

In this chapter we consider a specific model-based control method called model predictive control (MPC). MPC-based controllers use a model to predict the evolution of the traffic flows, and use this prediction to determine the optimal control signals. MPC [100] has been developed for the process industry, and the first applications for traffic control are described in [58]. Traffic controllers that explicitly use MPC are proposed in [10, 64], and other controllers that use similar schemes are, e.g., [49, 129, 140, 179]. These controllers also use models and predictions to obtain the control settings, but they are not explicitly formulated corresponding to the MPC structure. When MPC is applied, the controller in Figure 5.1 contains a state estimation algorithm, a prediction model, and an optimization algorithm. The measurements from the real network are used to obtain the estimated state of the network. Next, the controller predicts the evolution of the traffic flows over a prediction period that has a length of N_p controller steps. Based on this prediction, the optimization algorithm is used to obtain the optimal (according to pre-defined objectives and constraints) settings for the control signals up to a control period of N_c controller steps, with $N_\mathrm{c} \leq N_\mathrm{p}$. During the control period N_c the control signals vary, while during the remainder of the prediction period N_p, the control signals are kept constant. The optimal values for the control signals during the current control time step are applied to the real network. At the next control time step the procedure is started again, with the horizon shifted one time step into the future. This is called the rolling horizon approach. For further information on MPC, we refer the interested reader to [25, 57, 100].

5.2.3 Controller development

The process of developing an advanced controller is a combination of design issues and implementation issues. Figure 5.2 presents an overview of the required steps. The process starts with the design issues, consisting of policy issues and technical issues. The policy issues consider the objectives and constraints of the controllers, the selection of the network, and the design of the measurement structure. Then technical steps consider the selection of the control method, and of the model. Guidelines for how the steps of the design process can be applied to real situations are presented in [110].

When the design issues have been considered, the general design of the controller is available. Now some steps have to be made that are more closely related to the implementation. The model should be calibrated and validated, meaning that values for the parameters in the model should be selected. Further, a procedure should be developed to estimate the state of the traffic flows in network, and to estimate the demand. With these issues settled, the controller can be applied in a simulation environment, to investigate the effects of policy/economical choices, to investigate the effects of choices regarding, e.g., the number of measurements, control measures, and objectives. The simulation environment can also be used to tune the controller. If problems are encountered, parts of the design process should be redone. In general, multiple iterations will be necessary before all problems are solved. When the simulation gives good results, the controller is ready to be implemented in the real network. When the controller is implemented, its performance can be evaluated by comparing measurements of the controlled situation with measurements of the uncontrolled situation, and with the results of the simulation experiments..

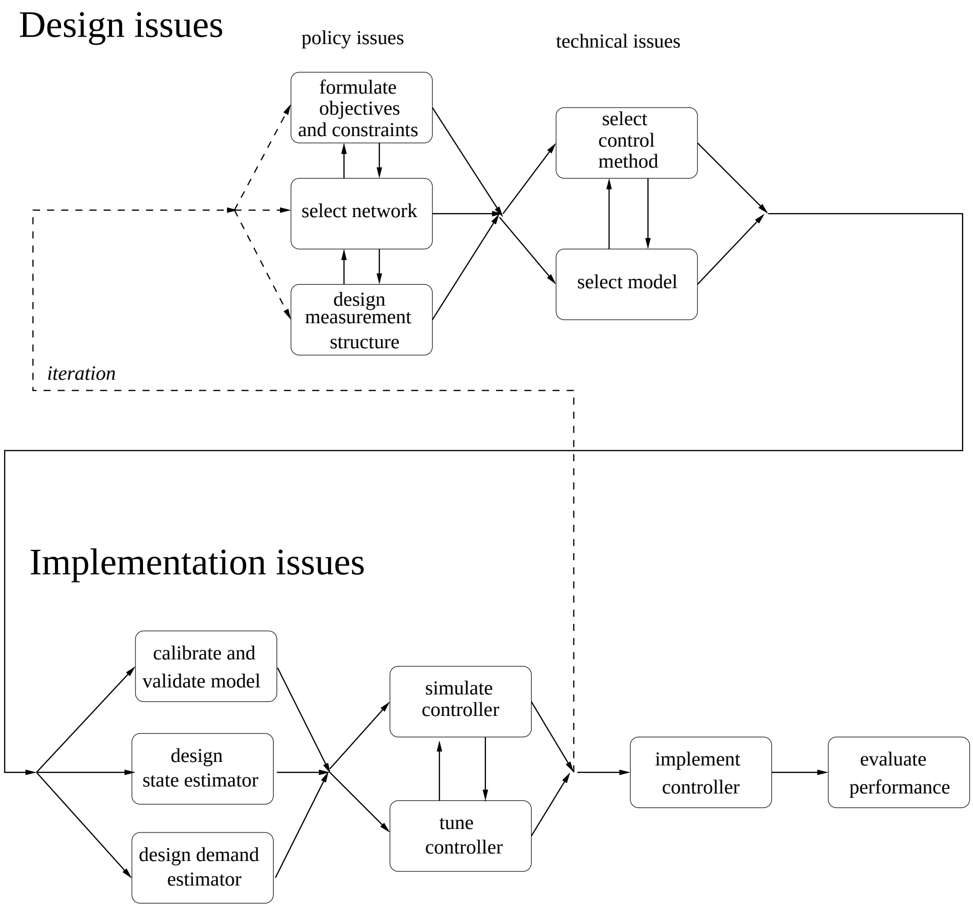


Figure 5.2: Overview of the controller development process.

5.3 Design issues

In this section we will further discuss the design issues that have been shortly introduced in Section 5.2.3 (see also Figure 5.2).

5.3.1 Formulation of the objectives and constraints

The objectives of a controller describe the goals that the controller should reach and have to be determined by the road authorities. Selecting an objective is often a trade-off between the desires of different parties (e.g., drivers, road managers, pedestrians, environmentalists). Accordingly, controller objectives can focus on different topics:

Efficiency: The available road capacity should be used as efficiently as possible. Possible objectives in this context are: reducing the total time spent, reducing the total travel time, increasing the throughput, and reducing delays.

Safety: Traffic controllers can improve safety by, e.g., reducing speeds, creating homogeneous flows, increasing intersection clearance times, and lowering flows in residential areas.

Environment: Traffic generates noise, and air pollution, and it consumes fuel. The environment benefits from, e.g., reducing the number of stops, smoothing the flows, decreasing the waiting time in the queues, and reducing the fuel consumption.

Location: When the traffic demand is so large that congestion cannot be prevented, the controller can try to put the congestion at a specific location where it causes the least problems, which can improve the situation in, e.g., residential areas and nature reserves.

In general, the overall objective of the controller will be formulated as a multi-objective criterion, composed of several of the objectives mentioned above. It is also possible to add penalties related to the expected behavior of the controller, which can target, e.g., variations in the control signal, and the traffic situation at the end of the prediction period. The use of multiple objectives in a controller results in a multi-objective optimization problem. Some methods to handle this kind of problems are: the weighted-sum method, the ϵ-constraint method, and the goal attainment method [108]. The weighted-sum method constructs a weighted sum of all the objectives, which is minimized with a standard constrained optimization method. The ϵ-constraint method selects a primary objective that is optimized, while the other objectives are included in the optimization problem via constraints on their values. The goal attainment method selects a target value for each objective, and minimizes the weighted deviation from the selected targets.

Another way to implement the requirements resulting from traffic policies, such as service levels, protection/safety of traffic participants, safety around schools, etc., is to formulate them as constraints for the optimization problem. This results in, e.g., maximum or minimum values for travel times, flows, speeds, intersection clearance times, or queue lengths. It is also possible to formulate physical constraints for the controller, that consider the limitations of the control measures and can result in, e.g., minimum or maximum values of the control signal.

5.3.2 Selection of the network

The decision to develop an advanced traffic controller is often induced by a traffic network in which a problem appears. However, the extent of the network that should be controlled is not always evident. Some problems can be solved within a small network, while others require a larger area to be solved efficiently. The extent of the required network can depend on, e.g., the ratio between local traffic and long distance traffic, the available measurements and their locations, the available traffic control measures, and the area on which the effects of the control measures appear. For some guidelines to select the network size available literature in the area of hierarchical control can be used, see, e.g., [77], where large systems are divided into subsystems based on the influence that parts of the system have on each other.

5.3.3 Design of the measurement structure

After the network has been selected, the measurement structure should be designed. This includes selecting the measurement technology, the locations of the measurements, the communication structure, the storage database, and the method for data polishing.

Often used measurement technologies for traffic networks include, e.g., pneumatic sensors, radar detectors, infra-red sensors, video cameras, or inductive loops [84]. The most commonly used sensors are inductive loop detectors. These detectors consist of inductive loops in the pavement, and measure the presence of a vehicle. They typically count the number of passing vehicles and average this over a time span (between 1 and 15 minutes). When double loops are used the speed of each vehicle can be determined. Pneumatic sensors located on the road can detect the presence of a vehicle. They are cheap but they are wearing fast, and thus they are mainly used for temporary measurements. Radar detection determines the presence and speed of vehicles via radar waves. These detectors are mainly used to determine the speed of vehicles. Infrared detectors determine the presence of a vehicle using infrared light. There are passive sensors which determine the radiations of the vehicles, and active sensors that send out a pulse and determine whether there is a vehicle based on the reflection of this pulse. Video images can be used to measure the traffic flows as well. The advantages of video imaging are that many different measurements can be obtained, e.g., space mean speeds, occupancy[1], vehicle positions, and vehicle types. The disadvantages are the sensitivity to rain, mist, or snow, and the relatively high costs due to the image processing that necessary to obtain the desired information (speeds, densities) from the video images.

Further, the locations of the detectors should be determined. On freeways, detectors are often located every 500 meters, and near bottlenecks such as on-ramps, off-ramps, lane-drops, and weaving areas. In urban areas, queue length detectors can be located at controlled intersections, and at the beginning and end of each link the number of entering and leaving vehicles respectively can be measured.

Then, the communication structure should be selected. The detectors and controllers can exchange data with their neighbors, or can cummunicate with a central controller. The obtained data should be stored in a database. The structure of the database and the desired contents should be determined.

At last, before measured data can be used, it should be polished [99]. Methods should be developed to remove outliers and sensor failures, and to address the uncertainty of the obtained measurements.

Since measurements form a basis for the controller design, and thus significantly influence the controller performance, we will investigate the effect of measurement methods more extensively in Section 5.5.

5.3.4 Selection of the control method

To determine the settings for the control measures, a control method should be selected. In the area of traffic control there exist methods that use no models, of which ALINEA [123] is the most well known. Further, there are methods based on fuzzy learning or neural networks,

[1] The occupancy is the percentage of time that the detector is occupied, which is representive for the density at the location of the detector.

see e.g., [35, 119, 188]. In this chapter however, we focus on model-based approaches. Examples of methods that use such an approach are presented in, e.g., [24, 49, 64, 76, 89, 129].

When the model-based approaches are used for optimal control (and also for MPC), an optimization algorithm should be selected to determine the optimal values for the control measures. Which algorithm should be selected depends on the type of optimization problem, which in its turn depends on the selected model, the objective function, and the constraints. For linear or convex problems many algorithms are available that will yield the global optimum [12, 126, 133]. However, a traffic control problem is nearly always non-convex and nonlinear, and can thus have many local optima. Hence, a global optimization method is required such as genetic algorithms, simulated annealing, pattern search, or multi-start local optimization [19, 44, 53, 61, 126]. These algorithms cannot guarantee that the global optimum is obtained, but they usually can obtain acceptable values. The use of these algorithms however increases the computation time, which is undesired for on-line computations. The selection of an optimization algorithm is thus based on the trade-off between the accuracy of the solution and the required computational effort.

5.3.5 Selection of the model

Many models are available for the use in model-based traffic controllers. An overview of traffic flow modeling in general is given in, e.g., [41, 72, 124]. Traffic models can be divided in categories based on the properties of the models. First of all the modeled application can be used as criterion, e.g., traffic flow models [40, 68, 106, 115], travel time models [26, 175], and traffic assignment models [15, 56, 130]. Second, the models can be stochastic [27, 101], or deterministic [15, 106]. Third, the models can be grouped based on the level of detail. Three categories that can be distinguished are:

Microscopic models that describe the behavior of individual vehicles in relation to the other vehicles and the infrastructure. Examples of commercially available models are Paramics [137], Vissim [136], and Aimsun [7], while an overview of more theoretical models is given in [72, 124].

Mesoscopic models that describe the traffic in probabilistic terms, using probability distribution functions, see for example the gas-kinetic model of [71]. Some mesoscopic models use a mix between detailed descriptions of important properties and a more general overall formulation, see, e.g., [31, 78].

Macroscopic models that describe the traffic flows using aggregated values, e.g., average speeds, and average densities. Early macroscopic models are formulated in [96, 139]. More recent models are METANET [106], INDY [15], and the Cell Transmission Model [40]. An overview of macroscopic models is presented in [125].

When a model should be selected for a model-based controller, attention should be paid to the features that are modeled. All features that are important for the controller should be modeled, including, e.g., traffic flows, influence of control actions, and properties affecting the objective of the controller. Further, the required computational effort should be taken into account. For a controller, a specific time is available for simulation of the model. The model should be able to run within this time. The accuracy and the computation time cannot be optimized at the same time, which makes a trade-off between the two criteria necessary. For on-line traffic controllers often macroscopic models are selected because they yield a reasonable accuracy within an acceptable computation time.

5.4 Implementation issues

When the design of the controller has been completed, more practical issues should be investigated. All these issues require measurement data. The design of the measurement structure has already been discussed in Section 5.3.3, and in Section 5.5 we will focus on the influence of speed measurement methods. In this section we present implementation issues that involve the use of measured data sets: calibration and validation, state estimation, demand estimation, and performance evaluation.

For the calibration and validation procedure the whole set of measurements is divided into two parts. One part is used for the calibration, and the other part for the validation. These data sets should be gathered under free-flow conditions as well as under congested conditions, in order to increase the persistency of exitation of the dataset, see [99].

State estimation requires real-time measurements of the system, while demand estimation can be done based on real-time measurements, or based on a dataset with historical measurements. For performance evaluation a set of data of the original situation is necessary, just as a data set obtained after the tuning and implementation of the controller.

5.4.1 Calibration and validation

Calibration is the process of selecting values for the parameters θ of a model, such as, e.g., the critical density, the desired speed, or the reaction time. The optimal parameter set minimizes the difference between the measured output (y^{measured}) and the output predicted by the model ($y^{\text{predicted}}$):

$$\min_{\theta} ||y^{\text{measured}} - y^{\text{predicted}}(\theta)||_2^2$$

Considering traffic flow models, macroscopic models are relatively easy to calibrate due to the limited number of variables. However, for large networks the required computation time can also increase up to the point where the problem becomes intractable. Calibration of macroscopic traffic models is described in, e.g., [39]. The authors of [39] consider an input that consists of the traffic demand, and the outputs are the measurements performed in the traffic network. For the nonlinear parameter identification problem the least-squares output error method is used, combined with the complex optimization algorithm of Box [20] for constrained problems. An automated calibration procedure for macroscopic traffic flow models is described in [116]. This procedure uses the Nelder-Mead Simplex algorithm to determine parameter values that optimize the total error between model output and measured data.

The calibration of microscopic models is more elaborate due to the large number of parameters that (in principle) can differ for each vehicle type. The calibration of microscopic models is considered in, e.g., [23, 38], where different available calibration methods are compared.

In modern traffic surveillance systems calibration can be performed on-line (and is also called parameter estimation), see, e.g., [3, 128, 178]. The advantage of this is that the

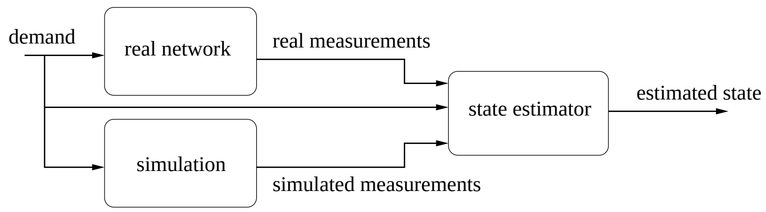


Figure 5.3: The state estimation procedure.

difference between the predictions and the real traffic situation will be as small as possible. For the on-line calibration the measurements of the real network during the last period (for example the last 15 minutes) are compared with the values that are predicted by the model. The difference between the two is minimized by optimizing the model parameters. These parameters are then used in the model, until a calibration for the next period is completed. An even more general tool that includes on-line calibration, traffic prediction, travel time estimation, queue length estimation, and incident detection is described in [180].

When a model is calibrated, the next step is to validate the model to determine the quality of the obtained parameters. During the validation, the parameters that are obtained during the calibration are implemented in the model. Then the model is used to predict the traffic variables corresponding to a different data set than the one used for the calibration. The difference between the simulation data and the real data set for the considered situation gives an indication of the correctness of the obtained parameter values, and of the generalizability of the simulation results obtained with the selected parameter values.

5.4.2 State estimation

In order to be able to make predictions with a model, the current state of the network should be known. This state is determined based on the available measurements, and is then used as initial state for the model predictions. In Figure 5.3 the general state estimation procedure is shown. The measurements of the real network are compared with the measurements that are obtained with the simulation. Based on this comparison, the state estimator determines the estimated state. State estimation is in most cases done based on Kalman filtering or one of its extensions [79]. For linear models, a Kalman filter adapts the estimated state in such a way that the mean of the error between prediction and measurements is minimized. For non-linear models, an extended Kalman filter should be used, which linearizes the model around the current estimate at every step. For traffic flow models this method is described in, e.g., [60, 178]. A more recent method that can be used for state estimation is based on particle filtering, see [67, 109]. Particle filters use probabilistic models, and start with a distribution of possible states. For all of these possible states the likelihood that it corresponds to the current state is computed, based on a measurement function and on Bayes' rule. With each new set of measurements these likelihoods are updated. The most likely state is selected to be the estimated state.

5.4.3 Demand estimation

To make a prediction of future traffic flows, the future demand must be known. This demand can be obtained based on Origin-Destination (OD) matrices, or based on upstream measurements:

OD-matrices OD-matrices contain the demand (in veh/h) from each origin to each destination. OD-matrices can be determined off-line, and can afterwards be adapted in an on-line setting. Off-line OD-matrices are obtained from, e.g., surveys, historical measurements, and estimations based on the surroundings (residential areas, shopping centers, business areas) and the expected amount of drivers that want to visit these places [30, 92]. On-line updating of the OD-matrix is done based on measurements using a method that is similar to methods for on-line calibration. When the OD-matrix is determined, a traffic assignment algorithm can be used to divide the traffic over the network. This results in the expected flows on each link. This procedure is described in, e.g., [4, 191].

Upstream measurements Traffic flows measured upstream of the controlled road section will arrive at the controlled section with a delay approximately equal to the expected travel time from the measurement location to the beginning of the controlled stretch. With measurements of the upstream flows estimations of the flows at the controlled road section can be made, as presented in [120]. The accuracy of the estimation is influenced by the distance between the upstream measurement and the controlled location, and by the number of intersections, on-ramps, and off-ramps on this stretch.

5.4.4 Controller tuning

The controller also has parameters that have to be tuned. For MPC controllers important parameters are the horizons, and the possible weights in the cost function. In the process industry methods for tuning MPC controllers have been developed [1, 94, 172]. There are methods for off-line as well as on-line tuning. However, most of these methods focus on linear MPC and thus cannot be used for advanced traffic controllers.

Controller tuning starts with the selection of initial parameters, which are manually adapted based on simulation results or real measurements. Related to traffic control, the initial values for the parameters can be selected as follows. An initial value for the prediction horizon is the time that a vehicle needs to drive through the selected network. This ensures that all the effects of the control actions on this vehicle are taken into account. The length of the control horizon mainly depends on the computational effort required to optimize the cost function. A longer control horizon leads to more parameters, which leads to longer computation times. However, when the control horizon is too short, the possible impact of the control actions will decrease. Some more detailed tuning rules for the horizons are presented in [64]. The weight of each part of the cost function should be based on the relative importance of the different parts, which should be determined by the road authorities. The weights have to be normalized, which can be obtained by dividing each part of the cost function by its nominal value.

5.4.5 Performance evaluation

To evaluate the performance of the controller, the controlled situation should be compared with the uncontrolled situation. This means that first, before the deployment of the controller, the initial situation should be measured. Then, when the controller is installed a period should be selected during which the traffic can adapt to the controller. During this period the behavior of the drivers can change, and if necessary the controller should be changed too. After this period the performance of the controller can be determined. This can be done by comparing initial measurements (without the controller) with measurements in the controlled situation. For this comparison a performance evaluation function has to be defined. The most logical choice for such a function is based on the cost function that is selected for the controller. This function can be extended with a penalty for constraint violations. For each situation (uncontrolled/controlled) the costs are computed, and the relative difference in the costs between the two situations represents the performance of the controller. When more controllers are compared in a case study, for all controllers the resulting traffic situation should be measured. These measurements can then be used to compare the performance of the different controllers.

5.4.6 Other issues

We will now mention some topics that are also relevant for the implementation of advanced controllers, but that will not be discussed in detail in this chapter:

Fault detection and fault tolerant control: The availability and the failure probability of the equipment used is important for the functioning of the controller. Missing measurements and wrong representations of the control signals can significantly influence the performance of the controller [14, 135]. By monitoring, the equipment failures can be noticed (or even predicted) and the controller can take the effects of the failures into account. This allows the controller to reduce the influence of the failure and to prevent a large decrease in the performance of the network.

Robustness: A model is never an exact representation of reality. The sensitivity of the controller for errors in the model structure should be accounted for, as well as the other uncertainties in the design, such as the error in the demand prediction, the state estimation, and the values of the model parameters. The effect of errors and uncertainties can be reduced by, e.g., including demand prediction, using a smaller controller time step to decrease the deviation between the real state and the predicted state, by on-line calibration of the parameters, and by using robust control techniques [93, 181].

Stability: The control actions influence the traffic flows. The control actions should lead to a stable traffic situation, without fast fluctuating control signals [141, 189]. Stable traffic situations are situations in which the traffic flows will stay around the same level, even if small disturbances of the flows occur. Fast fluctuating control signals can result in a fast changing traffic situation and thus in unstable traffic flows. Fast changing control signals can be prevented by including a penalty on changes in the control signal in the cost function, or by using larger thresholds with respect to the reactions on changes in the measurements.

5.5 Investigation of speed measurements

Since a major part of the steps described above is based on measurements, we now investigate the influence of different measurement averaging methods. We will first discuss inductive loop detectors, and next describe different averaging methods for speed measurements. An overview of symbols used in this section is given in Appendix 5.A.

