Model Predictive Control for Integrating Traffic Control Measures

András Hegyi

Model Predictive Control for Integrating Traffic Control Measures

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. J.T. Fokkema,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op dinsdag 3 februari 2004 om 13:00 uur
door András HEGYI
elektrotechnisch ingenieur
geboren te Leeuwarden.

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. J. Hellendoorn

Toegevoegd promotor: Dr. ir. B. De Schutter

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. J. Hellendoorn	Technische Universiteit Delft, promotor
Dr. ir. B. De Schutter	Technische Universiteit Delft, toegevoegd promotor
Prof. dr. ir. R. Boel	Universiteit Gent
Prof. ir. B. Immers	Katholieke Universiteit Leuven
Ir. F. Middelham	Ministerie van Verkeer en Waterstaat
Prof. Dr.-Ing. M. Papageorgiou	Technical University of Crete
Prof. dr. H. J. van Zuylen	Technische Universiteit Delft
Prof. dr. R. Babuška, M.Sc.	Technische Universiteit Delft, reservelid

TRAIL Thesis Series T2004/2, The Netherlands TRAIL Research School

Published and distributed by: András Hegyi
E-mail: a.hegyi@dcsc.tudelft.nl

ISBN 90-5584-053-X

Keywords: integrated traffic control, model predictive control

Printed in The Netherlands

Acknowledgments

The work reported in this thesis was supervised by Prof. Hans Hellendoorn and Dr. Bart De Schutter, at the Delft Center for Systems and Control, Delft University of Technology. I was always impressed by the speed, accuracy, and quality of the comments that Bart gave on my work. I am also grateful to both Hans and Bart for their suggestions on several versions of the manuscript of this thesis. I always felt their full support in my work and enjoyed much the freedom that they gave me. Their supervising style included paying attention to both the professional and the personal side of the process, which I appreciate very much. I could not have wished better supervisors! I hope our cooperation will continue for many years.

The financial support of AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management, and of the *Mobility of People and Transportation of Goods* spearhead program of the Delft University of Technology is gratefully acknowledged.

I also would like to thank the other members of the Ph.D. committee: Prof. R. Babuška, Prof. R. Boel, Prof. B. Immers, Ir. F. Middelham, Prof. M. Papageorgiou en Prof. H. J. van Zuylen. I would like to thank them for their interesting and useful comments on the manuscript of this thesis. In addition, my thanks to Dr. Serge Hoogendoorn and Dr. Steven Logghe who also have commented on my manuscript, and with whom we had several interesting discussions during the last four years.

I am grateful to all the students whom I had the chance to supervise[1] during their final work: Pascual Breton, Monique van den Berg, and Abdessadek Karimi. Their work provided a substantial contribution to this thesis.

My special thanks to Ir. Ronald van Katwijk, who asked me first about the conditions under which traffic control measures can improve the performance. His question was the trigger I needed to start thinking about Chapter 5.

I am also grateful to Suk-Han, for her support, care, and patience throughout the entire period of my Ph.D. work.

[1] Most of the time jointly with Hans Hellendoorn and/or Bart De Schutter.

Glossary

List of Symbols

Symbols related to METANET

k, k_f	freeway time step counter
k_c	controller time step counter
m, μ	link index
i	segment index
T_f	time step size of the freeway simulation (in hours; a typical value is about 10/3600 h = 10 s)
$\rho_{m,i}(k_\mathrm{f})$	density of segment i of freeway link m at time step k_f (veh/km/lane)
$v_{m,i}(k_\mathrm{f})$	speed of segment i of freeway link m at time step k_f (km/h)
$q_{m,i}(k_\mathrm{f})$	flow leaving segment i of freeway link m at time step k_f (veh/h)
N_m	number of segments in freeway link m
λ_m	number of lanes in freeway link m
L_m	length of the segments in link m (km)
τ	time constant of the METANET speed relaxation term (h)
κ	METANET speed anticipation term parameter (veh/km/lane)
η	METANET speed anticipation term parameter (km^2/h)
a_m	parameter of the fundamental diagram
$\rho_{\mathrm{crit},m}$	critical density of link m (veh/km/lane)
$V(\rho_{m,i}(k_\mathrm{f}))$	speed of segment i of link m on a homogeneous freeway as a function of the density $\rho_{m,i}(k_\mathrm{f})$ (km/h)
ρ_max	maximum density (veh/km/lane)
$v_{\mathrm{free},m}$	free-flow speed of link m (km/h)
o	origin (on-ramp or main-stream) index
$wo(k_\mathrm{f})$	length of the queue on on-ramp o at time step k_f (veh)
$q_o(k_\mathrm{f})$	flow that enters the freeway at time step k_f (veh/h)
$d_o(k_\mathrm{f})$	traffic demand at origin o at time step k_f (veh/h).
$r_o(k_\mathrm{f})$	ramp metering rate of on-ramp o at time step k_f

C_o	capacity of on-ramp o (veh/h)
δ	METANET parameter for the speed drop term caused by merging at an on-ramp
ϕ	METANET parameter for the speed drop term caused by weaving at a lane drop
n	node index
Q_n	total flow that enters freeway node n (veh/h)
I_n	set of link indexes that enter node n
O_n	set of link indexes that leave node n
$\beta_{n,m}(k_\mathrm{f})$	fraction of the traffic that leaves node n via link m
$\gamma_{m,i,j}(k_\mathrm{f})$	fraction of traffic in segment i of link m that has destination j at time step k_f
$\rho_{m,i,j}(k_\mathrm{f})$	partial density of traffic in segment i of link m that has destination j at time step k_f (veh/km/lane)
$w_{o,j}(k_\mathrm{f})$	partial queue at on-ramp o with destination j (veh)
$\gamma_{o,j}(k_\mathrm{f})$	fraction of traffic at on-ramp o that has destination j at time step k_f
$Q_{n,j}$	total flow that enters freeway node n with destination j (veh/h)
$\beta_{n,m,j}(k_\mathrm{f})$	fraction of the traffic with destination j that leaves node n via link m
$v_{\mathrm{control},m,i}(k_\mathrm{f})$	speed limit applied in segment i of link m (km/h)
α	parameter expressing the non-compliance of drivers with the displayed speed limits
$r_{msm}(k_\mathrm{c})$	main-stream metering rate at time step k_c
$q_{\mathrm{cap},m}$	