Remark 5.1 Within this section, we consider four different time steps: the controller time step T_{c} with index k_{c}, the sample (measurement) time step T_{s} with index k_{s}, the simulation time step T with index k, and the step between two images that are obtained with video imaging T_{t} with index k_{t}. These time steps are related as follows: $T_{\mathrm{c}} = M_{\mathrm{c}} T_{\mathrm{s}}$, $T_{\mathrm{s}} = M_{\mathrm{s}} T$, and $T_{\mathrm{s}} = M_{\mathrm{t}} T_{\mathrm{t}}$, where M_{c}, M_{s}, and M_{t} are integers. □

5.5.1 Speed measurements

In Section 5.3.3 we have discussed several measurement technologies for traffic networks. In this section we will consider inductive loop detectors, and focus on speed measurements. When double loop detectors are used the speed of each vehicle can be determined. Now consider detector d. The number of vehicles that are observed at this detector during the period $[k_{\mathrm{s}} T_{\mathrm{s}}, (k_{\mathrm{s}}+1) T_{\mathrm{s}})$ is equal to $N_d(k_{\mathrm{s}})$, and we index them as $1, 2, \ldots, N_d(k_{\mathrm{s}})$. The obtained individual vehicle speeds measured by the detector d are denoted by $u_{d,n}(k_{\mathrm{s}})$, where n is the vehicle index. With the obtained measurements the flow of the traffic can be obtained by dividing the number of observed vehicles $N_d(k_{\mathrm{s}})$ by the sampling time interval T_{s}. Hence, the flow passing the detector can be determined using

$$q_d^{\mathrm{sample}}(k_{\mathrm{s}}) = \frac{N_d(k_{\mathrm{s}})}{T_{\mathrm{s}}} \quad . \tag{5.1}$$

The density follows from the flow $q_d^{\mathrm{sample}}(k_{\mathrm{s}})$, the mean speed $v_d^{\mathrm{sample}}(k_{\mathrm{s}})$, and the number of lanes λ, as

$$\rho_d^{\mathrm{sample}}(k_{\mathrm{s}}) = \frac{q_d^{\mathrm{sample}}(k_{\mathrm{s}})}{v_d^{\mathrm{sample}}(k_{\mathrm{s}})\lambda} \quad . \tag{5.2}$$

In the next section we will formulate several methods to determine the mean speed $v_d^{\mathrm{sample}}(k_{\mathrm{s}})$.

5.5.2 Various speed averaging methods

In the field of freeway traffic flow modeling, two representations of mean speeds are often used: *time mean* speed and *space mean* speed [41, 105]. Traffic flow models that are used in model-based controllers often use the space mean speed which is the average speed of all vehicles present in the considered freeway stretch at a specific time instant. This space mean speed is used since most models describe the average traffic conditions on such a stretch. Video images can be used to obtain space mean speeds [43]. Unfortunately, the loop detectors that are often present in the road network typically return time mean speeds. Time mean speeds are based on measurements at a specific location averaged over a certain time span. When the time mean speeds are measured, it is impossible to calculate the exact space mean speeds due to the relatively low number of measurement locations. However, they can be approximated using various averaging methods, see [41, 138, 175]. In this chapter we will investigate the influence of different averaging methods on the performance of a model-based speed limit controller.

We discuss six different methods to calculate mean speeds. Within the description we assume that in the sample period $[k_\mathrm{s}T_\mathrm{s},(k_\mathrm{s}+1)T_\mathrm{s})$ the subsequent vehicles passing the detector d are numbered $1,2,\ldots,N_d(k_\mathrm{s})$. For the ease of notation we do not mention d in the equations, but note that the values can be determined for every detector in the freeway network.

Time mean speed The time mean speed is calculated using the **arithmetic mean** of the $N(k_\mathrm{s})$ locally measured vehicle speeds $u_n(k_\mathrm{s})$, measured over the sampling time interval $[k_\mathrm{s}T_\mathrm{s},(k_\mathrm{s}+1)T_\mathrm{s})$ as [41]:

$$v^{\mathrm{tms}}(k_\mathrm{s})=\frac{1}{N(k_\mathrm{s})}\sum_{n=1}^{N(k_\mathrm{s})}u_n(k_\mathrm{s}) \tag{5.3}$$

Estimated space mean speed For estimating the space mean speed, the **harmonic mean** of the locally measured vehicle speeds is used by Daganzo [41], given by

$$\hat{v}^{\mathrm{sms}}(k_\mathrm{s})=\left(\frac{1}{N(k_\mathrm{s})}\sum_{n=1}^{N(k_\mathrm{s})}\frac{1}{u_n(k_\mathrm{s})}\right)^{-1} \tag{5.4}$$

Geometric mean speed Besides the arithmetic and harmonic mean, there is a third 'classic' Pythagorean mean, namely the **geometric mean** [132]. This mean can be calculated as

$$v^{\mathrm{geo}}(k_\mathrm{s})=\sqrt[N(k_\mathrm{s})]{\prod_{n=1}^{N(k_\mathrm{s})}u_n(k_\mathrm{s})} \tag{5.5}$$

Estimated space mean speed using the instantaneous speed variance A method to estimate the space mean speed, based on locally measured vehicle speeds $u_n(k_\mathrm{s})$, is proposed by Van Lint [175]. The method uses the empirical relation between the time mean speed $v^{\mathrm{tms}}(k_\mathrm{s})$ (defined by (5.3)), and the space mean speed $v^{\mathrm{sms}}(k_\mathrm{s})$ (as described in [182]), which is given by

$$v^{\mathrm{tms}}(k_\mathrm{s})=\frac{\sigma_\mathrm{i}^2(k_\mathrm{s})}{v^{\mathrm{sms}}(k_\mathrm{s})}+v^{\mathrm{sms}}(k_\mathrm{s}) \tag{5.6}$$

where $\sigma_\mathrm{i}^2(k_\mathrm{s})$ is the variance of the instantaneously measured vehicle speeds. Since the instantaneous speed variance $\sigma_\mathrm{i}^2(k_\mathrm{s})$ cannot be determined exactly by local measurements, the following estimation used:

$$\hat{\sigma}_\mathrm{i}^2(k_\mathrm{s}) = \frac{1}{2N(k_\mathrm{s})}\sum_{n=1}^{N(k_\mathrm{s})}\frac{\hat{v}^{\mathrm{sms}}(k_\mathrm{s})}{u_n(k_\mathrm{s})}\left(u_{n+1}(k_\mathrm{s})-u_n(k_\mathrm{s})\right)^2$$

where $\hat{v}^{\mathrm{sms}}(k_\mathrm{s})$ is given by (5.4). The estimated space mean speed using the instantaneous speed variance then becomes

$$\hat{v}^{\mathrm{sms},\hat{\sigma}_\mathrm{i}}(k_\mathrm{s})=\frac{1}{2}\left\{v^{\mathrm{tms}}(k_\mathrm{s})+\sqrt{\left(v^{\mathrm{tms}}(k_\mathrm{s})\right)^2-4\hat{\sigma}_\mathrm{i}^2(k_\mathrm{s})}\right\} \tag{5.7}$$

Estimated space mean speed using the local speed variance In [138] an estimate of the space mean speed is proposed that is based on the time mean speed and variance of the locally measured speeds. The space mean speed estimate using the local speed variance is determined by

$$\hat{v}^{\text{sms},\sigma_1}(k_s) = v^{\text{tms}}(k_s) - \frac{\sigma_1^2(k_s)}{v^{\text{tms}}(k_s)} \tag{5.8}$$

where $v^{\text{tms}}(k_s)$ is computed using (5.3), and

$$\sigma_1^2(k_s) = \frac{1}{N(k_s)} \sum_{n=1}^{N(k_s)} \left(u_n(k_s) - v^{\text{tms}}(k_s)\right)^2$$

Time average space mean speed Using, e.g., video images [43] it is possible to obtain the space mean speed $v^{\text{sms,s}}(k_t)$ on a freeway segment for every imaging time k_t (recall that $T_s = M_t T_t$). Since the previous mean speed methods are all based on a time period $[k_s T_s, (k_s{+}1)T_s)$, we will take a time average of the obtained space mean speeds during this period:

$$\bar{v}^{\text{sms}}(k_s) = \frac{1}{M_t} \sum_{k_t = M_t k_s}^{M_t(k_s+1)} v^{\text{sms,s}}(k_t) \tag{5.9}$$

5.6 Case study

We now illustrate the effect of different speed averaging methods on the performance of a speed limit controller. The six variants for calculating the mean speed discussed in Section 5.5.2 are used to calibrate the model that is used by the controller, resulting in six different parameter sets. For each of these parameter sets we determine how good the model predicts future traffic states, and we determine the difference between the measured and the predicted total time spent (TTS).

In the remainder of this section we first introduce the network and traffic scenario, and then develop a variable speed limit controller according to the steps described in Sections 5.3 and 5.4. Next, we use the six different averaging methods within the speed limit controller, and perform a simulation case study in which we first calibrate the prediction model, and then compare the performance of the corresponding controllers.

5.6.1 Network and traffic scenario

For the traffic network, a part of the Dutch freeway A12 is selected, see Figure 5.4. The total length of the considered stretch is 17422 m. There are two on-ramps, near Veenendaal and near Maarsbergen. The major cause of delay on this stretch are shock waves. Shock waves are traffic jams that propagate in the opposite direction of the traffic flows, and often emerge from on-ramps and other types of bottlenecks. The outflow of a shock wave is usually about 70% of the freeway capacity [83], and resolving shock waves can significantly improve the freeway traffic flow.

The selected freeway stretch is modeled with Paramics v5.1 from Quadstone [137], a microscopic traffic simulation model. The resulting model will be used as representation of

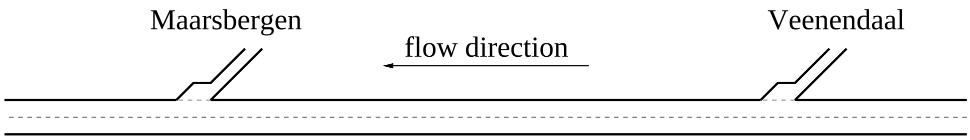


Figure 5.4: Network of the case study, part of the Dutch freeway A12.

the real world. Since Paramics is a stochastic model, each simulation should be performed several times to obtain statistically significant results. For our model, we perform 5 runs for each simulation.

During the simulation of the freeway a shock wave is introduced by simulating an incident downstream of the considered stretch. One vehicle is stopped downstream of the on-ramp near Maarsbergen for a period of 5 minutes, during which one of the two lanes is blocked. This will create a traffic jam, which expands while the lane is blocked, and the shock wave starts to move upstream when both lanes are accessible again. The traffic demand $q^{\text{dem}}(k)$ on the freeway is set to a constant value of 4400 veh/h, at which a shock wave will remain existent in the network when no control is applied.

We start the simulation with a full network, and simulate a period of one hour. Figure 5.5 shows the simulated measurements on the network. On the horizontal axis, the time is shown, and on the vertical axis, the locations are given. The first segment is at the bottom of the figure, and the last segment at the top. In the top subplot, the measured mean speeds are given. Lighter colors represent higher mean speeds. The shock wave is clearly shown as the thick, dark stripe going upstream as time elapses. Also in the density plot (the middle subplot), the shock wave is clearly visible as the thick, light stripe representing high densities. The bottom subplot shows the flow, where it can be seen from the dark color that due to the shock wave the flow decreases. Note that the thin dark stripes that are going downstream as time elapses, are caused by differences in speed between individual vehicles.

5.6.2 Design of a dynamic speed limit controller

The influence of different averaging methods on performance of a controller will be illustrated with a variable speed limit controller. We will now follow the design steps for the controller as described in Sections 5.3 and 5.4. Note however that since the case study is performed in a simulation environment, not all steps are necessary.

Formulation of the objectives and constraints

The policy objective of the controller is selected to be the reduction of the travel time on the freeway stretch. The long travel time in the uncontrolled situation is mainly due to shock waves. Shock waves can be reduced or dissolved by applying dynamic speed limits on the freeway, see [64]. Traffic upstream of the shock wave can be slowed down, thereby limiting the inflow to the shock wave. Since the outflow of the shock wave will stay constant, this will reduce the length of the shock wave, and can even dissolve it. This effect can be reached by selecting the total time spent (TTS) as cost function, which should be minimized.

It is also possible to define constraints for the controller. A possible policy constraint

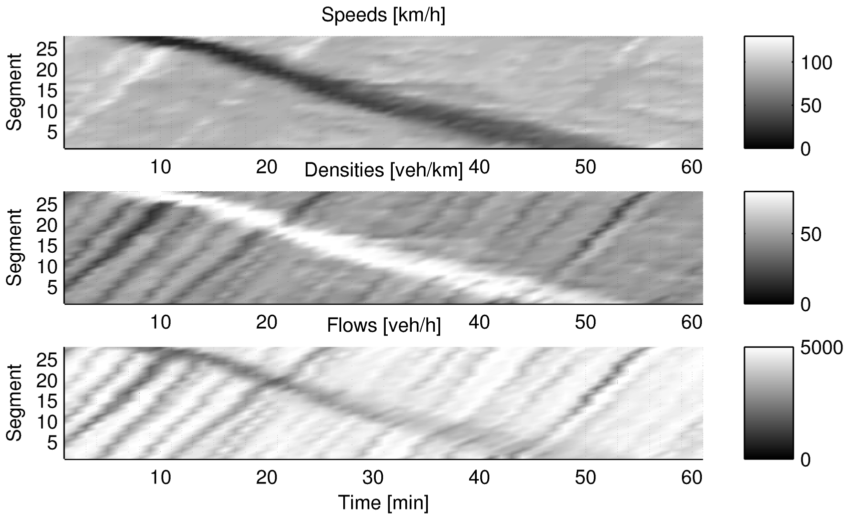


Figure 5.5: Traffic condition without control.

that we can select is a maximum queue at the origins:

$$w_o(k) \leq w_o^{\max} \quad \forall o \in O$$

where w_o is the queue length at origin o, and O is the set of all origins in the network. When we apply speed limit control, physical constraints on, e.g., the speed limits can be given by:

$$v^{\min} \leq v^{\text{control}}(k_\text{c}) \leq v^{\max}$$

where $v^{\min}$ and $v^{\max}$ are respectively the minimum and maximum allowed speed, and $v^{\text{control}}(k_\text{c})$ is the speed limit applied at control time step k_c. For the case study we use minimum and maximum values for the speed limits of 40 km/h and 120 km/h respectively. The speed limits vary between these bounds. Further, a penalty on signal variations is added to the cost function to reduce speed limit oscillations.

When the speed limits are applied to the Paramics model they are rounded to steps of 10 km/h to mimic reality more closely. This increases the mismatch between the predicted and measured states, which decreases the performance of the controller. However, in Paramics the actual speed of drivers will vary stochastically around the presented speed limit, which decreases the negative effect of the rounding operation on the performance of the controller. Moreover, in [66] it was found that when using the round operation the speeds that are obtained during a simulation are approximately equal to the speeds that are obtained during a simulation where no rounding is applied.

Selection of the network

The complete network is shown in Figure 5.6. Measurements will be taken on the whole stretch, over the length of 17422 m. We will control the part between the on-ramps at Veenendaal and Maarsbergen, which results in a controlled stretch of 9775 m.

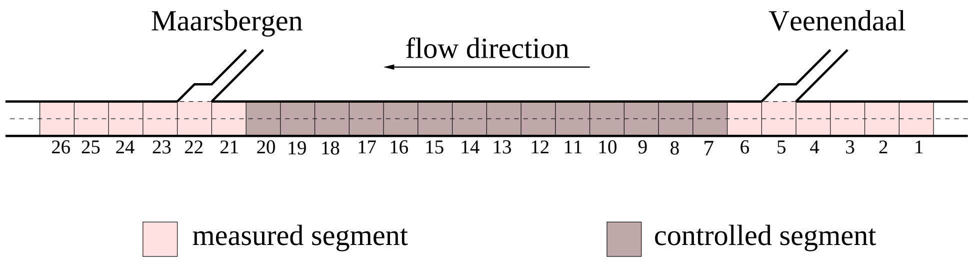


Figure 5.6: Detailed network of the case study.

Obtaining measurements

To obtain the measurements we use loop detectors. The loop detectors in the micro-simulator are placed at the locations of the existing loop detectors on the freeway. The distances between subsequent loop detectors is varying between 545 m and 810 m.

Selection of the control method

As control method we select model predictive control (MPC), as explained in Section 5.2.2. This control method requires a model and an optimization algorithm. The model selection will be described below. As optimization algorithm we select sequential quadratic programming (SQP) [19] as implemented in the MATLAB function `fmincon` [154], with a multi-start approach. This algorithm is selected since it can handle the nonlinear, non-convex, bounded optimization problem that should be solved by the MPC controller. At each controller step k_c, 16 distinct initial value sets are used.

Model selection

For the speed limit controller we will use the macroscopic traffic flow model METANET, as described in [106]. The METANET model introduces the division of a freeway network into multiple links and segments. Each freeway link m is divided into several segments i, see Figure 5.7. For the case study, we consider three links, with in total 26 segments, see Figure 5.6. Segments 1 to 5 belong to link 1, segments 6 to 21 belong to link 2, and segments 22 to 26 belong to link 3. The segments are chosen such that the loop detectors are near the downstream boundary, in order to obtain accurate measurements of the outflow $q_{m,i}(k_s)$ of the segments. This means that the segments have lengths $L_{m,i}$ and thus that each segment should have its own values for the model parameters, which is not conform the original formulation of METANET. However, for simplicity, in this case study we assume that within the given link all segments use the same value for the model parameters. The on-ramp near Veenendaal is connected to segment 5, and the on-ramp near Maarsbergen to segment 22. Segments 7 to 20 will be controlled via variable speed limits.

Within the METANET model the state of segment i of link m during the period $[kT,(k+1)T)$ is given in terms of the density $\rho_{m,i}(k)$, mean speed $v_{m,i}(k)$, and outflow $q_{m,i}(k)$ of the segment. Here k denotes the simulation step, with simulation time interval T. Each segment i of link m has a length $L_{m,i}$, while the number of lanes λ_m is equal for all segments in link m.

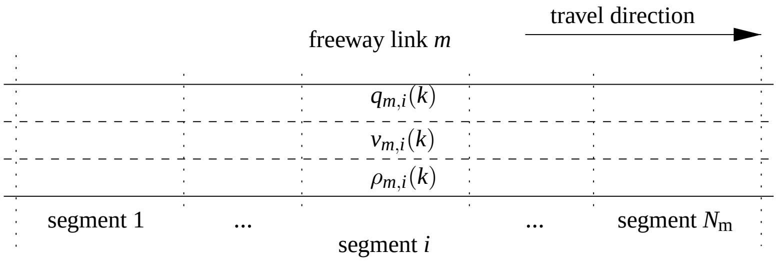


Figure 5.7: In the METANET model, a freeway link is divided into segments.

The METANET model equations are given by [106]:

$$
\begin{aligned}
q_{m,i}(k) &= \rho_{m,i}(k) v_{m,i}(k) \lambda_m, \\
\rho_{m,i}(k+1) &= \rho_{m,i}(k) + \frac{T}{L_{m,i}\lambda_m}\left(q_{m,i-1}(k) - q_{m,i}(k)\right), \\
v_{m,i}(k+1) &= v_{m,i}(k) + \frac{T}{\tau_m}\left(V\left(\rho_{m,i}(k)\right) - v_{m,i}(k)\right) \\
&\quad + \frac{T}{L_{m,i}} v_{m,i}(k)\left(v_{m,i-1}(k) - v_{m,i}(k)\right) \\
&\quad - \frac{\eta_m T}{\tau_m L_{m,i}} \frac{\rho_{m,i+1}(k) - \rho_{m,i}(k)}{\rho_{m,i}(k) + \kappa_m},
\end{aligned}
\tag{5.10}
$$

and

$$
V\left(\rho_{m,i}(k)\right) = v_m^{\text{free}} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_m^{\text{crit}}}\right)^{a_m}\right],
\tag{5.11}
$$

where v_m^{free} is the free flow speed in link m, ρ_m^{crit} is its critical density (i.e. threshold between free and congested traffic flow), and τ_m, η_m, κ_m, and a_m are model fitting parameters.

At the locations of the on-ramps, the flows arriving at the freeway and at the on-ramp are added to determine the flow downstream of the on-ramp:

$$
q_{m,1}(k) = q_{\mu,N_\mu}(k) + q_r(k)
$$

where $q_{m,1}(k)$ is the flow entering the freeway link m downstream of on-ramp r, $q_{\mu,N_\mu}(k)$ the flow leaving the freeway link μ upstream of on-ramp r, and $q_r(k)$ the flow leaving on-ramp r. In order to account for the speed drop caused by merging phenomena at the on-ramps, the following term is added to (5.10) [89, 106]:

$$
-\frac{\delta_m T q_r(k) v_{m,1}(k)}{L_{m,i}\lambda_m \rho_m^{\text{crit}}}
\tag{5.12}
$$

where δ_m is a model parameter of link m and $v_{m,1}(k)$ the speed at the first segment of the freeway link downstream of the on-ramp. The virtual entering speed $q_{m,0}(k)$ of leaving link m downstream of the on-ramp is given by:

$$
v_{m,0}(k) = \frac{v_{\mu,N_\mu}(k) q_{\mu,N_\mu}(k) + v_r(k) q_r(k)}{q_{\mu,N_\mu}(k) + q_r(k)} \quad ,
$$

where $v_r(k)$ is the speed of the vehicles that leave on-ramp r. This virtual entering speed is used in the speed update equation (5.10) to compute the speed of the traffic that enters the first segment of link m.

Origins are modeled using a simple queue model. The number of vehicles w_o in the queue at origin o evolves as follows:

$$w_o(k+1) = w_o(k) + T\left(q_o^{\mathrm{dem}}(k) - q_o(k)\right), \tag{5.13}$$

where $q_o^{\mathrm{dem}}(k)$ is the demand at simulation step k, and $q_o(k)$ is the outflow given by

$$q_o(k) = \min\left\{q_o^{\mathrm{dem}}(k) + \frac{w_o(k)}{T},\, Q_o\frac{\rho^{\max} - \rho_{m,1}(k)}{\rho^{\max} - \rho^{\mathrm{crit}}m}\right\}, \tag{5.14}$$

where Q_o is the capacity of origin o under free flow conditions, and $\rho^{\max}$ is the maximum density. For on-ramps (5.13) and (5.14) are also valid.

In [64] some extensions are described which, among others, formulate the effect of variable speed limits by replacing (5.11) by:

$$V\left(\rho_{m,i}(k)\right) = \min\left(v_m^{\mathrm{free}} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_m^{\mathrm{crit}}}\right)^{a_m}\right], (1+\alpha)v_{m,i}^{\mathrm{control}}(k_{\mathrm{c}})\right) \quad . \tag{5.15}$$

The parameter α expresses the obedience of the drivers with respect to the applied speed limit. When the speed limits are enforced α will be smaller since drivers will not exceed the speed limits as much as without enforcement. The index k_{c} counts the control time steps, as introduced in Remark 5.1. The interval of simulation time steps that correspond to the control time step k_{c} is given by $[k_{\mathrm{c}}M_{\mathrm{c}}M_{\mathrm{s}}, (k_{\mathrm{c}}+1)M_{\mathrm{c}}M_{\mathrm{s}}) - 1]$.

As cost function we selected the total time spent (TTS), see Section 5.6.2. When using the METANET model, the TTS can be computed as follows:

$$J^{\mathrm{TTS}}(k_{\mathrm{c}}) = T\sum_{k=k_{\mathrm{c}}M_cM_s+1}^{(k_{\mathrm{c}}+N_{\mathrm{p}})M_cM_s}\left(\sum_{(m,i)\in\mathscr{M}}\rho_{m,i}(k)\lambda_m L_{m,i} + \sum_{o\in O} w_o(k)\right) \tag{5.16}$$

where $\mathscr{M}$ is the set of pairs (m,i) of link indices and the corresponding segment indices.

Calibration and validation

Calibration is done by off-line numerical optimization using an objective function given by

$$J^{\mathrm{cal}}(\theta) = \frac{1}{K_{\mathrm{c}} - N_{\mathrm{p}}}\sum_{k_{\mathrm{c}}=1}^{K_{\mathrm{c}}-N_{\mathrm{p}}}\ \sum_{k_{\mathrm{s}}=M_{\mathrm{c}}k_{\mathrm{c}}}^{M_{\mathrm{c}}(k_{\mathrm{c}}+N_{\mathrm{p}})}\ \sum_{k=M_{\mathrm{s}}k_{\mathrm{s}}+1}^{M_{\mathrm{s}}(k_{\mathrm{s}}+1)} \check{J}^{\mathrm{cal}}(\theta, k_{\mathrm{s}}, k) \quad , \tag{5.17}$$

where θ is the set of model parameters consisting of v_m^{free}, ρ_m^{crit}, τ_m, η_m, κ_m, and a_m for each of the three links, K_{s} is the number of sample steps for which measurement data is available, and where $\check{J}^{\mathrm{cal}}(\theta, k_{\mathrm{s}}, k)$ is given by:

$$\check{J}^{\mathrm{cal}}(\theta, k_{\mathrm{s}}, k) = \sum_{(m,i)\in\mathscr{M}}\left\{\left(\frac{v_{m,i}^{\mathrm{sample}}(k_{\mathrm{s}}) - \tilde{v}_{m,i}(k)}{\bar{v}(k_{\mathrm{s}})}\right)^2 + \left(\frac{\rho_{m,i}^{\mathrm{sample}}(k_{\mathrm{s}}) - \tilde{\rho}_{m,i}(k)}{\bar{\rho}(k_{\mathrm{s}})}\right)^2\right\},$$

where $\bar{v}(k_s)$ and $\bar{\rho}(k_s)$ are the average speed and density of the measured data from controller step k_c to k_c+N_p. The error in the predictions that are made by the controller at every control time step is computed, and then the average of the errors of all prediction periods is taken. A lower value of the objective function (5.18) means a better fit of the measured states $\{v_{m,i}^{\text{sample}}; \rho_{m,i}^{\text{sample}}\}$ by the predicted states $\{\tilde{v}_{m,i}; \tilde{\rho}_{m,i}\}$ reproduced by the traffic flow model.

Another option is to compare the two data sets with respect to the value of the cost function. When the cost function is selected to be the TTS (as in (5.16)), the difference between measured and predicted TTS can be determined to judge the performance of the parameter values. The error is given by

$$E^{\text{TTS}}(\theta) = \frac{1}{K_c - N_p} \sum_{k_c=1}^{K_c-N_p} \left| \frac{\tilde{J}^{\text{TTS}}(k_c) - J^{\text{TTS}}(k_c)}{J^{\text{TTS}}(k_c)} \right| \tag{5.18}$$

which gives the average percentage of mismatch between the TTS for the measured data J^{TTS}, and the TTS for the predicted data $\tilde{J}^{\text{TTS}}$ for the parameter set θ, and where K_c is the final control time step in the scenario.

This offline calibration of the METANET model is performed with the MATLAB function `fmincon` [154] which implements SQP, which is the same algorithm that will be used in the on-line controller. We use the algorithm in a multi-start configuration with 100 different initial values, which increases the probability of finding the global (or best local) optimum. To deal with the stochasticity of Paramics (see Section 5.6.1) we use 5 different random seeds. The cost functions $J^{\text{cal}}(\theta)$ and $E^{\text{TTS}}(\theta)$ are determined at each control time step, for the prediction that is made at this time step. For the calibration of the total model the average value $\bar{J}^{\text{cal}}(\theta)$ and $\bar{E}^{\text{TTS}}(\theta)$ over the results of all control time steps are determined.

The calibration is be performed for each of the six mean speed calculation methods described in Section 5.5.2 separately, which results in six different parameter sets as presented in Section 5.6.3.

State estimation

The state of the network consists of the average speed, average density, and average flow. The speeds are measured by loop detectors, and averaged with the different methods described in Section 5.5.2. The density and flow can be calculated from these measurements using (5.1) and (5.2), see Section 5.3.5. At every controller time step, a new estimation of the current state is obtained, based on the last available measurements.

Demand estimation

The Paramics model uses a demand of 4400 veh/h, which means that it randomly introduces vehicles, with a mean of 4400 veh/h. To make predictions with the METANET model, an estimation of this stochastic demand should be made. However, for simplicity we use the known average value of 4400 veh/h as estimation of the demand during the case study.

Controller tuning

For the controller we have selected the following parameters. The METANET model (5.11) uses a simulation time interval of T=10 s. This period is small enough to ensure that the vehicles cannot drive through a whole segment in one simulation step, and large enough to prevent unnecessary long computation times. The controller time interval equals T_c=60 s, which prevents fast switching between control values and which is a reasonable time to perform the on-line optimization. The prediction horizon equals N_p=20 steps, which is the time required to drive through the network when a shock wave is present. The control horizon is N_c=10 steps, which forms a trade-off between the required computation time and the detectability of the effects of the control actions.

Performance evaluation

Now we will illustrate the selection of the evaluation function, using the speed limit control example. Since the objective is to minimize the TTS, we determine the improvement in traffic conditions by comparing the obtained values for the TTS. However, only comparing the TTS based on the number of vehicles in the measured area is not a good measure, since the controller will not only increase the outflow, but also the inflow when the shock wave is resolved successfully. This is due to the fact that the incoming vehicles will not be blocked anymore, when the shock wave is dissolved. To take into account this change in the demand a different formulation of the TTS is used, based on the demand $q^{\text{dem}}(k)$ and the outflow $q^{\text{out}}(k)$ [64]:

$$J^{\text{TTS}} = TN_0M_cM_sK_c + T^2 \sum_{k=0}^{M_cM_sK_c-1} (M_cM_sK_c - k)\left(q^{\text{dem}}(k) - q^{\text{out}}(k)\right) \tag{5.19}$$

where N_0 is the initial number of vehicles in the measured area of the freeway.

5.6.3 Results

The results of the case study are presented here. First, we discuss the calibration of the various mean speeds, and then we simulate the network using three of the different averaging methods within the controller and compare the results obtained with the different methods.

Comparison of calibration results for the various mean speeds

We first perform the calibration as described in Section 5.6.2. The results are shown in Table 5.1. The average calibration errors $\bar{J}_{\text{cal}}$ computed with (5.18) are presented, and next the average error between the measured and predicted TTS $\bar{E}_{\text{TTS}}$ is shown, as obtained using (5.18).

The lower the values for $\bar{J}_{\text{cal}}$ and $\bar{E}_{\text{TTS}}$, the better the fit between the model predictions and the measurements. Based on the average calibration errors $\bar{J}_{\text{cal}}$ the estimated space mean speed using the instantaneous speed variance $\hat{v}^{\text{sms},\hat{\sigma}_i}$ gives the best result, followed by the time mean speed v^{tms} and the geometric mean speed v^{geo}. Based on the average error between measured and predicted TTS the geometric mean speed v^{geo} performs the best, followed by the time mean speed v^{tms} and estimated space mean speed using the instantaneous speed variance $\hat{v}^{\text{sms},\hat{\sigma}_i}$, which yields the same performance.

Since the time mean speed v^{tms}, the geometric mean speed v^{geo}, and the estimated space mean speed using the instantaneous speed variance $\hat{v}^{\text{sms},\hat{\sigma}_i}$ perform better than the other averaging methods for both criteria, we will implement these averaging methods in the speed limit controller.

Table 5.1: Mean speed performance for calibration.

		$\bar{J}_{\text{cal}}$	$\bar{E}_{\text{TTS}}$
Time mean speed	v^{tms}	38.9	5.5
Estimated space mean speed	$\hat{v}^{\text{sms}}$	42.0	5.6
Geometric mean speed	v^{geo}	39.4	5.1
Estimated space mean speed (instantaneous)	$\hat{v}^{\text{sms},\hat{\sigma}_{\text{i}}}$	38.3	5.5
Estimated space mean speed (local)	$\hat{v}^{\text{sms},\sigma_{\text{l}}}$	42.3	6.5
Time average space mean speed	$\bar{v}^{\text{sms}}$	42.5	6.0

Performance evaluation results

Now the selected mean speed variants are used in the model predictive speed limit controller, to investigate which variant will give the largest improvement in traffic conditions.

For the comparison of improvement in traffic conditions, the TTS is used. A lower value of J^{TTS} as given by (5.19) represents better traffic conditions, since on average vehicles are spending less time in a certain area, indicating that the flow is higher. In the uncontrolled situation, as shown in Figure 5.5, the TTS is 1068.0 veh·h.

Using the MPC-based traffic controller, the shock waves are dissolved for all averaging methods that are used, see for example Figure 5.8 where the time mean speed is used. The speed limits lower the flow that enters the shock wave by delaying the upstream traffic. In this way the inflow of the shock wave is lower than the outflow, which reduces the shock wave. Using time mean speeds $v^{\text{tms}}(k_{\text{s}})$ as state variables for the controller gives the largest improvement compared to the uncontrolled situation (J_{TTS}=901.1 veh·h, i.e., 15.6% improvement), followed by the estimated space mean speed using the instantaneous speed variance $\hat{v}^{\text{sms},\hat{\sigma}_{\text{i}}}(k_{\text{s}})$ (J_{TTS}=916.1 veh·h, i.e., 14.2% improvement), the geometric mean speed $v^{\text{geo}}(k_{\text{s}})$ (J_{TTS}=928.0 veh·h, i.e., 13.1% improvement).

The different values for J_{TTS} show that the selected averaging method can make a difference of 2.5% of the controller performance, which illustrates that it is important to consider the used averaging technique when designing a controller.

5.7 Conclusions

The use of advanced traffic control systems can significantly improve the performance of the traffic network. However, implementing such controllers is not straightforward. Therefore, we have investigated issues that are important when advanced traffic control systems are applied in practice. First, we have presented a theoretical overview of the process of developing and implementing such a controller, and next we have investigated the effect of different averaging methods for speed limit controllers.

Within the literature overview, we have first discussed issues related to the design of a model-based controller. In particular we have considered the selection of the network, measurements, the selection of the objective, the formulation of constraints, the selection of the model, and selection of the control method. Next, we have described issues related to the implementation of model-based controllers. More specifically, we have addressed calibration and validation, state estimation, demand estimation, controller tuning, and performance

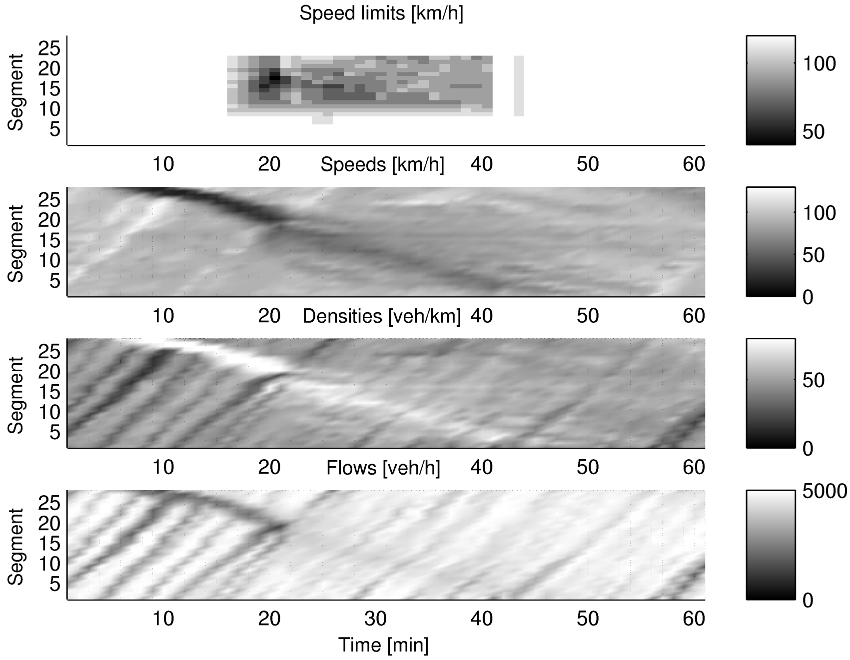


Figure 5.8: Controlled traffic flow using time mean speeds as state variable. Note that for the simulations with the other averaging methods the plots look similar. The speed limits are only present at segments 7 to 20, and active during a period of 25 minutes, as shown in the top plot.

evaluation.

Since many of the investigated topics are based on measurements, we have investigated the effect of measurements on the controller performance. In particular, we have focused on the effect of using different methods for calculating mean speeds. For a specific case study, namely simulating a part of the A12 freeway in The Netherlands, we have illustrated the influence of the averaging method for speed measurements in the performance of a dynamic speed limit controller, that applies model predictive control (MPC) with METANET as its prediction model. During the calibration of the prediction model, the most accurate prediction of the TTS has been obtained by using the geometric mean speed. During the simulation of the traffic with the controller, the time mean speed has resulted in the best controller performance. With the controller, the TTS could be reduced significantly. Improvements up to 15.6% compared to the uncontrolled situation are reached, which means that in this case the geometric mean speed is the most suitable averaging method for the speed measurements. Reducing the shock waves also has a positive effect on the flow, which is increased by 4.8%.

5.A List of symbols

Metanet model

k	simulation period counter, for period $[kT,(k+1)T)$
T	simulation time step (h)
M_s	ratio between the sample time step T_s and the simulation time step T
$\mathscr{M}$	set of pairs (m,i) freeway links m and corresponding segments i
λ_m	number of lanes of freeway link m
$\rho_{m,i}(k)$	density at segment i of freeway link m at simulation time k (veh/km/lane)
$q_{m,i}(k)$	outflow of segment i of freeway link m at simulation time k (veh/h)
$v_{m,i}(k)$	mean speed on freeway segment i of link m at simulation time k (km/h)
$L_{m,i}$	length of segment i of freeway link m
$w_o(k)$	queue length at origin o at simulation time k (veh)
$q_o^\mathrm{dem}(k)$	traffic demand on origin o during simulation period k (veh/h)

Measurements

k_s	sampling period counter, for period $[k_\mathrm{s}T_\mathrm{s}, (k_\mathrm{s}+1)T_\mathrm{s})$
T_s	sampling time interval
M_c	ratio between the control time step T_c and the sample time step T_s
K_s	number of sample steps for which measurement data is available
$q_d^\mathrm{sample}(k_\mathrm{s})$	flow on freeway near detector d during the sampling period k_s (veh/h)
$v_d^\mathrm{sample}(k_\mathrm{s})$	mean speed on freeway near detector d during sampling period k_s (km/h)
$\rho_d^\mathrm{sample}(k_\mathrm{s})$	density on the freeway stretch near detector d during sampling period k_s (veh/km/lane)
$u_{d,n}(k_\mathrm{s})$	speed of vehicle n measured by detector d during sampling period k_s (veh/h)
$N_d(k_\mathrm{s})$	number of observed vehicles on detector d during sampling period k_s
$v^\mathrm{tms}(k_\mathrm{s})$	time mean speed during the sampling period k_s (veh/h)
$\hat{v}^\mathrm{sms}(k_\mathrm{s})$	estimated space mean speed during the sampling period k_s (veh/h)
$v^\mathrm{geo}(k_\mathrm{s})$	geometric mean speed during the sampling period k_s (veh/h)
$\sigma_\mathrm{i}^2(k_\mathrm{s})$	variance of the instantaneously measured vehicle speeds during the sampling period k_s (veh/h)
$\hat{v}^{\mathrm{sms},\hat{\sigma}_\mathrm{i}}(k_\mathrm{s})$	estimated space mean speed using the instantaneous speed variance during the sampling period k_s (veh/h)
$\hat{v}^{\mathrm{sms},\sigma_\mathrm{l}}(k_\mathrm{s})$	estimated space mean speed using local speed variance during the sampling period k_s (veh/h)
$\sigma_\mathrm{l}^2(k_\mathrm{s})$	variance of the locally measured vehicle speeds during the sampling period k_s (veh/h)
$\bar{v}^\mathrm{sms}(k_\mathrm{s})$	time average space mean speed during the sampling period k_s (veh/h)

Calibration and control

N_p	prediction horizon (control time steps)
N_c	control horizon (control time steps)
θ	set of model parameters that should be calibrated
k_c	controller time step counter, for period $[k_\mathrm{c}T_\mathrm{c}, (k_\mathrm{c}+1)T_\mathrm{c})$
T_c	controller time step (h)
M_c	ratio between the control time step T_c and the sample time step T_s
$J^{\mathrm{TTS}}(k_\mathrm{c})$	total time spent in the network and at origin queues at control time step k_c (veh·h)
$J^{\mathrm{cal}}(\theta)$	calibration cost for parameter set θ
$E^{\mathrm{TTS}}(\theta)$	difference between measured and predicted TTS for parameter set θ (veh·h)

Chapter 6

Conclusions and future research

During peak periods the traffic demand exceeds the available road capacity on several parts of the road network. This leads to congestion with negative effects such as delay, increased fuel consumption, and pollution. This problem can be reduced by constructing new roads, but this is expensive and time-consuming. Another way to reduce the amount of congestion is to make more efficient use of the available roads. The objective of the research conducted in this thesis is to develop control methods to increase the efficiency of road use, and thus to reduce the negative effects of congestion. The focus lies on roadside control systems for networks that contain freeways as well as urban roads. Methods are developed to integrate available control measures, and to include the effect of the drivers' route choices in the controllers.

In this chapter we give an overview of the contributions of the research described in this thesis, and we present recommendations for future research projects.

6.1 Research contributions

The main goal of the research described in this thesis is to develop control systems that allow for a more efficient use of the available road capacity. In general, the developed controllers should provide road authorities with means to improve the performance of their networks, to coordinate control actions of different measures, and to incorporate the effects of information providing. Within this general goal, several sub-goals have been selected:

- Design of a control method for mixed freeway-urban networks,

- Development of controllers that influence route choice,

- Presenting an overview of implementation related issues.

Each of these topics is considered in one or two chapters. Below we will summarize the research described in the whole thesis. We first describe the developed controllers, and then consider implementation issues.

6.1.1 Controllers developed

In this thesis, controllers for three different processes that appear in traffic networks are developed. All controllers are based on model predictive control (MPC), a control method that uses a prediction model to predict the future evolution of the traffic flows, and an optimization algorithm to determine the optimal control signals. The objective of the controllers is to reduce the total time that the vehicles spend in the network, for some controllers combined with other factors, such as minimizing the deviation from a desired flow, or the reduction of the travel times. Note however that the developed control methods can also be used with different cost functions, e.g., related to pollution, waiting times, number of stops, etc. As optimization method for the nonlinear non-convex optimization problems that appear within the MPC-based controllers, multi-start Sequential Quadratic Programming (SQP) is used as optimization algorithm. For complex objective functions multi-objective optimization techniques are necessary, as described in, e.g., [108].

We will now shortly discuss each of the controllers, considering the process that they focus on, and describe the corresponding prediction model that is used.

Controller for mixed networks A controller is developed that controls the traffic flows on networks that consist of freeways and urban roads. The goal of the controller is to reduce the total time spent in the network. Since the controller considers the whole network, congestion that appears on on-ramps and off-ramps can be reduced by coordinating the control actions of the control measures in the urban area and on the freeway. Also, the performance of the urban and freeway networks individually is in principle improved, due to the integration of all available control measures. The control measures that are considered are ramp metering installations at on-ramps, variable speed limits at freeways, and traffic signals at urban intersections. A prediction model is developed for the controller by adapting and extending the urban queue length model developed by Kashani [82], and coupling this model with the macroscopic traffic flow model Metanet. This has resulted in a model that describes the traffic flows on mixed networks, and provides a good trade-off between accuracy and computational effort, and thus is suitable for the use in on-line controllers.

The integrated control of mixed networks has been illustrated with a network consisting of two freeways and an urban road. The performance of the controller has been compared with the performance of systems that approximate the existing control methods SCOOT and UTOPIA. For different traffic scenarios the integrated controller obtained improvements between 2% and 7%.

Anticipative controllers Three model-based controllers are developed that take the influence of route choice on freeway networks into account. As prediction model for the evolution of the traffic flows all controllers use the Metanet model.

The first controller uses an equilibrium-based dynamic traffic assignment (DTA) algorithm to determine the within-day route choice. To obtain this equilibrium-based dynamic assignment model, a static assignment model based on the Method of Successive Averages is combined with a method that adapts the current assignment toward the computed static assignment via a learning factor. This anticipative controller based on an equilibrium-based DTA model has been applied to a network with two routes and an on-ramp metering installation. With respect to the situations where no control was applied, and where the ramp metering strategy ALINEA was used, improvements of respectively 11% and 9% have been obtained.

The second controller used a route-choice-based model that describes the day-to-day route choice as well as the within-day route choice. The day-to-day route choice is based on a Bayesian learning algorithm, using the travel times experienced by the drivers as main factor in the route choice process. The within-day route choice is based on a density-dependent lookup table. This anticipative controller that uses a route choice model has been used for on-ramp as well as off-ramp metering in a network with two routes. Compared with the situation without control, the total time spent has been improved by 3% for on-ramp metering, and the mean urban density has been decreased with 60% in the situation where off-ramp metering has been applied.

Third, the reaction on provided travel time information is included in the route-choice-based method, based on the difference between the provided and expected travel times, and on the correctness of the provided travel time. This anticipative route-choice-based controller that uses information providing and variable speed limits for a network with two routes improved the traffic situation with 3% compared to the situation without control.

Day-to-day route choice controller A controller is designed that actively influences the drivers' route choice. For the controller a basic route choice model is developed, with three different versions that differ with respect to the network properties they can model. The first version, for networks with separate routes, leads to a mixed integer linear programming (MILP) problem when it is applied in an MPC controller. This is an advantage, because for this type of problems fast and efficient solvers are available, which is a requirement for on-line control. The second version of the model includes piecewise constant demands, and when it is used in an MPC-based controller the resulting optimization problem can be approximated with an MILP problem. The last version of the model includes overlapping routes and piecewise constant demands, and results in a general non-linear optimization problem.

The day-to-day route choice control for a network with two routes resulted in improvements between 45% and 72% with respect to the no-control case, depending on the selected optimization algorithm and the number of initial values.

The day-to-day route choice control is further illustrated with a network consisting of three routes where the Braess paradox appears, meaning that adding a link to the network decreases the performance. The controller illustrates that closing this extra link indeed improves the performance with 15%.

6.1.2 Implementation issues considered

Before the developed controllers can be implemented in practice, many steps should be taken. This has led to the following topics:

Implementation issues A short literature overview of issues related to the implementation of model-based controllers in practice is given. Design issues as well as implementation issues are discussed.

Influence of speed averaging methods The controllers developed in this thesis base their control actions on measurements obtained from the road network. These measurements are used to determine the state of the network, which in its term is used to predict the evolution of the traffic flows. The measurements are averaged over a certain time period. Different methods can be used to average the measurements, which leads to different average values. To investigate the effect of these different values on the performance of the controllers, we compare different speed averaging methods using a speed limit controller that has the objective to reduce shock waves on a freeway.

The speed limit controller that has been developed to determine the influence of measurement averaging methods, resulted in a reduction of the total time spent on a freeway stretch of 16% for the time mean speed, 14% for the estimated space mean speed, and 13% for the geometric mean speed. This illustrates that the influence of the averaging method is small for the selected scenario.

6.2 Future research

During the research performed for this thesis new theoretical questions were encountered which form the basis for interesting topics for future research. In this section we will present ideas for challenging research projects related to controller design, model development, and policy issues.

6.2.1 Controller-related issues

In this thesis we have developed several controllers. During the development of these controllers problems and theoretical questions were encountered that could not be solved within this research project. This has lead to the formulation of the following research topics related to controller design:

Develop efficient optimization algorithms The applicability of model-based predictive control methods largely depends on their ability to obtain optimal control settings within the available time. The controllers designed in this thesis all use multi-start SQP as optimization method for non-linear non-convex problems. Other algorithms might be available, such as, e.g., genetic algorithms [44, 62], pattern search [16, 85], tabu search [61] and SNOPT [75]. These algorithms are so called 'global' optimization methods that in general obtain good results in a reasonable amount of time. These algorithms do not use gradients or Hessians and thus are better suitable for optimization problems related to traffic control. Also, the development of new optimization algorithms can significantly improve the applicability of the developed controllers and it can also allow coordination of a larger amount of control measures, or the control of larger networks. The tailor-made algorithms should be suitable for the specific type of traffic-related optimization problems and make use of, e.g., the form of the cost function and the type of the available control signals. Special attention should be paid to the speed and scalability of the algorithms. Since there is only a limited amount of time available for optimization the intermediate solutions of the algorithm should be feasible. This allows the controller to terminate the optimization process and return a feasible control signal at the moment that the available time has elapsed.

Another possibility to lower the computation time for each optimization step is to develop distributed optimization methods, which divide the optimization problem in several smaller problems, which can be solved simultaneously. In this case algorithms should be developed that can handle the boundary conditions appearing due to the connections between the different problems.

Multi-objective control The objective of most controllers developed in this thesis is to reduce the total time spent. Many other objectives can be considered, e.g., total throughput, total delay, fuel consumption, emissions, queue lengths, etc. Some of these objectives require extensions of the prediction models, or even the development of new models. When a controller is developed, the total objective almost always has to be a combination of different goals. The combination of different goals leads to a multi-objective optimization problem, for which a suitable solution method should be selected, such as, e.g., combining the objectives into one function using a weighted sum approach, using goal attainment, or coupling the objectives by introducing a common variable, see [108]. Another issue related to the objective function is that some goals can better be left out of the function itself, but can be included into the control problem as constraints. As a result, for each goal of the controller the most suitable way of handling it should be determined, choosing between making it a constraint or including it as part of the objective function.

Translate MPC into a faster control method MPC uses a prediction model and an optimization algorithm to determine the optimal settings for the control measures. This can lead to significant improvements with respect to the traffic network performance. However, the required computational effort will remain an issue that strongly reduces the possibilities of MPC controllers in practice. A solution for this problem can be the development of faster control methods that approximate the results that can be obtained with MPC-based controllers. For example, a possible method is to simulate and optimize many different traffic scenarios with an MPC-based controller off-line, approximate the nonlinear results with a large look-up table, and then develop a controller that compares the current scenario with the scenarios in the look-up table and applies the control actions that the MPC controller has determined during the off-line simulation. When the current situation is not available in the look-up table the controller interpolates between situations in the look-up table that are relatively close to the current situation, see [73]. Another option is to develop rule-based controllers that approximate the MPC-based controllers by formulating several rules that express the relation between the current traffic situation and the optimal settings for the traffic control measures.

Develop hierarchical control methods Controlling large traffic networks with one central controller will lead to large computation times. The maximum allowed computation time limits the size of the network that can be controlled. This problem can be overcome by applying hierarchical control, where small parts of the network are optimized by low-level controllers that use detailed models, and where higher-level controllers optimize the whole network using a model with a low level of detail, see

[24, 127, 129]. In this way the processes that have an influence on the total network, such as, e.g., the route choice, and changes in the demand, can be considered by the higher-level controller, while the fast traffic flow processes, e.g., vehicle behavior on the freeway, and queue length evolution at intersections, can be handled by the low-level controllers. Challenging research questions within the area of hierarchical control are, e.g., the selection of the number of layers, the areas that should be considered by each controller, the required communication, the coordination between the controllers, and the influence that the higher-level controllers should have on the lower-level controllers.

Include future control measures In the future, new control measures or new applications of existing measures will become available. These new measures should be included in the control methods. First of all, the use of traffic information as control measure can be investigated. Initially, traffic information was provided to inform drivers about the current traffic situation. Nowadays, the information starts to be used to influence the drivers, by considering providing information as control measure, see Chapter 3 and [76, 81, 107].

Second, control methods including in-car systems can be developed. The technology that is located in cars is developing rapidly. The in-car systems can be used for data gathering, which allows for a more detailed estimation of the current traffic situation. Algorithms should be developed that can handle all the information that can be obtained by in-car devices, and that combine all this information into an estimated state of the network. Further, when the possibilities of the in-car systems are combined with road-side systems in one control approach, the performance of the road network can be increased significantly. Therefore, controllers that can handle these in-car devices as well as the road-side equipment are being developed, as in [9]. This research can be extended to the development of intelligent vehicle highway systems, in which the road-side equipment steers the in-car controllers in such a way that the vehicles drive completely automated without intervention from the drivers. Technical problems in this area are the large number of devices that should be controlled, the unknown effects of the control actions, the dependence of the controller on the penetration rate of the in-car systems, and the required computational effort. Another problem is the introduction of the system. The transition from the current situation to a fully automated situation should go smoothly. This means that the automation should be introduced in phases, and that the developed systems should be compatible with the current situation. Further, the social acceptance should be considered. Drivers might not willingly surrender the authority over their car to an automated system.

6.2.2 Model-related issues

The controllers developed in this thesis are based on models. The selected models determine the possibilities of the controller, have a large influence on the controller performance, and influence the required computation time. This has lead to the formulation of the following model-related research topics.

Develop model selection methods It is not trivial to decide which model should be used for a specific application. The goal of the model should be selected first, and next the effects that should be modeled must be determined. When a model is used within a model-based traffic controller, the model must at least be able to describe the effects of the traffic control measures. However, the level of detail in which these effects should be modeled depends on many factors, including the required accuracy and the available computation time. As a result, model selection nearly always includes a trade-off between accuracy and required computational effort. For example, when a controller uses a fast model the control time step can be smaller, which might make up for the less accurate results due to the faster feedback loop which increases the robustness with respect to measurement errors. On the other hand, a more detailed model with a longer time step might yield more accurate predictions and thus can result in better control actions. In many cases this dilemma will lead to the development of new models, or to the adaptation of existing models. Guidelines should be determined that can help to select useful features for the model, and to make the trade-off between the accuracy and the computation time.

Develop and use multi-class traffic models Traffic can be divided into different user classes, like trucks and personal cars, business and leisure drivers, male and female, etc. Each of these classes will react differently on the available traffic control measures, or can be controlled by special measures like dynamic truck lanes. When these different reactions and special measures are taken into account via including a multi-class model into the controller, the controller performance can be further improved. Issues that should be considered when designing multi-class controllers are the selection of efficient and accurate models, the possibly larger computation times of these multi-class models, selecting which properties of the classes are important for the control method, and the effect of using 'binary' control measures such as opening and closing an HOV lane.

Develop and use multi-modal models Each transportation mode (car, train, lightrail, ship) uses its own network but these networks do often intersect, at, e.g., railway passages, bridges, or urban intersections. Opening a bridge for one ship during the peak hour might cause disproportionally large delays for the road traffic, and delaying a train at a railway passage might lead to a missed connection for many travelers. This illustrates that the performance for all networks can be increased by coordinating the control actions on the different networks. Therefore multi-modal models that are suitable for the use within controllers should be developed. Issues that could be investigated are the available control methods for the different modes, and the relation between the costs for different modes. The different time scales at which the different modes operate can also form a problem for the design. Moreover, the different actions of the available control measures might require the use of integer nonlinear optimization methods.

Investigate stochastic models Traffic flows are inherently stochastic. This means that the controllers should find an optimal solution for the whole set of traffic situations that might occur in the near future. Including stochastic models in the controllers can lead to the development of robust traffic control methods that can handle the variations in the traffic situation. When stochastic models are investigated, one should take into account the number of runs that should be performed to get statistically significant results, and the required computational effort. It should also be considered that robust controllers are in general conservative, and thus find less optimal solutions.

6.2.3 Policy-related issues

The use of traffic controllers is a policy decision. Politicians determine whether the controllers are used, and what the control objectives are. Further, the potential effect of controllers can be used when long-term policies are developed. This leads to the following research topics.

Develop desired flow patterns With the controllers developed in this thesis it is possible to approximate desired flows, which allows the introduction of optimal flow patterns in the network. This means however that a method should be developed to determine what the optimal flows are for different roads. The network, the main origins and destinations, and the surroundings of the network should be analyzed to define important routes and corresponding demands. Then the locations at which low flows are desired should be determined together with the routes that can handle larger flows. Based on this information it should be possible to define the desired traffic assignment. This desired assignment can be used in the formulation of the objectives of a controller, for example to steer toward the desired flows. Another situation in which the desired flows are of interest appears when maintenance works are performed. On the one hand, the level of service should as high as possible during these works, while on the other hand the decrease in capacity of the available roads and safety considerations require lower flows. In this case, the order in which the works are performed and the control actions that are taken should be optimized with respect to the difference between the generated flows and the desired flows.

Investigate social and legal aspects of future control measures Future control methods can only be used when they are accepted by the drivers, and when the control actions performed are legal. This is not automatically the case, as can be seen when considering, e.g., information providing or in-car control systems. When using information providing on variable message signs, the attitude of the drivers toward the presented information is important since reacting on the provided information is voluntary, and thus the drivers should consider the provided information to be correct, otherwise they will ignore the messages. This also leads to an investigation of the legal issues with respect to information providing. In principle, the presented travel times should correspond to the real travel times, otherwise it is seen as false informing. However, when providing information and other control measures are combined into one control method, the other control measures can be used to make the provided information true. Is this considered as false informing or not?

When considering in-car systems, the acceptance of the drivers plays a large role. The drivers should partly or completely surrender their car to the control system, which limits the possibilities of the driver and might cause distrust with respect to the control system. Also, the legal aspects of automated highway systems should be considered. Who will be responsible for, e.g., incidents?

Influence road layout and spatial scheduling The layout of the road network has a significant influence on the results that can be obtained with traffic control measures. The number of lanes at urban intersections influences the possibilities of traffic signal control, the number of on-ramps and off-ramps of a freeway influences the performance of variable speed limits, and the number of alternative routes determines the effect of route choice control. Guidelines should be developed for improving the road layout in such a way that the performances of traffic controllers can be increased, and a method for simultaneously developing the road layout and control measures can be designed. This will lead to an optimal road layout, which also offers a trade-off between control possibilities and, e.g., road area, number of ramps, number of lanes, and number of roads. On the long term, not only the road layout can be influenced, but also the spatial scheduling can be adapted based on the expected traffic demands and control possibilities. The traffic demand on the network largely depends on the origins and destinations in the network. The traffic demand thus can be influenced by changing important origins and destinations. Existing origins and destinations cannot be changed within a short period, but via granting construction licenses the location of buildings and houses can be influenced, e.g., by allowing companies at special locations only, and residential buildings at others only. When a new residential or business area is developed, the location of the main origins and destinations can be selected beforehand. This can lead to the integrated design of the road network and the location of the main origins and destinations, which can prevent a mismatch between road capacity and demand, and reduce the negative effects of the traffic on the surroundings.

Appendix A

Route guidance during maintenance works

In this appendix we perform a simulation case study, in which we consider the road network around the Dutch city of Eindhoven. During the years 2008 to 2013 maintenance works will be performed on this network, combined with a reconfiguration of the freeway part of the network. It can be expected that this will lead to large traffic problems. In this appendix we propose a route guidance method to reduce these traffic problems by re-routing the traffic flows. With this method, we will illustrate the possibilities of influencing the route choice of drivers on purpose during construction works.

We first give a general overview of maintenance/reconstruction works and the traffic management during such works. Next, we describe the network of Eindhoven and focus on the expected traffic problems during the works. Then we make a detailed model of the network, which is used as a basis for the case study. Then we design a basic system for route guidance, which will reduce the travel time influencing the route choice of the drivers. Finally, we present some simulation results. The work described in this appendix is based on [69] and [70].

A.1 Maintenance works

Maintenance or reconstruction works are necessary to keep the road network safe and up-to-date, but they always come with negative side-effects for the traffic operation during the period that the works are performed. Lanes or complete roads have to be narrowed or closed, the maximum speed has to be reduced, and the neighboring roads have to deal with an increase in demand. All this leads to longer travel times and more delay for the road users. In The Netherlands the deterioration of the traffic operation in a construction zone has direct consequences for the traffic operation in a large part of the surrounding network due to the large number of roads in a relatively small area. Currently, the difference between available capacity and traffic demand is small, and much more inconvenience due to maintenance works cannot be tolerated. However, many parts of the Dutch road network need maintenance or reconfiguration to be able to cope with the current and future traffic demands. Therefore in The Netherlands the tendency is that more and more effort is put into the investigation and analysis of the traffic operation during maintenance or reconstruction projects, see, e.g., [52, 151, 171].

Also in the US the traffic operation during road maintenance works is a topic of current research. The US Department of Transportation, and especially the Federal Highway Department, has developed a general optimization process to minimize the impact of reconstruction projects on the traffic operation, such as extra congestion and an increase in the number of incidents [117]. Further, a strategy for finding an optimal maintenance plan is described in [34], and the reaction of road users to dynamic route information at freeway work zones is investigated in [74]. In [74] the estimated travel time to the end of work zone was presented to the drivers. Showing these travel times actually resulted in the selection of an alternative route by 7 to 10% of the road users.

A.2 The network around Eindhoven

For the case study, we consider the network around Eindhoven. The city of Eindhoven is located in the south of The Netherlands. The network around the city is selected since the city attracts a large amount of traffic due to the presence of large company areas and a large shopping center, and since routes for long-distance traffic toward Belgium and Germany pass the city. This combination of local and long-distance traffic forms a challenge for the road authorities, since the two types of traffic use the same road network, but have different needs. Furthermore, the capacity of the road network is not enough to handle the current demand, and thus a reconstruction is required to increase the capacity. During this reconstruction the road capacity will be reduced even more. This will increase the importance of traffic control measures, which should be used to keep the efficiency of the road use at an acceptable level.

The selected network has several road administrators. The main road network containing the freeways is administered by the Dutch Ministry of Transport, Public Works and Water Management, and the underlying network is administered by the provincial government, the city of Eindhoven, and the governments of the surrounding villages, depending on the location of the different roads. These road administrators are responsible for the maintenance, traffic operation, and safety on the roads. There exists an organization that helps the different parties with issues concerning the road administration. This organization is the SRE ('Samenwerkingsverband regio Eindhoven', the cooperation association of the region Eindhoven). The SRE initiates, stimulates, and coordinates the cooperation between the 21 communities that form the region of Eindhoven.

The current road network is shown in Figure A.1. For this case study we consider origins/destinations T1, T2, and T3, junctions P and S, and on-ramps/off-ramps Q and R, as marked in the figure. The main freeways in the network are the A58 located at the north side of the city, the A2 connecting T2 and T3, and the A67 connecting T1 and T3, with junctions at Batadorp (T2), De Hogt (P), and Leenderheide (S). Important on-ramps and off-ramps are Veldhoven Zuid (U), Waalre (R), and the High Tech Campus (Q). Major local roads are the ring road around the center of Eindhoven, the connections between this ring road and the freeways, and the roads to the surrounding villages. We mainly consider the southern part of the network during the case study. A simplified representation of this part

Figure A.1: The network of Eindhoven.

of the network is presented in Figure A.2, which shows the origins/destinations T1, T2, and T3, junctions P and S, and on/off-ramps Q, R, and U, and the connecting freeways.

There are three important routes in the network:

- T1 - T2, which is actually the connection from the A67 to the A2, for traffic traveling from the south to the north of the Netherlands and vice versa.

- T1 - T3, the A67 from west to east and vice versa.

- T2 - T3, the A2 from north to south and vice versa.

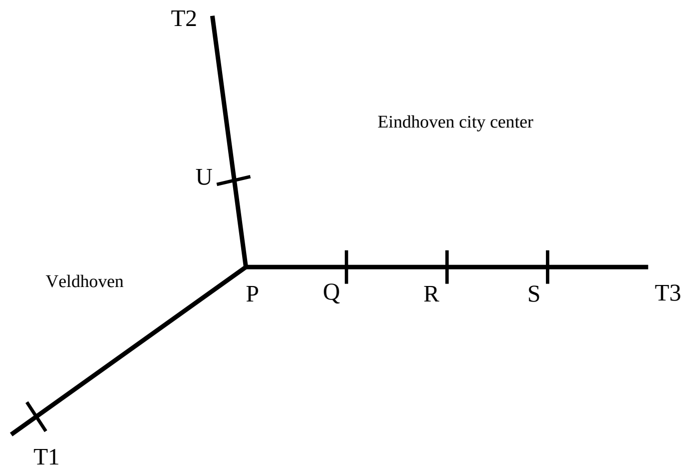


Figure A.2: Schematic representation of the part of the network that is considered during the case study.

A.2.1 Paramics model of the network of Eindhoven

The network of Eindhoven, as shown in Figure A.1, will be modeled with a detailed microscopic model. The obtained model will be used as a basis for the case study. To create this model the microscopic modeling and simulation tool Paramics [137] is used. Paramics is a suite of software tools used to simulate the movement and behavior of individual vehicles on urban and freeway road networks. It contains three models, a car-following model, a lane-changing model, and a route choice model. Based on these models the individual vehicle movements and the interactions between the vehicles in the network are determined. This results in a detailed simulation of the traffic flows on the whole network.

Before actually starting to construct the model, the number of roads that should be included in the model of the network of Eindhoven must be determined. We have decided to model all freeways. The number of modeled urban roads is based on a trade-off between accuracy and complexity, since including more roads means increasing the complexity, which increases the required computational effort. We have selected the major urban roads that handle a large part of the local traffic, and the roads that are used as secondary route by the long-distance traffic in case of incidents on the freeway. These are the roads N265, N69, N270, and connections between Eindhoven and Veldhoven, Eersel, Waalre, and Best. Further, we model the network south-west of the freeways more thoroughly, since this part is used for the simulation study. The final network as modeled in Paramics is shown in Figure A.3.

Information about links, nodes, and geometric properties of the network of Eindhoven is obtained form a static Omnitrans model [118, 146], owned by the SRE. The demand is also derived from the demands of this Omnitrans model. In the Omnitrans model static demand information is aggregated over four time periods: morning peak, evening peak, rest of the day, and twenty-four hours. For each origin in the network, the mean demand during the selected period is given. For the case study we consider the morning peak between 6:30

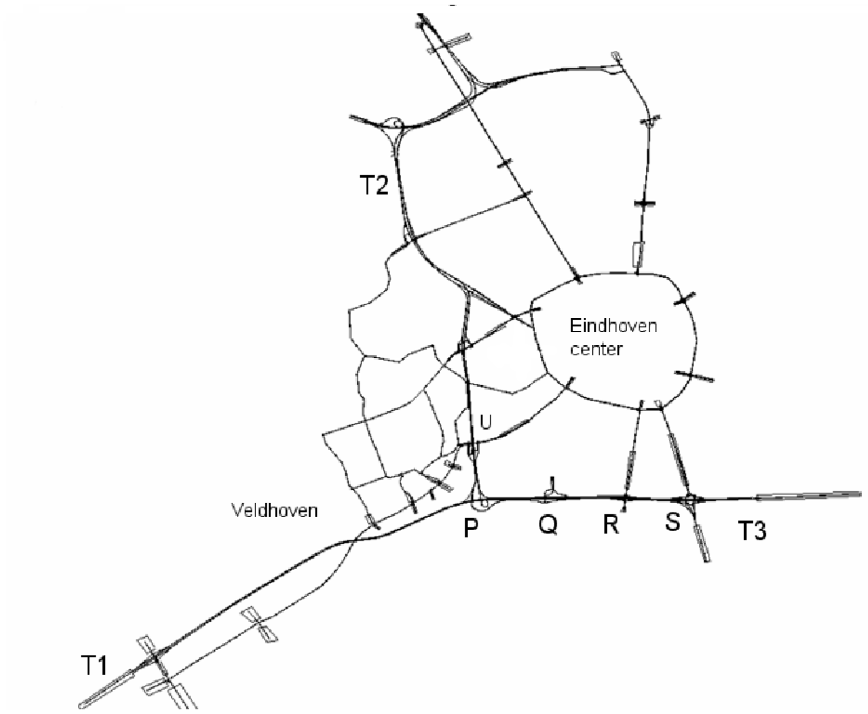


Figure A.3: Network that is modeled with Paramics.

a.m. and 9:00 a.m., since during this period the demand is the highest, and thus many traffic problems can be expected. Within Paramics, the demand on the network is represented with a static Origin-Destination (OD) matrix, which contains mean values of the demand for each origin-destination pair. The demands of the Omnitrans model are used to determine the entries in this OD-matrix. To obtain a more dynamic demand, Paramics multiplies these static demands with a demand profile, which is equal for all origins. The profile that we have selected is based on flow measurements on the A67 freeway, and shown in Figure A.4.

The Paramics software offers three different methods to assign the traffic flows to the network: all-or-nothing assignment, stochastic assignment, and dynamic feedback assignment. For the case study we select the dynamic feedback assignment, which assumes that drivers will adapt their route choice based on the current traffic situation on the network. The feedback time is set to 5 minutes, which is short enough to obtain accurate reactions on changes in the traffic situation, and long enough to prevent fast switching. Factors that influence the route choice of the drivers are the travel time for a trip and the distance traveled, where the travel time is selected to be twice as important as the traveled distance, since according to [18] travel time is one of the most important aspects influencing the route choice.

Two types of measurement data from the network of Eindhoven are available to validate the results obtained with the model. The Dutch Ministry of Transport, Public Works, and Water Management gathers measurements with loop detectors on the freeway, with a detec-

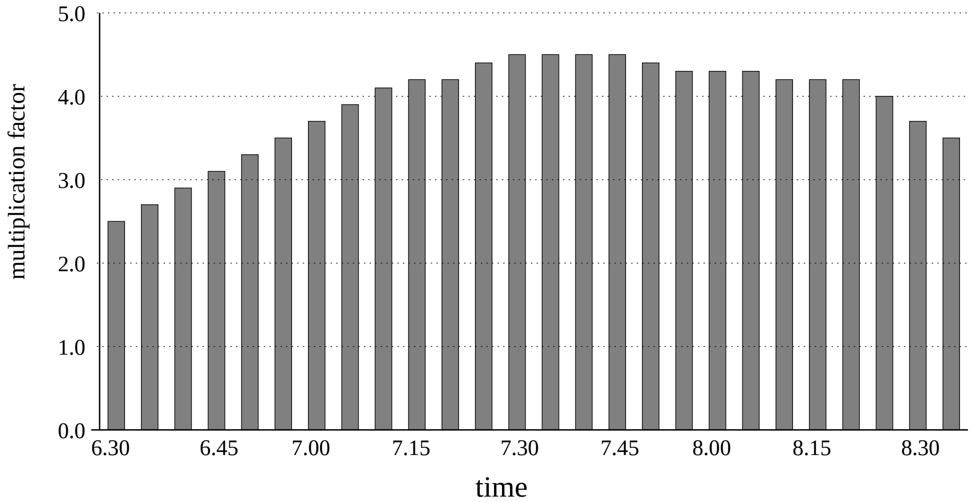


Figure A.4: Demand profile during the morning peak. For each origin a static demand is defined, which is multiplied with the factor that is specified in the demand profile to obtain a dynamic demand.

tor at every stretch of 500 m. These detectors measure, among others, the average flows for every minute. In the urban area manual counts are performed by the SRE on the ring road around the center, and on the provincial road connecting T1 and U via Veldhoven. These counts can be translated into average flows for these roads. The obtained average flows are compared with the simulated flows of the Paramics model. The difference between the flows was between 2.47% and 13.89%. This is not an excellent result, and before the model can be used in practice, a more extensive calibration should be performed. However, the results obtained with this study can be used to illustrate the possible effects of route guidance, but the actual improvement will differ from the improvement that would have been obtained with a more accurate model or in reality.

A.3 The network during the maintenance works

During the reconstruction works, the configuration of the network of Eindhoven will be changed such that the current freeway configuration, with two lanes and a peak lane in each direction of the freeways, will be reconfigured into a configuration with two times two lanes in each direction. In this new configuration, the capacity of the freeways will be improved by separating the long-distance and the local traffic. The currently existing two-lane freeways will serve as long-distance freeways and the newly created two-lane freeways, parallel to the existing freeways, will serve as freeways for local traffic. These new freeways will be connected to the underlying network with many on-ramps and off-ramps, to facilitate traffic entering and leaving the city. The three major junctions (T2, P, and S) will facilitate the transition between the long-distance and the local freeways.

The case study focuses on the maintenance works around junction P. The link from T1

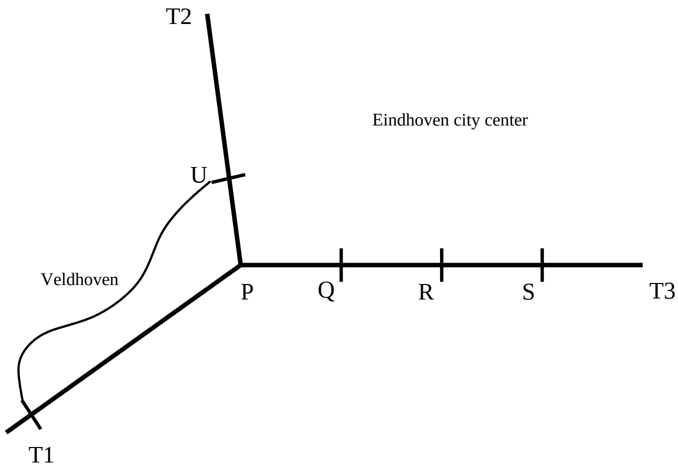


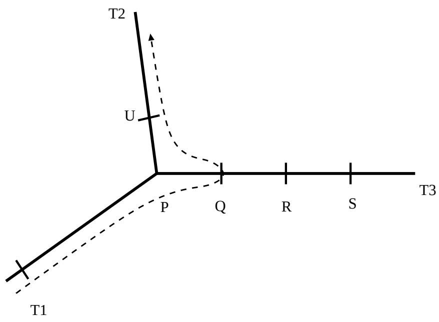
Figure A.5: Part of the network that is considered during the case study.

to T2 will not be available for a period of three years, from 2009 to 2011. We propose a dynamic route guidance strategy that can be used to reduce the delay that occurs due to this closure.

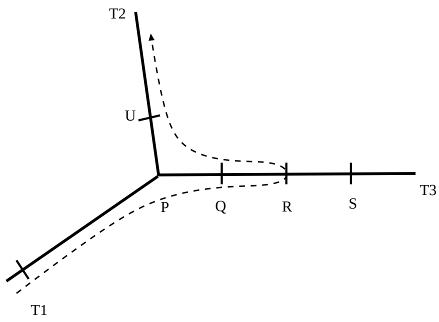
Junction P, see Figure A.5, facilitates the traffic from T1 to T2 as well as from T1 to T3. When the connection between T1 and T2 will be removed, the traffic on the route T1 → T2 will have to select another route. The most logical alternative route follows the urban road through Veldhoven. The capacity of this road however is not high enough to facilitate the resulting large demands, and the queues that will appear in the residential areas around this road will cause discomfort and extra pollution. For the case study, we assume that the use of this road will be prevented by closing it for long-distance traffic. Then three different routes remain available for the traffic from T1 to T2. For these routes, the drivers should travel from T1 in the direction of T3, and then make a U-turn at exit Q, R, or S, and then continue toward T2, as shown in Figure A.6.

Due to the layout of the network, a problem will appear when all drivers from T1 to T2 use exit Q to make their U-turn, see Figure A.7. The resulting traffic flows will lead to a large flow from T1 to Q, which will have to cross the already existing large flow of drivers that travel from T2 to T3. The available space for the weaving between the two traffic flows is very limited, and thus the weaving will strongly influence the traffic flows and create congestion. This congestion will spill back into the upstream directions, toward T1 and T2.

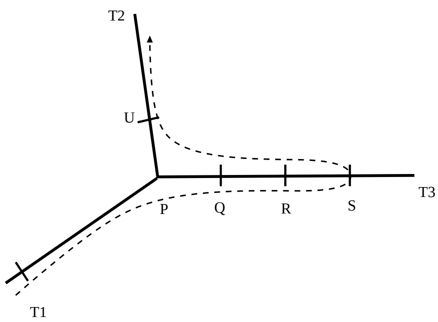
The other two exits that are available to make a U-turn do not have this negative effect. Re-routing the traffic along these routes will reduce the delay that is experienced during the maintenance works. However, the travel distance on these routes is longer, so a control strategy has to be developed to incite drivers to use these routes.



(a) Route 1, via exit Q.



(b) Route 2, via exit R.



(c) Route 3, via exit S.

Figure A.6: Alternative routes that are available when the direct connection from T1 to T2 via P is under construction.

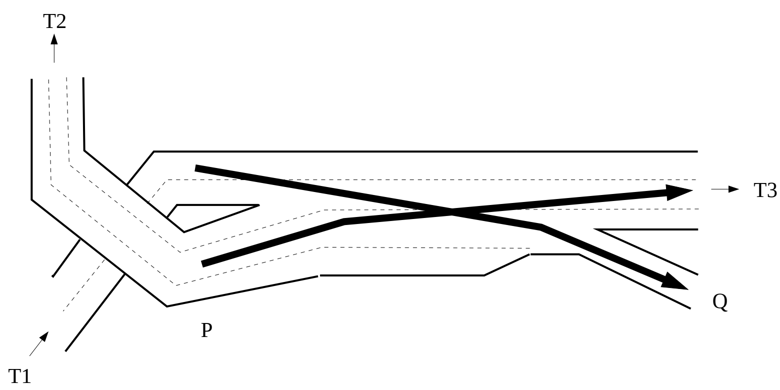


Figure A.7: Road stretch between P and Q where the traffic from T1 to Q crosses the traffic from T2 to T3.

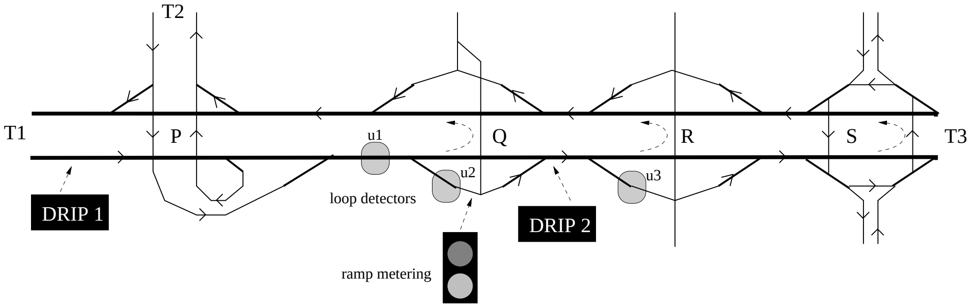


Figure A.8: Controlled freeway stretch with locations of the DRIPs, the ramp metering installation, and the loop detectors.

A.4 Proposed route guidance system

To incite drivers that travel from T1 to T2 to use exits R and S of the freeway to make their U-turn, we propose a route guidance system that is a combination of two measures, off-ramp metering and providing information on dynamic route information panels (DRIPs). The measures are located in the network as shown in Figure A.8. The first DRIP is located at T1 before junction P, the second DRIP is located between exits Q and R, and the off-ramp metering installation is located at exit Q.

A basic vehicle-actuated control method is used. With this control method, we will show the possibilities of dynamic route guidance, and justify the development of advanced route choice controllers, as done in Chapters 3 and 4. The main goal of the control strategy is to reduce the congestion in the network. A large part of this congestion is caused by the weaving behavior on the stretch P → Q. Improving the flow passing this stretch will reduce the congestion significantly, and is thus a goal of the control methods.

We will now describe the control strategies that are used for the control measures that have been selected. The methods are basic switched-control methods, which are used in practice for, e.g., on-ramp metering installations [150].

The off-ramp metering installation uses measurements of the detection loop $u1$, located between P and Q. The control strategy that is applied is shown in Figure A.9. If there is no congestion, the off-ramp metering installation is off. When the measured speed v_{u1} at the detector location drops below 30 km/h the ramp metering installation starts metering with a fixed metering rate. This static metering rate is selected such that a queue appears at the off-ramp, which will discourage drivers to make a U-turn at exit Q. When fewer vehicles take the exit, the congestion due to the weaving will decrease, and finally the measured speed will be above 60 km/h, and the off-ramp metering installation will be turned off.

For the DRIPs we consider two different control strategies. The first strategy uses only DRIP 1. This strategy is presented in Figure A.10. The traffic is measured with detector $u2$ at exit Q and with detector $u3$ at exit R. If there is no congestion, the DRIP gives the advice to take exit Q. If the speed at detector $u2$ drops below 30 km/h, the DRIP advises to take exit R. If the speed at detector $u3$ also drops below 30 km/h, the advice switches to exit S. If the speed at detector $u3$ becomes higher than 60 km/h, the advice is to take exit R, and

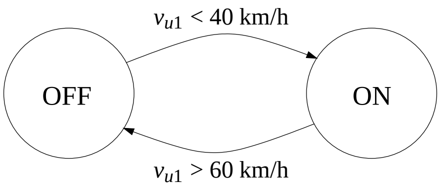


Figure A.9: State diagram for the ramp metering installation.

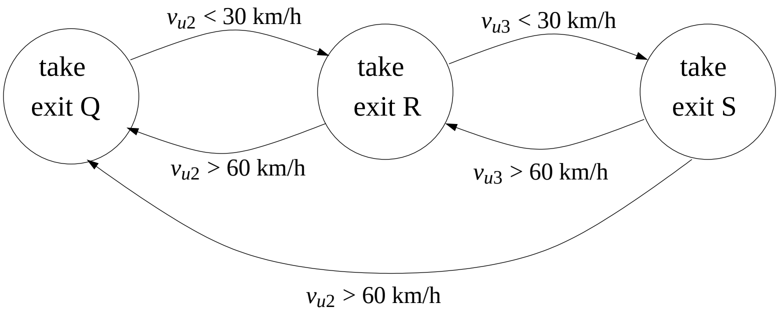


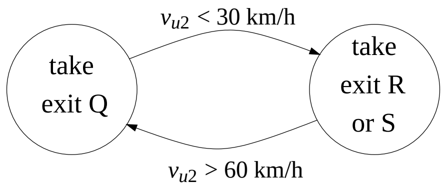
Figure A.10: State diagram for the DRIP when only 1 DRIP is active.

if the speed at detector $u2$ becomes higher than 60 km/h, the advice is to take exit U. With this algorithm a queue appears at exit Q, and when this queue reaches the detector, traffic is encouraged to take exit R. When the queue at this exit reaches the detector before exit R, traffic is guided toward exit S.

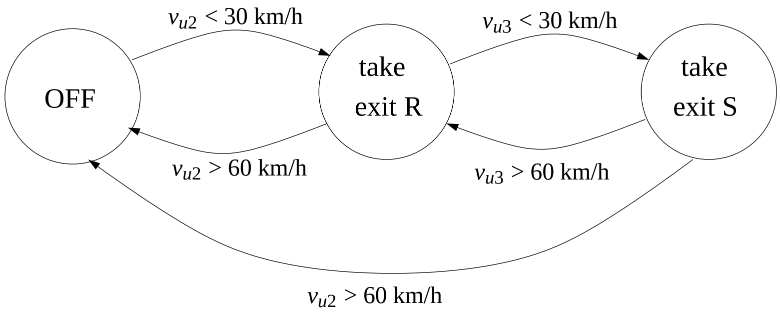
The second control strategy for the DRIPs uses both DRIPs. The first DRIP switches between two advises: 'take exit Q' and 'take exit R or S', see Figure A.11(a). Which advice is selected depends on the measurements obtained with detector $u2$. The second DRIP specifies this advice. When the first DRIP advises to take exit Q, the second DRIP is off, see Figure A.11(b). If the first DRIP advises to take exit R or S, the second DRIP advises to take exit R if the speed at detector location $u3$ is above 60 km/h, while it advises to take exit S if this speed is below 30 km/h. The difference with the first control strategy is that the delay between the moment that the choice between making a U-turn at exit R or at junction S is made and the moment that the drivers have reached the selected exit is smaller for the second strategy, which means that the controllers can more effectively react on the measurements.

The values that are selected for the thresholds have a large impact on the performance of the controllers. The values that are used for this case study are shown in the figures. The value of the lower threshold is selected lower than the speed corresponding to the critical density. This to prevent that the density will exceed the critical density, since when the critical density is reached congestion will appear. The upper threshold is selected in such a way that it is approximately equal to the speed at which free flow appears. To improve the performance of the controllers the values for the thresholds should be optimized with respect to the performance of the controllers. However, for the current case study we used these initially selected values.

It might be expected that the strategy with 2 DRIPs will always perform better than the



(a) DRIP 1



(b) DRIP 2

Figure A.11: State diagram for the DRIPs when both DRIPs are active.

strategy that uses only 1 DRIP. However, the purchase and maintenance costs of DRIPs are high, and the costs of the second DRIP might not even out against the obtained improvements in the travel time. This makes it useful to consider both strategies, and to compare them with respect to their costs and benefits.

A.5 Simulation study

With a simulation study we illustrate the effect of the route guidance system. For this study, the control measures have been implemented in the Paramics model of Eindhoven by means of agents, using the test bed for multi agent systems for dynamic traffic management, which is described in [174]. We first describe the set-up of the simulation study, and then we present the obtained simulation results.

A.5.1 Set-up of the simulation study

The designed control system has been tested for the traffic demand of the morning peak between 07:00 a.m. and 09:00 a.m. During this period the demand from T1 to T2 is the largest, and thus the congestion is the most severe. In the network there are already control measures present, e.g., speed limits for incident detection, and traffic signals at local roads near the exits. These control measures are implemented in the model, and perform their control actions simultaneously with the newly developed dynamic route guidance strategy.

The percentage of the vehicles that comply to the route advice on the DRIPs should be selected. In literature, values between 7% and 14% are presented [51, 74, 90]. Since we expect that during maintenance works drivers are more willing to comply, we select a value of 25% for our case study.

We will perform three different simulations. During the first simulation no control will be applied. During the second simulation the off-ramp metering installation is active, and the first strategy for the DRIPS, with only DRIP 1, is applied. During the third simulation, the off-ramp metering installation and both DRIPs are active, using the second strategy for the DRIPs.

The performance of the control strategies will be evaluated with respect to different criteria. First of all, the outflow of the freeway stretch between P and Q will be considered. The controller should increase this flow. Second, the delays experienced due to the maintenance works on the stretches between P and S, S and P, and T1 and P are considered, since at these stretches the most traffic problems appear. Third, the travel times on important routes in the network will be considered. The controllers should decrease the travel time from T1 to T2 and T3, without increasing the travel times in other directions too much.

A.5.2 Simulation results

We will now describe the results of the three different simulations. First, we give an overview of each simulation, and next we compare the results.

In the first simulation, where no control is applied, the expected congestion due to the weaving behavior appears at the stretch P $\rightarrow$ S. The conflicting streams T1 $\rightarrow$ T2 and T2 $\rightarrow$ T3 cause this congestion, which spills back for several kilometers into the freeway sections upstream of the weaving section, blocking the traffic flow originating from T1 as well as the traffic flow originating from T2.

The second simulation uses off-ramp metering and only one DRIP. At the start of this simulation DRIP 1, located upstream of the junction gives the advice to *take exit Q*. When the demand increases, the weaving behavior between the junction, and the first exit causes the traffic to slow down, which announces upcoming congestion. When the speed at this weaving section drops below the threshold of 30 km/h the off-ramp metering installation is activated. At the off-ramp of exit Q, the flow leaving the freeway is metered such that a traffic jam is created on the off-ramp. When this jam reaches the end of the off-ramp, but does not block the slip-lane of the freeway, the DRIP notices the congestion on exit Q and gives the advise to *take exit R*. When due to the extra traffic making use of exit R, this exit becomes also fully loaded, the DRIP advises to *take exit S*. When an exit queue becomes empty, the advice switches back to the previously displayed advice.

In the third simulation in general the same scenario as in the second simulation occurs, the only difference being that vehicles that have already passed the first DRIP can be timely informed to change their route. This means that for these vehicles the time between the moment that the advise is given and the moment of the actual route choice is reduced. This results in a decrease of the number vehicles entering the queue when the off-ramp of the exits R is already full and therefore decreases the congestion.

The flows that are present in the network during the simulations are presented in Table A.1. In the first column of the table the outflow of the weaving area between P and Q is listed. The other columns present the flows using exits Q, R, and S. The first row gives the results without control, while the second and third row present the results with control using off-ramp metering and respectively one and two DRIPs. The goal of the control strategies

	P → Q	Q	R	S
No control	9221	2177	1551	440
RM + 1 DRIP	9852	2021	1646	918
RM + 2 DRIPs	9703	2047	1448	1036

Table A.1: Vehicle counts (number of vehicles per two hours) at different locations for the no control case and the two control methods, where RM is the abbreviation for ramp metering installation.

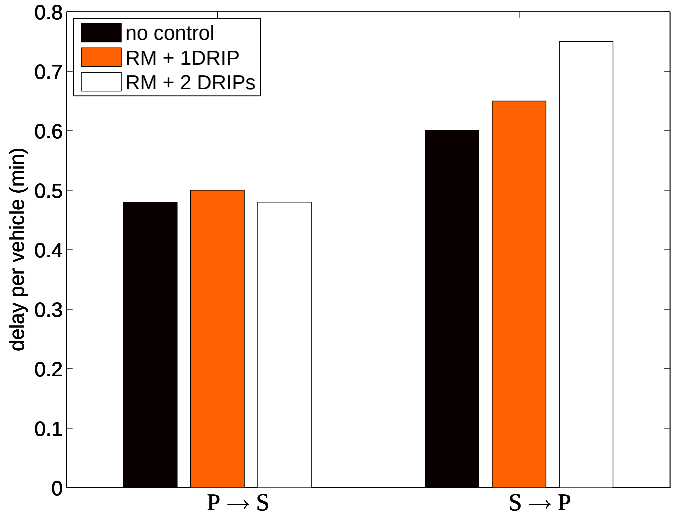


Figure A.12: Delays per vehicle on the stretches P → S and S → P.

was to increase the outflow of the stretch P → Q, which both of them do. They reach this by leading more traffic to the other exits, which indeed reduces the influence of the weaving behavior and thus increases the outflow. The strategy with two DRIPs incites more drivers to take exit S compared to the strategy with one DRIP, and thus obtains the best performance.

The flow at the stretch P → Q increases when the controllers are used. However, the delay on the stretch P → S is not reduced, as shown in Figure A.12, where the first vertical bar presents the delay for the simulation where no control is applied, the second bar shows the situation where off-ramp metering and 1 DRIP are used, and the last bar gives the delay for the simulation with off-ramp metering and two DRIPs. The delay at the stretch P → S is equal or even higher in the situations where control is applied. This is due to the fact that more drivers make a U-turn at exits R and S, and thus stay longer at the stretch P → S increasing the delay. Also, the delay into the direction S → P is higher than without control. This is due to the fact that the drivers that make a U-turn at intersections R and S also use the stretch S → Q longer, causing a larger delay in this direction. This effect is the largest for the controller with two DRIPs, since with this controller the largest number of drivers takes exit S. This illustrates that for the stretch P → S itself the controllers are not able to prevent the congestion. However, the controllers can reduce the spill-back of the traffic jam

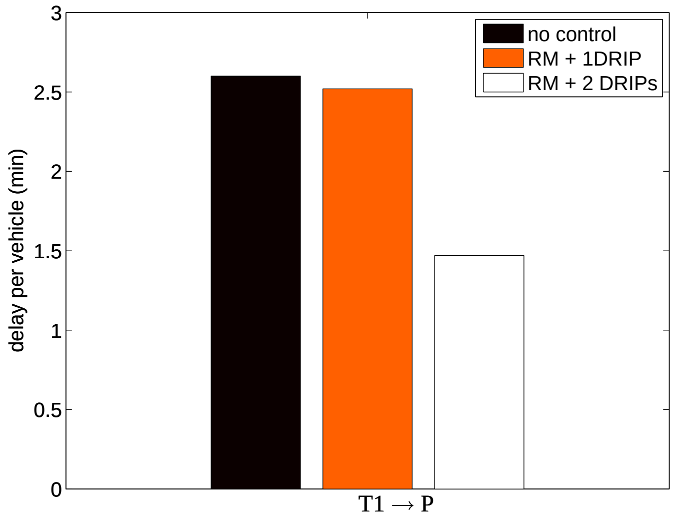


Figure A.13: Delays on the stretch T1 → P.

into the upstream direction by increasing the number of vehicles that passes the stretch P → Q. This increases the outflow of the upstream stretches, and thus reduces the delay on these stretches, as shown for the stretch T1 → P in Figure A.13.

The actions of the controllers influence the traffic flows in all directions. Figure A.14 shows the travel times for important routes in the network. The travel time on the route T1 → T2 is the highest, since the drivers on this route have to make the U-turn. The presented travel time is the weighted average over the different exits that are used for the U-turn. Other routes that have high travel times are T1 → T3 and T2 → T3. The traffic flows on these routes encounter the congestion that spills back from the stretch P → Q, which increases the travel times on these routes. The first controller with 1 DRIP is not able to significantly reduce the travel times on these routes, mostly due to the delay that is caused by the extra vehicles that cause a queue on the stretch Q → R. The second controller with 2 DRIPs is able to divert more vehicles from this queue toward exit S, which reduces the travel times on these routes. The routes T3 → T1 and T3 → T2 overlap with the routes T1 → T2 only after the U-turns. The travel time on route T3 → T2 is only influenced by the extra traffic that uses the stretch S → P, and thus increases when the controllers are used. The traffic flow on route T3 → T1 has to cross the traffic flow that originates from on-ramp Q traveling to T2. This means that the travel time in the uncontrolled situation is high, and that the use of the controllers can reduce the weaving behavior and decrease the travel time significantly. The second controller slightly increases the travel time due to the fact that more vehicles use the route.

In general, both control methods are able to improve the outflow of the freeway link between P and Q, and thus reduce the delay on the route from T1 to T2. This however leads to an increase in delay and travel times on other routes. Which routes experience the largest delay depends on the use of one or two DRIPs.

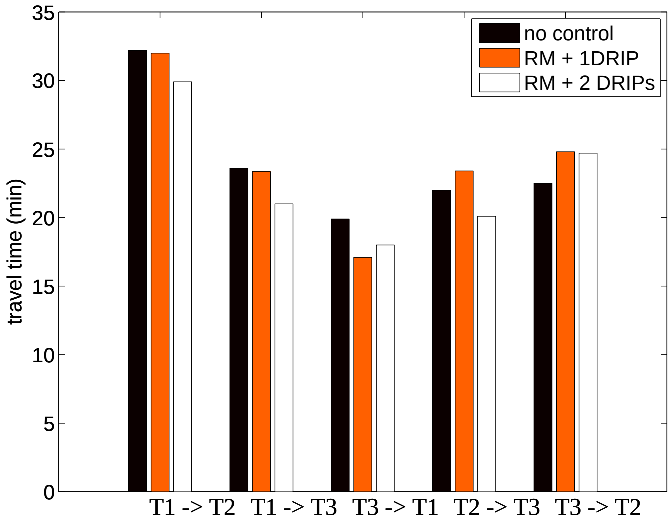


Figure A.14: Travel times on important routes in the network.

A.6 Cost-benefit trade-off

Installing DRIPs in the network is expensive. This means that before installing a DRIP, the costs should be compared to the benefits of the DRIP.

The costs of a DRIP consist of purchase and installation costs, and maintenance costs. The purchase and installation costs of a DRIP are around € 290000, and the maintenance of a DRIP costs € 5000 each year according to the Dutch Ministry of Transport, Public Works, and Water Management. Assuming that the DRIPs are used during the three years that junction P will be closed, this results in a total cost of € 305000 for the duration of the reconstruction works at junction P (3 years). The loop detectors and the ramp metering installation are already present in the current network, or should be used in the network after the reconstruction. So to determine the costs of the control methods, only the costs of the DRIPs should be considered. This means that the first control strategy, which uses only one DRIP, costs € 305000, and that the second control strategy with two DRIPs costs € 610000.

The benefits of a traffic control method are expressed in the reduction of the delay that they obtain. Each vehicle delay hour costs 20 Euro. The total delay in the uncontrolled situation is 672 h per moring peak, which is 1.466310^7 h for all morning peaks during the considered period of three years, which amounts to 14.7 million euro. The total delay for the situation where the first route guidance method with off-ramp metering and one DRIP are used is 633 h per morning peak, which is $4.67 \cdot 10^5$ h, amounting to 13.8 million euro. The daily delay for the second controller with off-ramp metering and two DRIPs is 527.4 h per day, which costs in total 11.5 million euro for the considered period.

An overview of all costs and benefits is presented in Table A.2. The first column presents the costs of the DRIPs, the second the costs of the vehicle delay hours. The third column presents the benefits that are obtained with respect to the no control case. The last column

	equipment costs	delay costs	benefits	factor
no control	-	14.7	-	-
RM + 1 DRIP	0.305	13.8	0.81	2.7
RM + 2 DRIPs	0.610	11.5	3.1	5.1

Table A.2: Costs and benefits of the dynamic route guidance methods (m€).

gives the ratio between the benefits and the costs. Normally, a measure is implemented by the Dutch Ministry of Transport, Public Works, and Water Management when its benefits are four times as much as that it costs. This means that the method with one DRIP provides not enough benefits to justify the installation, while the method with two DRIPs does. But note that in other countries other factors can be used, and that the possibilities to use the DRIPs after the reconstruction period can have an impact on the choice to install one or two DRIPs.

A.7 Conclusions

We have performed a case study on the network of Eindhoven, which will undergo a major reconstruction in the coming years. The situation during this reconstruction has been used to illustrate the potential benefits of dynamic route guidance. First, a microscopic model of the network has been made using the modeling software Paramics. Then, a basic route guidance method has been developed, which gives an indication of the effects that could be reached with route guidance. Also, a cost-benefit trade-off is performed to compare different variants of the developed route guidance method. The obtained results are encouraging, and form a basis for the development of advanced route choice controllers.

Bibliography

[1] A. Al-Ghazzawi, E. Ali, A. Nouh, and E. Zafiriou. On-line tuning strategy for model predictive controllers. *Journal of Process Control*, 11(3):265–284, June 2001.

[2] A. Alessandri, A. Di Febbraro, A. Ferrara, and E. Punta. Nonlinear optimization for freeway control using variable-speed signaling. *IEEE Transactions on Vehicular Technology*, 48(6):2042–2052, November 1999.

[3] C. Antoniou, M. Ben-Akiva, and H.N. Koutsopoulos. Online calibration of traffic prediction models. *Transportation Research Record*, 1934:235–245, 2005.

[4] K. Ashok and M.E. Ben-Akiva. Alternative approaches for real-time estimation and prediction of time-dependent Origin-Destination flows. *Transportation Science*, 34 (1):21–36, February 2000.

[5] A. Atamtürk and M.W.P. Savelsbergh. Integer-programming software systems. *Annals of Operations Research*, 140(1):67–124, November 2005.

[6] TPA Traffic & Parking Automation. Toptrac; verkeersafhankelijke regeling voor netwerken. Technical report, TPA Traffic & Parking Automation, NHTV Breda, 2002. in Dutch.

[7] J. Barceló and J.L. Ferrer. *AIMSUN2: Advanced Interactive Microscopic Simulator for Urban Networks, User's Manual*. Universitat Politècnica de Catalunya, 1997.

[8] J. Barcelo, J. Casas, J.L. Ferrer, and G. Funes. Heuristic dynamic assignment based on microsimulation with AIMSUN 2. In *AFRICON, 1999 IEEE*, volume 1, pages 21–26. IEEE, October 1999.

[9] L. Baskar. *Traffic Management and Control in Intelligent Vehicle Highway Systems*. Ph.D. thesis, TRAIL thesis series T2009/12, Delft University of Technology, Delft, The Netherlands, November 2009.

[10] T. Bellemans. *Traffic Control on Motorways*. PhD thesis, Faculty of Applied Sciences, K.U.Leuven, Leuven, Belgium, May 2003.

[11] T. Bellemans, B. De Schutter, and B. De Moor. Anticipative model predictive control for ramp metering in freeway networks. In *Proceedings of the 2003 American Control Conference*, pages 4077–4082, Denver, Colorado, June 2003.

[12] A. Bemporad and M. Morari. Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35(3):407–427, March 1999.

[13] M. Ben-Akiva, A. De Palma, and I. Kaysi. Dynamic network models and driver information systems. *Transportation Research Part A*, 25(5):251–266, September 1991.

[14] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki. *Diagnosis and Fault-tolerant Control*. Springer-Verlag, Berlin Heidelberg, second edition, 2006. ISBN 97-835-40010-562.

[15] M.C.J. Bliemer. A quasi-variational inequality approach to multi-user class dynamic traffic assignment. *Transportation Research Record*, 1710:11–19, 2000.

[16] C. Bogani, M.G. Gasparo, and A. Papini. Generalized pattern search methods for a class of nonsmooth optimization problems with structure. *Journal of Computational and Applied Mathematics*, 229(1):283–293, July 2009.

[17] E. Bogers. *Traffic Information and Learning in Day-to-Day Route Choice*. Ph.D. thesis, TRAIL thesis series T2009/5, Delft University of Technology, Delft, The Netherlands, June 2009.

[18] E. Bogers, F. Viti, and S.P. Hoogendoorn. Joint modeling of ATIS, habit and learning impacts on route choice by laboratory simulator experiments. *Traveler Behavior and Values*, 1926:189–197, 2005.

[19] P.T. Boggs and J.W. Tolle. Sequential quadratic programming. *Acta Numerica*, 4(1): 1–51, January 1995.

[20] M.J. Box. A new method of constrained optimization and a comparison with other methods. *Computer journal*, 8(1):42–52, 1975.

[21] B. Braess. Uber ein Paradoxen der Verkehrsplanung. *Mathematical Methods of Operations Research*, 12(1):258–269, December 1968.

[22] P. Breton, A. Hegyi, B. De Schutter, and J. Hellendoorn. Shock wave elimination/reduction by optimal coordination of variable speed limits. In *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems (ITSC 2002)*, pages 225–230, Singapore, September 2002.

[23] E. Brockfeld, R.D. Kuhne, and P. Wagner. Calibration and validation of microscopic models of traffic flow. *Transportation Research Record*, 1934:179–187, 2005.

[24] F. Busch and G. Kruse. MOTION for SITRAFFIC - a modern approach to urban traffic control. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pages 61–64, Oakland, California, 2001.

[25] E.F. Camacho and C. Bordons. *Model Predictive Control in the Process Industry*. Springer-Verlag, Berlin, Germany, 1995. ISBN 1-85233-694-3.

[26] M. Carey and Y.E. Ge. Comparing whole-link travel time models. *Transportation Research Part B*, 37(10):905–926, December 2003.

[27] E. Cascetta. A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research Part B*, 23(1):1–17, February 1989.

[28] E. Cascetta, P. Coppola, and V. Adamo. A day-to-day dynamic model of departure time and pre-trip route choice in presence of advanced travellers information systems. In *Proceedings of the 7th World Congres on Intelligent Transportation Systems*, Turin, Italy, November 2000.

[29] M.J. Cassidy and J. Rudjanakanoknad. Increasing the capacity of an isolated merge by metering its on-ramp. *Transportation Research Part B*, 39(10):896–913, December 2005.

[30] E. Castillo, J.M. Menéndez, and P. Jiménez. Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B*, 42(5):455–481, june 2008.

[31] H.B. Celikoglu and M. DelÓrco. Mesoscopic simulation of a dynamic link loading process. *Transportation Research Part C*, 15(5):329–344, October 2007.

[32] C.G.Chorus, T.A. Arentze, and H.J.P. Timmermans. Traveler compliance with advice: A bayesiannext term utilitarian perspective. *Transportation Research Part E*, 45(3):486–500, May 2009.

[33] I. Chabibi and Y. He. Dynamic traffic assignment: An analytical tool and its application. In *Proceedings of the International Conference on Intelligent Transportation Systems*, pages 32–38, Washington, D.C., 2000.

[34] C.H. Chen and P. Schonfeld. Work zone lengths for a four-lane road with an alternate route. *Journal of Transportation Engineering-ASCE*, 131(10):780–789, October 2005.

[35] L.L. Chen, A.D. May, and D.M. Auslander. Freeway ramp control using fuzzy set theory for inexact reasoning. *Transportation Research Part A*, 24(1):15–25, January 1990.

[36] R.B. Chen and H.S. Mahmassani. Travel time perception and learning mechanisms in traffic networks. In *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, volume 1894, pages 209–221, Washington, D.C., January 2004.

[37] C.G. Chorus, E.J.E. Molin, and B. van Wee. Use and effects of advanced traveller information services (ATIS): A review of the literature. *Transport Reviews*, 26(2): 127–149, March 2006.

[38] B. Ciuffo, V. Punzo, and V. Torrieri. Comparison of simulation-based and model-based calibrations of traffic-flow microsimulation models. *Transportation Research Record*, 2088:36–44, 2008.

[39] M. Cremer and M. Papageorgiou. Parameter-identification for a traffic flow model. *Automatica*, 17(6):837–843, November 1981.

[40] C.F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, 28B(4):269–287, August 1994.

[41] C.F. Daganzo. *Fundamentals of Transportation and Traffic Operations*. Pergamon Press, United Kingdom, third edition, 1997. ISBN 0-08-042785-5.

[42] C.F. Daganzo and Y. Sheffi. On stochastic models of traffic assignment. *Transportation Science*, 11:253–274, 1977.

[43] D.J. Dailey, F.W. Cathey, and S. Pumrin. An algorithm to estimate mean traffic speed using uncalibrated cameras. *IEEE Transactions on Intelligent Transportation Systems*, 1(2):98–107, June 2000.

[44] L. Davis, editor. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York, New York, 1991. ISBN 0-442-00173-8.

[45] B. De Schutter, H. Hellendoorn, A. Hegyi, M. van den Berg, and S.K. Zegeye. Model-based control of intelligent traffic networks. In R.R. Negenborn, Z. Lukszo, and H. Hellendoorn, editors, *Intelligent Infrastructures*, volume 42 of *Intelligent Systems, Control and Automation: Science and Engineering*, chapter 11, pages 277–310. Springer, Dordrecht, The Netherlands, 2010.

[46] F.P. Deflorio. Evaluation of a reactive dynamic route guidance strategy. *Transportation Research Part C*, 11(5):375–388, October 2003.

[47] C. Diakaki, M. Papageorgiou, and T. Mclean. Simulation studies of integrated corridor control in Glasgow. *Transportation Research Part C*, 5(3/4):211–224, August 1997.

[48] C. Diakaki, M. Papageorgiou, and T. McLean. Integrated traffic-responsive urban corridor control strategy in Glasgow, Scotland: Application and evaluation. *Transportation Research Record*, 1727:101–111, 2000.

[49] C. Diakaki, V. Dinopoulou, K. Aboudolas, and M. Papageorgiou. Signal management in real time for urban traffic networks. Technical report, Technical University of Crete - IST programme, Chania, Greece, July 2002.

[50] R.B. Dial. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5(2):83–111, June 1971.

[51] M. Dicke and K. Brookhuis. Graphical presentation of travel information: the use of full colour information panels. *Dutch Journal of Ergonomics*, 31:6–13, 2006.

[52] M.H.M. Ebben, G.J. Wassink, E. van den Hoogen, and M. Westerman. Samen werken op de A12, Verkeerskundige architectuur in de praktijk. *Verkeerskunde*, 53(3):56–61, 2002. in Dutch.

[53] R.W. Eglese. Simulated annealing: A tool for operations research. *European Journal of Operational Research*, 46(3):271–281, June 1990.

[54] C. Fisk. Some developments in equilibrium traffic assignment. *Transportation Research Part B*, 14(3):243–255, September 1980.

[55] R. Fletcher and S. Leyffer. Numerical experience with lower bounds for MIQP branch-and-bound. *SIAM Journal on Optimization*, 8(2):604–616, May 1998.

[56] M. Florian, M. Mahut, and N. Tremblay. A hybrid optimization-mesoscopic simulation dynamic traffic assignment model. In *Proceedings of the IEEE conference on Intelligent Transportation Systems*, pages 118–121, Oakland, California, 2001.

[57] C.E. García, D.M. Prett, and M. Morari. Model predictive control: Theory and practice — A survey. *Automatica*, 25(3):335–348, May 1989.

[58] N.H. Gartner. Development of demand-responsive strategies for urban traffic control. In P. Thoft-Christensen, editor, *Proceedings of the 11th IFIP Conference on System Modelling and Optimization*, pages 166–174, New York, New York, 1984. Springer-Verlag.

[59] N.H. Gartner, J.D.C. Little, and H. Gabbay. Simultaneous optimization of offsets, splits, and cycle time. *Transportation Research Record*, 596:6–15, 1976.

[60] D. Gazis and C. Liu. Kalman filtering estimation of traffic counts for two network links in tandem. *Transportation Research Part B*, 37(8):737–745, September 2003.

[61] F. Glover and M. Laguna. *Tabu Search*. Kluwer Academic Publishers, Boston, Massachusetts, 1997. ISBN 0-7923-9965-X.

[62] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, Massachusetts, 1989. ISBN 97-8020-115-7673.

[63] H. Haj-Salem and M. Papageorgiou. Ramp metering impact on urban corridor traffic: Field results. *Transportation Research Part A*, 29(4):303–319, Juli 1995.

[64] A. Hegyi. *Model Predictive Control for Integrating Traffic Control Measures*. Ph.D. thesis, TRAIL thesis series T2004/2, Delft University of Technology, Delft, The Netherlands, February 2004. ISBN 90-5584-053-X, http://www.hegyi.info/thesis/dep_hegyi_20040203.pdf.

[65] A. Hegyi, B. De Schutter, H. Hellendoorn, and T. van den Boom. Optimal coordination of ramp metering and variable speed control — An MPC approach. In *Proceedings of the 2002 American Control Conference*, pages 3600–3605, Anchorage, Alaska, May 2002.

[66] A. Hegyi, B. De Schutter, and J. Hellendoorn. Optimal coordination of variable speed limits to suppress shock waves. *IEEE Transactions on Intelligent Transportation Systems*, 6(1):102–112, March 2005.

[67] A. Hegyi, D. Girimonte, R. Babuška, and B. De Schutter. A comparison of filter configurations for freeway traffic state estimation. In *Proceedings of the 9th IEEE conferense on Intelligent Transportation Systems (ITSC2006)*, pages 1029–1034, Toronto, Canada, September 2006.

[68] D. Helbing, A. Hennecke, V. Shvetsov, and M. Treiber. Micro- and macro-simulation of freeway traffic. *Mathematical and Computer Modelling*, 35(5–6):517–547, March 2002.

[69] B. Heutinck, M. van den Berg, J. Hellendoorn, and L.H. Immers. Route guidance during maintenance works, a case study. In *Proceedings of the 11th IFAC Symposium on Control in Transportation Systems*, Delft, The Netherlands, Augustus 2006.

[70] B.H. Heutinck. Eindhoven under construction: Microscopic traffic simulation and dynamic alternative route guidance. Master's thesis, Delft University of Technology, Delft, The Netherlands, November 2005. TNO report 05-7N-289.

[71] S.P. Hoogendoorn and P.H.L. Bovy. Generic gas-kinetic traffic systems modeling with applications to vehicular traffic flow. *Transportation Research Part B*, 35(4): 317–336, May 2001.

[72] S.P. Hoogendoorn and P.H.L. Bovy. State-of-the-art of vehicular traffic flow modelling. *Journal of Systems Control Engineering — Proceedings of the Institution of Mechanical Engineers, Part I*, 215(14):283–303, 2001.

[73] S.P. Hoogendoorn, B. De Schutter, and H. Schuurman. Decision support in dynamic traffic management. Real-time scenario evaluation. *European Journal of Transport and Infrastructure Research*, 3(1):21–38, 2003.

[74] A.J. Horowitz, I. Weisser, and T. Notbohm. Diversion from a rural work zone owing to a traffic-responsive variable message signage system. *Transportation Research Record*, 2003.

[75] http://www.tomlab.biz/. *Tomlab*, 2008. Last visited on June 1, 2008.

[76] T.-Y. Hu and H.S. Mahmassani. Day-to-day evolution of network flows under real-time information and reactive signal control. *Transportation Research Part C*, 5(1): 51–69, February 1997.

[77] G.M. Huang and W.-L. Hsieh. Exact convergence of a parallel textured algorithm for data network optimal routing-problems. *IEEE Transactions on Parallel and Distributed Systems*, 6(11):1132–1146, November 1995.

[78] R. Jayakrishnan, H.S. Mahmassani, and T.Y. Hu. An evaluation tool for advanced traffic information and management systems in urban networks. *Transportation Research Part C*, 2(3):129–147, September 1994.

[79] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. New York: Academic Press, eleventh edition, 1970. ISBN 0123815509, 9780123815507.

[80] M. Jha, S. Madanat, and S. Peeta. Perception updating and day-to-day travel previous termchoicenext term dynamics in traffic networks with information provision. *Transportation Research Part C*, 6(3):189–212, June 1998.

[81] A. Karimi, A. Hegyi, B. De Schutter, H. Hellendoorn, and F. Middelham. Integration of dynamic route guidance and freeway ramp metering using model predictive control. In *Proceedings of the 2004 American Control Conference (ACC 2004)*, pages 5533–5538, Boston, Massachusetts, 2004.

[82] H.R. Kashani and G.N. Saridis. Intelligent control for urban traffic systems. *Automatica*, 19(2):191–197, March 1983.

[83] B.S. Kerner and H. Rehborn. Experimental features and characteristics of traffic jams. *Physical Review E*, 53(2):1297–1300, February 1996.

[84] L.A. Klein, M.K. Mills, and D.R.P. Gibson. Traffic detector handbook, third edition. Technical Report FHWA-HRT-06-108, Federal Highway Administration, Washington, D.C., October 2006.

[85] T.G. Kolda, R.M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Journal of Optimization*, 45 (3):385–482, September 2003.

[86] A. Kotsialos and M. Papageorgiou. Motorway network traffic control systems. *European Journal of Operational Research*, 152:321–333, January 2004.

[87] A. Kotsialos, M. Papageorgiou, and A. Messmer. Optimal coordinated and integrated motorway network traffic control. In *Proceedings of the 14th International Symposium of Transportation and Traffic Theory (ISTTT)*, pages 621–644, Jerusalem, Israel, July 1999.

[88] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham. Traffic flow modeling of large-scale motorway using the macroscopic modeling tool METANET. *IEEE Transactions on Intelligent Transportation Systems*, 3(4):282–292, December 2002.

[89] A. Kotsialos, M. Papageorgiou, M. Mangeas, and H. Haj-Salem. Coordinated and integrated control of motorway networks via non-linear optimal control. *Transportation Research Part C*, 10(1):65–84, February 2002.

[90] M. Kraan, N. van der Zijpp, B. Tutert, T. Vonk, and D. van Megen. Evaluating networkwide effects of variable message signs in The Netherlands. *Transportation Research Record*, 1689:60–76, 1999.

[91] P. Kronborg and F. Davidsson. MOVA and LHOVRA: traffic signal control for isolated intersections. *Traffic engineering & control*, 34(4):195–200, 1993.

[92] M. Kuwahara and E.C. Sullivan. Estimating origin-destination matricesnext term from roadside survey data. *Transportation Research Part B*, 21(3):233–248, June 1987.

[93] I.D. Landau. From robust control to adaptive control. *Control Engineering Practice*, 7(9):1113–1124, September 1999.

[94] J.H. Lee and Z.H. Yu. Tuning of model-predictive controllers for robust performance. *Computers & Chemical engineering*, 18(1):15–37, January 1994.

[95] H. Lenz, R. Sollacher, and M. Lang. Non-linear speed control for a continuum theory traffic flow. In *Proceedings of the 14th World Congress of IFAC*, volume Q, pages 67–72, Beijing, China, January 1999.

[96] M.J. Lighthill and G.B. Whitham. On kinematic waves, II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences*, 229(1178):317–345, May 1955.

[97] S. Lin, Y. Xi, and Y. Yang. Short-term traffic flow forecasting using macroscopic urban traffic network model. In *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems*, pages 134–138, Beijing, China, October 2008.

[98] J. Linderoth and T. Ralphs. Noncommercial software for mixed-integer linear programming. *Optimization Online*, January 2005. Available at www.optimization-online.org/DB_HTML/2004/12/1028.html.

[99] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, New Jersey, second edition, 1999. ISBN 97-8013-881-6407.

[100] J.M. Maciejowski. *Predictive Control with Constraints*. Prentice Hall, Harlow, United Kingdom, 2002. ISBN 0-201-39823-0.

[101] M. Maher. Algorithms for logit-based stochastic user equilibrium assignment. *Transportation Research Part B*, 32(8):539–549, November 1998.

[102] M.J. Maher and P.C. Hughes. A probit-based stochastic user equilibrium assignment model. *Transportation Research Part B*, 31(4):341–355, August 1997.

[103] H.S. Mahmassani, N.N. Huynha, K. Srinivasan, and M. Kraanc. Tripmaker choice behavior for shopping trips under real-time information: model formulation and results of stated-preference internet-based interactive experiments. *Journal of Retailing and Consumer Services*, 10(6):311–321, November 2003.

[104] V. Mauro and C. Di Taranto. Utopia. In *Proceedings of the 6th IFAC/IFIP/IFORS Symposium on Control, Computers, Communication on Transportation*, pages 245–252, Paris, France, 1989.

[105] A.D. May. *Traffic Flow Fundamentals*. Prentice-Hall, Englewood Cliffs, New Jersey, 1990. ISBN 0-13-926072-2.

[106] A. Messmer and M. Papageorgiou. METANET: A macroscopic simulation program for motorway networks. *Traffic Engineering and Control*, 31:466–470, September 1990.

[107] A. Messmer, M. Papageorgiou, and N. Mackenzie. Automatic control of variable message signs in the interurban Scottish highway network. *Transportation Research Part C*, 6(3):173–187, June 1998.

[108] K.M. Miettinen. *Nonlinear multiobjective optimization*, volume 12. Kluwer Academic Publishers, Boston, 1999. ISBN 0-7923-8278-1.

[109] L. Mihaylova, R. Boel, and A. Hegyi. Freeway traffic estimation within particle filtering framework. *Automatica*, 43(2):290–300, February 2007.

[110] Ministry of Transport, Public Works and Water Management. *HANDBOOK Sustainable Traffic Management - A Guide for Users*. AVV transport research center, Rotterdam, The Netherlands, first edition, November 2003.

[111] Ministry of Transport, Public Works and Watermanagement. Mobiliteitsonderzoek Nederland 2005, tabellenboek. Technical report, AVV transport research center, Rotterdam, The Netherlands, 2006. http://www.rvs-avv.nl, in Dutch.

[112] Ministry of Transport, Public Works and Watermanagement. *MaatregelenCatalogus Benutten*. AVV transport research center, Rotterdam, The Netherlands, 2005. ISBN 903693617 9. in Dutch.

[113] Ministry of Transport, Public Works and Watermanagement. Bereikbaarheidsmonitor hoofdwegennet 2007. Technical report, DVS transport research center, Rotterdam, The Netherlands, Juli 2008. Available at www.rijkswaterstaat.nl, in Dutch.

[114] J.D. Murchland. Braess's paradox of traffic flow. *Transportation Research*, 4(4): 391–394, December 1970.

[115] D. Ngoduy. *Macroscopic Discontinuity Modeling for Multiclass Multilane Traffic Flow Operations*. Ph.D. thesis, TRAIL thesis series T2006/3, Delft University of Technology, Delft, The Netherlands, April 2006.

[116] D. Ngoduy, S.P. Hoogendoorn, and J. van Zuylen. An automated calibration procedure for macroscopic traffic flow models. In *Proceedings of the 10th IFAC Symposium on Control in Transportation Systems*, pages 1–16, Tokio, Japan, 2003.

[117] US Department of Transportation. Application of analysis tools to evaluate the travel impacts of highway reconstruction with the emphasis on microcomputer applications. Technical Report FHWA-ED-89-02, Federal Highway Administration, March 1989.

[118] Omnitrans. *Introductory guide to OmniTRANS*. Omnitrans International B.V., March 2008. Available at www.omnitrans-international.com.

[119] P.V. Palacharla and P.C. Nelson. Application of fuzzy logic and neural networks for dynamic travel time estimation. *International Transactions in Operational Research*, 6(1):145–160, 1999.

[120] M. Papageorgiou. Certainty equivalent open-loop feedback control applied to multireservoir networks. *IEEE Transactions on Automatic Control*, 33(4):392–399, April 1988.

[121] M. Papageorgiou and A. Kotsialos. Freeway ramp metering: An overview. *IEEE Transactions on Intelligent Transportation Systems*, 3(4):271–280, December 2002.

[122] M. Papageorgiou and A. Kotsialos. Freeway ramp metering: An overview. In *Proceedings of the 3rd International Conference on Intelligent Transportation Systems (ITSC2000)*, pages 228–239, Dearborn, Michigan, October 2000.

[123] M. Papageorgiou, H. Hadj-Salem, and J.M. Blosseville. ALINEA: A local feedback control law for on-ramp metering. *Transportation Research Record*, 1320:58–64, 1991.

[124] M. Papageorgiou, M. Ben-Akiva, J. Bottom, P.H.L. Bovy, S.P. Hoogendoorn, N.B. Hounsell, A. Kotsialos, and M. McDonald. *Handbooks in Operations Research and Management Science*, volume 14: Transportation, chapter 11 ITS and Traffic Management, pages 714–774. Elsevier, Amsterdam, The Netherlands, 2007. ISBN 978-0-444-51428-8.

[125] M. Papageourgiou. Some remarks on macroscopic traffic flow modelling. *Transportation Research Part A*, 32(5):323–329, September 1998.

[126] P.M. Pardalos and M.G.C. Resende, editors. *Handbook of Applied Optimization*. Oxford University Press, Oxford, United Kingdom, 2002. ISBN 0-19-512594-0.

[127] E.S. Park, J.H. Lim, I.H. Suh, and Z. Bien. Hierarchical optimal control of urban traffic networks. *International Journal of Control*, 40(4):813–829, October 1984.

[128] A. Paz and S. Peeta. On-line calibration of behavior parameters for behavior-consistent route guidance. *Transportation Research Part B*, 43(4):403–421, May 2009.

[129] Peek Traffic. *UTOPIA/SPOT - Technical Reference Manual*. Peek Traffic, Amersfoort, The Netherlands, January 2002.

[130] S. Peeta and H.S. Mahmassani. Multiple user classes real-time traffic assignment for online operations: a rolling horizon solution framework. *Transportation Research Part C*, 3(2):83–98, April 1995.

[131] S. Peeta and A.K. Ziliaskopoulos. Foundations of dynamic traffic assignment: the past, the present and the future. *Networks and Spatial Economics*, 1(3-4):233–265, September 2001.

[132] D. Petz and R. Temesi. Means of positive numbers and matrices. *SIAM journal on Matrix Analysis and Applications*, 27(3):712–720, January 2005.

[133] D.A. Pierre. *Optimization Theory with Applications*. Courier Dover Publications, New York, New York, 1986. ISBN 97-804-86652-054.

[134] W.B. Powell and Y. Sheffi. The convergence of equilibrium algorithms with predetermined step sizes. *Transportation Science*, 16(1):45–55, January 1982.

[135] J. Prakash, S. Narasimhan, and S.C. Patwardhan. Integrating model based fault diagnosis with model predictive control. *Industrial and engineering chemistry research*, 44(12):4344–4360, June 2005.

[136] PTV. Vissim – traffic flow simulation. Technical report, PTV, Karlsruhe, Germany, 2003. http//:www.ptv.de.

[137] Quadstone. Paramics system overview. Technical report, Quadstone limited, Edinburgh, Scotland, 2002. www.paramics-online.com.

[138] H. Rakha and W. Zhang. Estimating traffic stream space mean speed and reliability from dual- and single-loop detectors. *Transportation Research Record*, 1925:38–47, 2005.

[139] P.I. Richards. Shock waves on the highway. *Operations Research*, 4:42–57, 1956.

[140] D.I. Robertson and R.D. Bretherton. Optimizing networks of traffic signals in real time - the SCOOT method. *IEEE Transactions on Vehicular Technology*, 40(1):11–15, 1991.

[141] L.O. Santos, L.T. Biegler, and J.A.A.M. Castro. A tool to analyze robust stability for constrained nonlinear mpc. *Journal of Process Control*, 18(3-4):383–390, March 2008.

[142] Y. Sheffi. *Urban Transportation Networks: Equilibrium Analysys with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, New Jersey, 1985. ISBN 97-801-39397-295.

[143] Y. Sheffi and W. Powell. An algorithm for the equilibrium assignment problem with random link times. *Networks*, 12(2):191–207, Summer 1982.

[144] Y. Sheffy and W. Powell. A comparison of stochastic and deterministic traffic assignment over congested networks. *Transportation Research Part B*, 15(1):53–64, 1981.

[145] S. Smulders. Control of freeway traffic flow by variable speed limits. *Transportation Research Part B*, 24(2):111–132, 1990.

[146] SRE. SRE verkeers- en milieumodel. Brochure, available at www.SRE.nl, 2006. in Dutch.

[147] K.K. Srinivasan and H.S. Mahmassani. Analyzing heterogeneity and unobserved structural effects in route-switching behavior unter ATIS: a dynamic kernal logit formulation. *Transportation Research Part B*, 37(9):793–814, November 2003.

[148] H. Taale. *Integrated Anticipatory Control of Road Networks. A game theoretical approach*. Ph.D. thesis, TRAIL thesis series T2008/15, Delft University of Technology, Delft, The Netherlands, December 2008.

[149] H. Taale and F. Middelham. Ten years of ramp-metering in The Netherlands. In *Proceedings of the 10th International Conference on Road Transport Information and Control*, IEE Conference Publication No. 472, pages 106–110, London, United Kingdom, April 2000.

[150] H. Taale and H.J. van Zuylen. Traffic control and route choice: Occurrence of instabilities. In *Proceedings of the 5th TRAIL Congress 1999: Five Years Crossroads of Theory and Practice*, volume 2, pages 1–19, Rotterdam, The Netherlands, November 1999.

[151] H. Taale, G. Bootsma, and M.T. te Wierik. Droomrenovatie ge-evalueerd, Leren van het groot onderhoud aan de A1. *Verkeerskunde*, 54(2):54–59, 2003. in Dutch.

[152] H. Taale, S. Hoogendoorn, M. van den Berg, and B. De Schutter. Anticiperende netwerkregelingen. *NM Magazine*, 1(4):22–27, December 2006. in Dutch.

[153] *METANET – A simulation program for motorway networks*. Technical University of Crete, Dynamic Systems and Simulation Laboratory and A. Messmer, July 2000.

[154] The MathWorks. *Optimization Toolbox User's Guide – Version 2.1*. Natick, Massachusetts, 2000. Available at www.mathworks.com.

[155] The MathWorks. *Genetic Algorithm and Direct Search Toolbox 2 — User's Guide*. Natick, Massachusetts, 2007. Available at www.mathworks.com.

[156] M. van den Berg, A. Hegyi, B. De Schutter, and J. Hellendoorn. A macroscopic traffic flow model for integrated control of freeway and urban traffic networks. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 2774–2779, Maui, Hawaii, December 2003.

[157] M. van den Berg, T. Bellemans, B. De Schutter, B. De Moor, and J. Hellendoorn. Anticipative ramp metering control using dynamic traffic assignment. In *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC 2004)*, pages 503–508, Washington, D.C., October 2004.

[158] M. van den Berg, B. De Schutter, A. Hegyi, and J. Hellendoorn. Model predictive control for mixed urban and freeway networks. In *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2004. Paper 04-3327.

[159] M. van den Berg, B. De Schutter, and H. Hellendoorn. Integrated model predictive control for mixed urban and freeway networks. In P.H.L. Bovy, editor, *Proceedings of the 8th TRAIL Congress 2004 — A World of Transport, Infrastructure and Logistics — Selected Papers*, pages 27–49, Rotterdam, The Netherlands, November 2004.

[160] M. van den Berg, T. Bellemans, B. De Schutter, B. De Moor, and J. Hellendoorn. Control of traffic with anticipative ramp metering. In *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2005. Paper 05-0252.

[161] M. van den Berg, B. De Schutter, and J. Hellendoorn. Information providing as a control measure in an integrated traffic control approach. In *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems (ITSC 2005)*, pages 91–96, Vienna, Austria, September 2005.

[162] M. van den Berg, B. De Schutter, and H. Hellendoorn. Effects of on-ramp and off-ramp metering on queue forming in urban traffic networks. In *Proceedings of the 11th IFAC Symposium on Control in Transportation Systems*, pages 130–135, Delft, The Netherlands, August 2006.

[163] M. van den Berg, A. Hegyi, B. De Schutter, and H. Hellendoorn. Influencing long-term route choice by traffic control measures, a basic model study. In *Proceedings of the 9th TRAIL Congress 2006 — TRAIL in Motion — CD-ROM*, Rotterdam, The Netherlands, November 2006.

[164] M. van den Berg, A. Hegyi, B. De Schutter, and J. Hellendoorn. Integrated traffic control for mixed urban and freeway networks: A model predictive control approach. *European Journal of Transport and Infrastructure Research*, 7(3):223–250, September 2007.

[165] M. van den Berg, B. De Schutter, H. Hellendoorn, and A. Hegyi. Day-to-day route choice control in traffic networks — A model predictive control approach based on mixed integer linear programming. In *Proceedings of the 10th TRAIL Congress 2008 — TRAIL in Perspective — CD-ROM*, Rotterdam, The Netherlands, October 2008.

[166] M. van den Berg, B. De Schutter, J. Hellendoorn, and A. Hegyi. Influencing route choice in traffic networks: A model predictive control approach based on mixed-integer linear programming. In *Proceedings of the 17th IEEE International Conference on Control Applications*, pages 299–304, San Antonio, Texas, September 2008.

[167] M. van den Berg, A. Hegyi, B. De Schutter, and J. Hellendoorn. Influencing long-term route choice by traffic control measures — A model study. In *Proceedings of the 17th IFAC World Congress*, pages 13052–13057, Seoul, Korea, July 2008.

[168] M. van den Berg, M. Burger, B. De Schutter, and J. Hellendoorn. Practical issues for model-based traffic control. Technical Report 09-032, Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands, 2009. Submitted for publication to a journal. July 2009.

[169] M. van den Berg, B. De Schutter, A. Hegyi, and H. Hellendoorn. Day-to-day route choice control in traffic networks with time-varying demand profiles. In *Proceedings of the European Control Conference 2009*, pages 1776–1781, Budapest, Hungary, August 2009.

[170] M. van den Berg, B. De Schutter, H. Hellendoorn, and A. Hegyi. Control of day-to-day route choice in traffic networks with overlapping routes. In *Proceedings of the 12th IFAC Symposium on Transportation Systems*, pages 556–561, Redondo Beach, California, September 2009.

[171] J. van den Bosch, W. van der Hoeven, and J. Roos. A28 bij Zwolle goed voorbereid op de schop, Dynasmart toont effect verkeersmaatregelen bij werk in uitvoering op A28. *Verkeerskunde*, 55(2):20–25, 2004. in Dutch.

[172] J.H. van der Lee, W.Y. Svrcek, and B.R. Young. A tuning algorithm for model predictive controllers based on genetic algorithms and fuzzy decision making. *ISA Transactions*, 47(1):53–59, 2008.

[173] R. van Katwijk. *Multi-Agent Look-Ahead Traffic Adaptive Control*. Ph.D. thesis, TRAIL thesis series T2008/3, Delft University of Technology, Delft, The Netherlands, January 2008.

[174] R.T. van Katwijk. Road traffic services on a navigational level: proof of concept. Technical report, TNO, January 2005.

[175] J.W.C. van Lint. *Reliable Travel Time Prediction for Freeways*. Ph.D. thesis, TRAIL thesis series T2004/3, Delft University of Technology, Delft, The Netherlands, June 2004.

[176] D. Van Vliet. The Frank-Wolfe algorithm for equilibrium traffic asignment viewed as a variational inequality. *Transportation Research Part B*, 21(1):87–89, February 1987.

[177] Y. Wang and M. Papageorgiou. A predictive feedback routing control strategy for freeway network traffic. In *Proceedings of the 2002 American Control Conference*, pages 3606–3611, Anchorage, Alaska, May 2002.

[178] Y. Wang and M. Papageorgiou. Real-time freeway traffic state estimationnext term based on extended Kalman filter: a general approach. *Transportation Research Part B*, 39(2):141–167, February 2005.

[179] Y. Wang, M. Papageorgiou, and A. Messmer. Predictive feedback routing control strategy for freeway network traffic. *Transportation Research Record*, 1856:62–73, 2003.

[180] Y. Wang, M. Papageorgiou, and A. Messmer. RENAISSANCE - a unified macroscopic model-based approach to real-time freeway network traffic surveillance. *Transportation Research Part C*, 14(3):190–212, June 2006.

[181] Y.J. Wang and J.B. Rawlings. A new robust model predictive control method. ii: examples. *Journal of Process Control*, 14(3):249–262, April 2004.

[182] J.G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers, Part II*, 1(2):325–378, 1952.

[183] A. Wilson. *Handboek verkeerslichtenregelingen*, volume 213. National Information and Technology Platform for Transport, Infrastructure and Public space (CROW), Ede, The Netherlands, January 2006. ISBN 90 6628 444 7.

[184] M.B. Wisten and M.J. Smith. Distributed computation of dynamic traffic equilibria. *Transportation Research Part C*, 5(2):77–93, April 1997.

[185] B. Wolson and W.C. Taylor. Analysis of intersection delay under real-time adaptive signal control. *Transportation Research Part C*, 7(1):53–72, October 1999.

[186] J. Wu and G.L. Chang. Heuristic method for optimal diversion control in freeway corridors. *Transportation Research Record*, 1667:8–15, 1999.

[187] H. Yang and S. Yagar. Traffic assignment and traffic control in general freeway-arterial corridor systems. *Transportation Research Part B*, 28(6):463–486, December 1994.

[188] H. Yin, S.C. Wong, J. Xu, and C.K. Wong. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C*, 10(2):85–98, April 2000.

[189] Y. Yin. Robust optimal traffic signal timing. *Transportation research Part B*, 42(10): 911–924, December 2008.

[190] H.M. Zhang, S.G. Ritchie, and R. Jayakrishnan. Coordinated traffic-responsive ramp control via nonlinear state feedback. *Transportation Research Part C*, 9(5):337–352, October 2001.

[191] X. Zhou and H.S. Mahmassani. A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework. *Transportation Research Part B*, 41(8):823–840, October 2007.

Samenvatting

Motivatie

De capaciteit van het huidige wegennetwerk is niet groot genoeg om de verkeersvraag aan te kunnen. Dit zorgt ervoor dat files op het wegennet een dagelijks terugkerend verschijnsel zijn geworden. Hoewel het een illusie is dat het fileprobleem binnen een paar jaar volledig opgelost kan worden, kan de huidige situatie wel verbeterd worden. Dit is noodzakelijk omdat files nadelige gevolgen hebben voor weggebruikers, zoals langere en onbetrouwbare reistijden en hogere kosten. Ook hebben files een nadelige invloed op de omgeving en de omwonenden: ze veroorzaken vervuiling en geluidsoverlast en zorgen voor onveilige situaties in bijvoorbeeld woonwijken en winkelcentra. Het doel van het onderzoek dat beschreven is in dit proefschrift is het ontwikkelen van methodes om het verkeer te regelen waarmee de situatie verbeterd kan worden voor zowel de weggebruikers als voor de omgeving.

Gecombineerde regeling van snelwegen en locale wegen

Doordat er steeds meer wegen komen op een klein oppervlak, wordt de invloed die de verschillende soorten wegen op elkaar hebben steeds groter. Om het verkeer op plaatsen waar snelwegen en lokale wegen sterk met elkaar verbonden zijn, effectief te kunnen blijven regelen, moeten de maatregelen op beide soorten wegen gekoppeld worden. Hiervoor ontwikkelen we een model dat zowel het snelwegverkeer als het stadsverkeer kan beschrijven. Voor het snelwegverkeer gebruiken we met het bestaande model METANET. Voor het stadsverkeer gebruiken we een wachtrijmodel ontwikkeld door Kashani en Saridis, uitgebreid met horizontale rijen, het blokkeren van kruispunten en een kleinere simulatietijdsstap. De twee modellen worden met elkaar verbonden door het modelleren van toe- en afritten. Het resultaat is een efficiënt model dat geschikt is voor gebruik in modelgebaseerde on-line verkeersregelmethoden. We ontwikkelen een regelmethode gebaseerd op ‘model predictive control’ (modelgebaseerd voorspellend regelen), waarbij een voorspelling gemaakt wordt van de verkeersstromen, en aan de hand van deze voorspelling de beste instellingen voor de maatregelen bepaald worden. Voor het maken van de voorspelling gebruiken we het hierboven beschreven model voor snelwegen en locale wegen.

Regelen van routekeuze

Verkeersmaatregelen beïnvloeden indirect de routekeuze van weggebruikers. Dit effect kan gebruikt worden om de efficiëntie van de maatregelen te vergroten. Het routekeuze-proces bestaat uit twee delen: de routekeuze binnen een dag en de routekeuze van dag tot dag. De routekeuze binnen een dag beschrijft de keuzes die weggebruikers maken terwijl ze onderweg zijn, terwijl de dag tot dag routekeuze beschrijft hoe de voorkeur voor een route verandert over verschillende dagen.

Eerst bekijken we de keuze binnen een dag, met als voorbeeld een toeritdoseringsinstallatie. Deze installatie zorgt voor een rij op de toerit die de reistijd via deze toerit langer maakt. Als gevolg van deze langere reistijd zullen sommige weggebruikers een andere route kiezen. We ontwikkelen twee verschillende modellen die dit effect beschrijven, en gebruikt kunnen worden in de regeling van de toeritdoseringsinstallatie: een dynamisch toedelingsmodel en een routekeuzemodel gebaseerd op een look-up tabel. Op deze manier kan de regelmethode voor de toeritdosering verbeterd worden.

De routekeuze van dag tot dag beschrijven we met behulp van het tweede model, gebaseerd op de look-up tabel. Hierbij worden de waardes in de tabel aangepast met behulp van ‘Bayesian learning’. Het model veronderstelt dat de routekeuze gebaseerd is op een combinatie van de huidige voertuigdichtheid en de reistijden op eerdere dagen. Met dit model bekijken we behalve toeritdosering ook de effecten van afritdosering en de resultaten die bereikt kunnen worden met het tonen van reistijdinformatie op dynamische route-informatiepanelen. Hierbij nemen we aan dat als de reistijd op deze panelen duidelijk verschilt van de reistijd die de weggebruikers verwachten, een deel van de weggebruikers een andere route zal kiezen. We ontwikkelen een modelgebaseerde regelmethode die de reistijdinformatie op dynamische route-informatiepanelen combineert met variabele snelheidslimieten, om zo de routekeuze van weggebruikers actief te kunnen beïnvloeden.

De routekeuzemodellen zoals hiervoor beschreven zijn rekenintensief. Daarom hebben we ook een eenvoudig routekeuzemodel ontwikkeld dat de routekeuze van dag tot dag beschrijft. Op basis van dit model kunnen relatief snel schattingen van het routekeuzegedrag gemaakt worden, waardoor het model geschikt is om een eerste indruk te krijgen van de verkeersverdeling, om gebruikt te worden in on-line optimalisatie-algoritmes of om als startpunt te gebruiken voor complexere optimalisatie-algoritmes. In dit proefschrift gebruiken we het model in een regelmethode voor variabele snelheden en intensiteitsbeperkingen. Het doel van deze methode is het beïnvloeden van de routekeuze zodat de prestatie van het verkeersnetwerk verbeterd kan worden.

Het installeren van maatregelen in de praktijk

Voordat verkeersmaatregelen in de praktijk toegepast kunnen worden, moet aandacht besteed worden aan verschillende praktische zaken. We geven een kort overzicht van onderwerpen die van belang zijn, en onderzoeken specifiek de invloed van het bepalen van het gemiddelde van de snelheidsmetingen. Hiervoor vergelijken we verschillende methoden om het gemiddelde te bepalen, namelijk het tijdsgemiddelde, harmonisch gemiddelde, geometrisch gemiddelde, plaatsgemiddelde en het geschatte plaatsgemiddelde gebaseerd op de variatie van de instantaan gemeten snelheden van voertuigen in een segment en op de variantie van de snelheden van voertuigen op het moment dat ze het meetpunt passeren. Elk van deze methoden is toegepast in een snelheidsregeling op de snelweg, om de prestaties van de regelaar te kunnen vergelijken.

Conclusies

Het verkeersnetwerk kan efficiënter benut worden als geavanceerde regelmethodes worden gebruikt. Deze methodes kunnen bestaande en nieuwe maatregelen gebruiken om de kosten voor de weggebruikers te verlagen en om de routekeuze van weggebruikers te beïnvloeden, waardoor files en wachtrijen verminderd of verplaatst kunnen worden. Dit leidt tot economische winst door de lagere reistijden, verbetert de leefbaarheid door het reduceren van vervuiling en geluidsoverlast, en verbetert de veiligheid door het realiseren van lagere voertuig-intensiteiten in woonwijken en stadscentra.

Summary

Motivation
The growth of our road infrastructure cannot keep up with the growing mobility of people, and the corresponding increase in traffic demand. This results in daily congestion on the freeways. It is an illusion that the problem of congestion can be solved completely within a few years, but it is possible to improve the current situation. This is necessary since the congestion on the roads has disadvantages for the drivers, including long travel times and high economic costs. It has also disadvantages for the surroundings of the roads, where the increased traffic load results in e.g., pollution, noise, and unsafety in residential areas. The goal of this thesis is to develop model-based traffic control methods that improve the situation for the drivers as well as for the environment.

Mixed urban and freeway control
Due to the growing density of the road networks, freeways and urban networks become tightly coupled. This requires that the control on the two types of roads should also be coupled. Therefore we develop a mixed urban-freeway model that combines a macroscopic freeway model with an urban queue length model. For the macroscopic model we use the traffic flow model METANET. The urban queue length model is based on a model developed by Kashani and Saridis, extended with horizontal queues, blocking effects, and a shorter time step. The two models are coupled via the modeling of on-ramps and off-ramps. The obtained macroscopic model can simulate traffic flows efficiently, and thus is suitable for the use in a model-based control setting. We develop such a model-based control method that uses model predictive control, with the mixed urban-freeway model as prediction model.

Route choice control
Control measures can also be used to influence route choice. Route choice is a complicated process that can be divided into two main processes with a different time scale. The within-day route choice focuses on the choices that drivers make during their trip, while the day-to-day route choice describes the change in route choice from one day to the next.

We first discuss the effect of ramp metering on within-day route choice. By installing a ramp metering installation at an on-ramp, the density – and thus the travel time – on the freeway as well as on the on-ramp itself is changed, which influences the route choice. We develop two different methods to include route choice in model-based controllers: a dynamic traffic assignment model, and a model based on a look-up table determined via Bayesian learning. Second, we investigate day-to-day route choice using the Bayesian learning model. We assume that drivers base their route choice on a combination of the

density and the corresponding travel times experienced on previous days. With the model based on Bayesian learning, in addition to the on-ramp metering, we also briefly explore the effects that can be obtained with the use of off-ramp metering. Another measure that can be used to influence the route choice is displaying travel time information on dynamic route information panels. Displaying travel times with a large enough difference can encourage drivers to change their route choice. We model the drivers' reaction on the route information, and develop a controller that actively influences the route choice of the drivers using information on dynamic route information panels in combination with variable speed limits.

Since route choice models as described above in general require large computational efforts, we also formulate a simplified route choice model for day-to-day route choice that can be used to obtain fast predictions of the route choice behavior, and that is suitable to obtain a first impression of the traffic assignment, for use in on-line optimization algorithms, or as initial value for more complex optimization algorithms. We use this model in a model-based control setting where the objective of the controller is to influence the route choice, and investigate in particular speed limit control and outflow control.

Practical control issues

To apply model based controllers in practice, several practical issues have to be considered. We present a short overview of interesting issues, and next we explicitly investigate the effect of averaging method that is used for the speed measurements. We compare the time mean speed, harmonic mean speed, geometric mean speed, time average space mean speed, and the estimated space mean speed based on instantaneous speed variance and based on local speed variance. All averaging methods are applied in a freeway speed limit control method, to investigate the influence of the averaging method on the controller performance.

Conclusions

The current traffic infrastructure can be used more effectively when advanced control algorithms are used. Existing traffic control measures can be used to decrease the costs for the drivers, and to relocate the traffic flows via influencing the route choice. This results in economical benefits due to shorter travel times, environmental benefits due to the reduction of pollution and noise, and safety benefits due to the lower flows in urban areas.

Curriculum vitae

Monique van den Berg was born on December 8, 1979 in Leiderdorp, The Netherlands. She finished her pre-university education in 1998, at the Vlietland College in Leiden, The Netherlands. After this, Monique van den Berg studied Electrical Engineering at Delft University of Technology, where she specialized in Control Engineering. She graduated in 2003, with her master thesis entitled ‘Model Predictive Control for Mixed Urban and Freeway Networks’.

Since 2003 Monique van den Berg has been working on the research project ‘Advanced Multi-agent Information and Control for Integrated multi-class networks’ at the Delft Center for Systems and Control. She worked on the Ph.D. project ’Development of advanced multi-agent control strategies for multi-class traffic networks’. The research of her Ph.D. project focused on three different topics: the integrated control of traffic networks consisting of freeways as well as urban roads, influencing the route choice of drivers with currently available control measures, and investigating implementation aspects of advanced control approaches. This research has been performed under the supervision of Prof.dr.ir. Hans Hellendoorn and Prof.dr.ir. Bart De Schutter.

In 2009 Monique van den Berg started at the Dutch Ministry of Transportation, Rijkswaterstaat, as advisor considering strategic traffic flow models.

TRAIL Thesis Series

A series of The Netherlands TRAIL Research School for theses on transport, infrastructure and logistics.

Berg, M. van den, *Integrated Control of Mixed Traffic Networks using Model Predictive Control*, T2010/4, April 2010, TRAIL Thesis Series, The Netherlands

Top, J. van den, *Modelling Risk Control Measures in Railways*, T2010/3, April 2010, TRAIL Thesis Series, The Netherlands

Craen, S. de, *The X-factor: A longitudinal study of calibration in young novice drivers*, T2010/2, March 2010, TRAIL Thesis Series, The Netherlands

Tarau, A.N., *Model-based Control for Postal Automation and Baggage Handling*, T2010/1, January 2010, TRAIL Thesis Series, The Netherlands

Knoop, V.L., *Road Incidents and Network Dynamics. Effects on driving behaviour and traffic congestion*, T2009/13, December 2009, TRAIL Thesis Series, The Netherlands

Baskar, L.D., *Traffic Management and Control in Intelligent Vehicle Highway Systems*, T2009/12, November 2009, TRAIL Thesis Series, The Netherlands

Konings, R., *Intermodal Barge Transport. Network Design, Nodes and Competitiveness*, T2009/11, November 2009, TRAIL Thesis Series, The Netherlands

Kusumaningtyas, I., *Mind Your Step. Exploring Aspects in the Application of Long Accelerating Moving Walkways*, T2009/10, October 2009, TRAIL Thesis Series, The Netherlands

Gong, Y., *Stochastic Modelling and Analysis of Warehouse Operations*, T2009/9, September 2009, TRAIL Thesis Series, The Netherlands

Eddia, S., *Transport Policy Implementation and Outcomes: the Case of Yaounde in the 1990s*, T2009/8, September 2009, TRAIL Thesis Series, The Netherlands

Platz, T.E., *The Efficient Integration of Inland Shipping into Continental Intermodal Transport Chains. Measures and Decisive Factors*, T2009/7, August 2009, TRAIL Thesis Series, The Netherlands

Tahmasseby, S., *Reliability in Urban Public Transport Network. Assessment and Design*, T2009/6, June 2009, TRAIL Thesis Series, The Netherlands

Bogers, E.A.I., *Traffic Information and Learning in Day-to-Day Route Choice*, T2009/5, June 2009, TRAIL Thesis Series, The Netherlands

Amelsfort, D. van, *Behavioural Responses and Network Effects of Time-varying Road Pric-*

ing, T2009/4, May 2009, TRAIL Thesis Series, The Netherlands

Li, H., *Reliability-based Dynamic Network Design with Stochastic Networks*, T2009/3, May 2009, TRAIL Thesis Series, The Netherlands

Stankova, K., *On Stackelberg and Inverse Stackelberg Games and their Applications in the Optimal Toll Design Problem, the Energy Markets Liberalization Problem, and in the Theory of Incentives*, T2009/2, February 2009, TRAIL Thesis Series, The Netherlands

Li, T., *Informedness and Customer-Centric Revenue*, T2009/1, January 2009, TRAIL Thesis Series, The Netherlands

Agusdinata, D.B., *Exploratory Modeling and Analysis. A promising method to deal with deep uncertainty*, T2008/17, December 2008, TRAIL Thesis Series, The Netherlands

Kreutzberger, E., *The Innovation of Intermodal Rail Freight Bundling Networks in Europe . Concepts, Developments, Performances*, T2008/16, December 2008, TRAIL Thesis Series, The Netherlands

Taale, H., *Integrated Anticipatory Control of Road Networks. A game theoretical approach* , T2008/15, December 2008, TRAIL Thesis Series, The Netherlands

Li, M., *Robustness Analysis for Road Networks. A framework with combined DTA models*, T2008/14, December 2008, TRAIL Thesis Series, The Netherlands

Yu, M., *Enhancing Warehouse Performance by Efficient Order Picking*, T2008/13, October 2008, TRAIL Thesis Series, The Netherlands

Liu, H., *Travel Time Prediction for Urban Networks*, T2008/12, October 2008, TRAIL Thesis Series, The Netherlands

Kaa, E.J. van de, *Extended Prospect Theory. Findings in Choice Behaviour from Economics and the Behavioural Sciences and their Relevance for Travel Behaviour*, T2008/11, October 2008, TRAIL Thesis Series, The Netherlands

Nijland, H.A., *Theory and Practice of the Assessment and Valuation of Noise from Roads and Rialroads in Europe*, T2008/10, September 2008, TRAIL Thesis Series, The Netherlands

Annema, J.A., *The Practice of Forward-Looking Transport Policy Assessment Studies*, T2008/9, September 2008, TRAIL Thesis Series, The Netherlands

Ossen, S., *Longitudinal Driving Behavior: Theory and Empirics*, T2008/8, September 2008, TRAIL Thesis Series, The Netherlands

Tu, H., *Monitoring Travel Time Reliability on Freeways*, T2008/7, April 2008, TRAIL Thesis Series, The Netherlands

D'Ariano, A., *Improving Real-Time Train Dispatching: Models, Algorithms and Applications*, T2008/6, April 2008, TRAIL Thesis Series, The Netherlands

Quak, H., *Sustainability of Urban Freight Transport*, T2008/5, February 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. 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 Transportation 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

Feijter, R. de, *Controlling High Speed Automated Transport Network Operations*, T2006/9,

December 2006, TRAIL Thesis Series, The Netherlands

Huisken, G., *Inter-Urban Short-Term Traffic Congestion Prediction*, T2006/8, December 2006, TRAIL Thesis Series, The Netherlands

Viti, F., *The Dynamics and the Uncertainty of Delays at Signals*, T2006/7, November 2006, TRAIL Thesis Series, The Netherlands

Yuan, J., *Stochastic Modelling of Train Delays and Delay Propagation in Stations*, T2006/6, October 2006, TRAIL Thesis Series, The Netherlands

Ongkittikul, S., *Innovation and Regulatory Reform in Public Transport*, T2006/5, September 2006, TRAIL Thesis Series, The Netherlands

Vanderschuren, M.J.W.A., *Intelligent Transport Systems for South Africa. Impact assessment through microscopic simulation in the South African context*, T2006/4, August 2006, TRAIL Thesis Series, The Netherlands

Ngoduy, D., *Macroscopic Discontinuity Modeling for Multiclass Multilane Traffic Flow Operations*, T2006/3, April 2006, TRAIL Thesis Series, The Netherlands

Lentink, R., *Algorithmic Decision Support for Shunt Planning*, T2006/2, February 2006, TRAIL Thesis Series, The Netherlands

Bontekoning, Y.M., *Hub exchange operations in intermodal hub-and-spoke networks. A performance comparison of four types of rail-rail exchange facilities*,T2006/1, February 2006, TRAIL Thesis Series, The Netherlands

Krogt, R.P.J. van der, *Plan Repair in Single-Agent and Multi-Agent Systems*, T2005/17, December 2005, TRAIL Thesis Series, The Netherlands

Valk, J.M., *Coordination among Autonomous Planners*, T2005/16, December 2005, TRAIL Thesis Series, The Netherlands

Groothedde, B., *Collaborative Logistics and Transportation Networks, a modeling approach to network design*, T2005/15, November 2005, TRAIL Thesis Series, The Netherlands

Pielage, B.A., *Conceptual Design of Automated Freight Transport Systems*, T2005/14, November 2005, TRAIL Thesis Series, The Netherlands

Boer, C., *Distributed Simulation in Industry*, T2005/12, October 2005, TRAIL Thesis Series, The Netherlands

Kager, R.M., *Design and implementation of a method for the synthesis of travel diary data*, T2005/11, October 2005, TRAIL Thesis Series, The Netherlands

Goverde, R., *Punctuality of Railway Operations and Timetable Stability Analysis*, T2005/10, October 2005, TRAIL Thesis Series, The Netherlands

Le-Duc, T., *Design and control of efficient order picking*, T2005/9, September 2005, TRAIL Thesis Series, The Netherlands

Oosten, W., *Ruimte voor een democratische rechtsstaat. Geschakelde sturing bij ruimtelijke investeringen*, T2005/8, September 2005, TRAIL Thesis Series, Sociotext, The Netherlands

Vromans, M., *Reliability of Railway Systems*, T2005/7, July 2005, TRAIL Thesis Series, The Netherlands

Koolstra, K., *Transport Infrastructure Slot Allocation*, T2005/6, June 2005, TRAIL Thesis Series, The Netherlands

Dekker, S., *Port Investment - Towards an integrated planning of port capacity*, T2005/5, June 2005, TRAIL Thesis Series, The Netherlands

Hoogendoorn-Lanser, S., *Modelling Travel Behaviour in Multimodal Networks*, T2005/4, May 2005, TRAIL Thesis Series, The Netherlands

Zuidgeest, M.H.P., *Sustainable Urban Transport Development: a Dynamic Optimization Approach*, T2005/3, April 2005, TRAIL Thesis Series, The Netherlands

Le-Anh, T., *Intelligent Control of Vehicle-Based Internal Transport Systems*, T2005/2, April 2005, TRAIL Thesis Series, The Netherlands

Rooij, R.M., *The Mobile City. The planning and design of the Network City from a mobility point of view*, T2005/1, February 2005, TRAIL Thesis Series, The Netherlands

Tampère, C.M.J., *Human-Kinetic Multiclass Traffic Flow Theory and Modelling: With Application to Advanced Driver Assistance Systems in Congestion*, T2004/11, December 2004, TRAIL Thesis Series, The Netherlands

Wees, K.A.P.C. van, *Intelligente voertuigen, veiligheidsregulering en aansprakelijkheid. Een onderzoek naar juridische aspecten van Advanced Driver Assistance Systems in het wegverkeer*, T2004/10, December 2004, TRAIL Thesis Series, The Netherlands

Versteegt, C., *Holonic Control For Large Scale Automated Logistic Systems*, T2004/9, December 2004, TRAIL Thesis Series, The Netherlands

Bos, D.M., *Changing Seats: A Behavioural Analysis of P&R Use*, T2004/8, November 2004, TRAIL Thesis Series, The Netherlands

Zoeteman, A., *Railway Design and Maintenance from a Life-Cycle Cost Perspective: A Decision-Support Approach*, T2004/7, November 2004, TRAIL Thesis Series, The Netherlands

Daamen, W., *Modelling Passenger Flows in Public Transport Facilities*, T2004/6, September 2004, TRAIL Thesis Series, The Netherlands

Verduijn, T. M., *Dynamism in Supply Networks: Actor switching in a turbulent business environment*, T2004/5, September 2004, TRAIL Thesis Series, The Netherlands

Tabibi, M., *Design and Control of Automated Truck Traffic at Motorway Ramps*, T2004/4, July 2004, TRAIL Thesis Series, The Netherlands

Lint, van, J.W.C., *Reliable Travel Time Prediction for Freeways*, T2004/3, June 2004, TRAIL Thesis Series, The Netherlands

Hegyi, A., *Model Predictive Control for Integrating Traffic Control Measures*, T2004/2, February 2004, TRAIL Thesis Series, The Netherlands

Langen, P.W, de, *The Performance of Seaport Clusters*, T2004/1, January 2004, TRAIL Thesis Series, The Netherlands

Weerdt, M.M., de, *Plan Merging in Multi-Agent Systems*, T2003/8, December 2003, TRAIL

Thesis Series, The Netherlands

Lindveld, Ch..D.R., *Dynamic O-D matrix estimation: a behavioural approach*, T2003/7, September 2003, TRAIL Thesis Series, Eburon, The Netherlands

Soto Y Koelemeijer, G., *On the behaviour of classes of min-max-plus systems*, T2003/6, September 2003, TRAIL Thesis Series, The Netherlands

Peeters, L.W.P., *Cyclic Railway Timetable Optimization*, T2003/5, June 2003, TRAIL Thesis Series, The Netherlands

Reeven, P.A., van, *Competition in Scheduled Transport*, T2003/4, April 2003, TRAIL Thesis Series, Eburon, The Netherlands

Riet, O.A.W.T., van de, *Policy Analysis in Multi-Actor Policy Settings. Navigating Between Negotiated Nonsense and Useless Knowledge*, T2003/3, March 2003, TRAIL Thesis Series, Eburon, The Netherlands

Lindeijer, D.G., *Controlling Automated Traffic Agents*, T2003/2, February 2003, TRAIL Thesis Series, Eburon, The Netherlands

Spek, S.C., van der, *Connectors. The Way beyond Transferring*, T2003/1, February 2003, TRAIL Thesis Series, Delft University Press, The Netherlands

Runhaar, H., *Freight transport: at any price? Effects of transport costs on book and newspaper supply chains in the Netherlands*, T2002/7, December 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Pol, P.M.J., *A Renaissance of Stations, Railways and Cities, Economic Effects, Development Strategies and Organisational Issues of European High-Speed-Train Stations*, T2002/6, October 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Van Nes, R., *Design of multimodal transport networks, a hierarchical approach,* T2002/5, September 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Veeneman, W.W., *Mind the Gap; Bridging Theories and Practice for the Organisation of Metropolitan Public Transport*, T2002/4, June 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Koppius, O.R., *Information Architecture and Electronic Market Performance,* T2002/3, May 2002, TRAIL Thesis Series, The Netherlands

Vis, F.A., *Planning and Control Concepts for Material Handling Systems*, T2002/2, May 2002, TRAIL Thesis Series, The Netherlands

Wolters, M.J.J., *The Business of Modularity and the Modularity of Business*, T2002/1, February 2002, TRAIL Thesis Series, The Netherlands

Bockstael-Blok, W., *Chains and Networks in Multimodal Passenger Transport. Exploring a design approach*, T2001/8, December 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Schaafsma, A., *Dynamisch Railverkeersmanagement, besturingsconcept voor railverkeer op basis van het Lagenmodel Verkeer en Vervoer*, T2001/7, October 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Rosmuller, N., *Safety analysis of Transport Corridors,* T2001/6, June 2001, TRAIL Thesis

Series, Delft University Press, The Netherlands

Binsbergen, A.J., van, J.G.S.N. Visser, *Innovation Steps towards Efficient Goods Distribution Systems for Urban Areas*, T2001/5, May 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Willems, J.K.C.A.S., *Bundeling van infrastructuur, theoretische en praktische waarde van een ruimtelijk inrichtingsconcept*, T2001/4, June 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Roodbergen, K.J., *Layout and Routing Methods for Warehouses*, T2001/3, May 2001, TRAIL Thesis Series, The Netherlands

Muilerman, G.J., *Time-based logistics: An analysis of the relevance, causes and impacts*, T2001/2, April 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Bliemer, M.C.J., *Analytical Dynamic Traffic Assignment with Interacting User-Classes: Theoretical Advances and Applications using a Variational Inequality Approach*, T2001/1, January 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Meer, J.R., van, *Operational Control of Internal Transport*, T2000/5, September 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Subiono, *On Classes of Min-max-plus Systems and their Applications*, T2000/2, June 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Marchau, V.A.W.J., *Technology Assessment of Automated Vehicle Guidance - Prospects for Automated Driving Implementation*, T2000/1, January 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoedemaeker, M., *Driving with Intelligent Vehicles: Driving Behaviour with Adaptive Cruise Control and the Acceptance by Individual Drivers*, T99/6, November 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoogendoorn, S.P., *Multiclass Continuum Modelling of Multilane Traffic Flow*, T99/5, September 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Minderhoud, M.M., *Supported Driving: Impacts on Motorway Traffic Flow*, T99/4, July 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Veenstra, A.W., *Quantitative Analysis of Shipping Markets*, T99/3, April 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Westrenen, F.C., van, *The Maritime Pilot at Work*: E*valuation and Use of a Time-to-boundary Model of Mental Workload in Human-machine Systems*, T99/2, May 1999, TRAIL Thesis Series, Eburon, The Netherlands

Nat, C.G.J.M., van der, *A Knowledge-based Concept Exploration Model for Submarine Design*, T99/1, March 1999, TRAIL Thesis Series, Delft University Press, The Netherlands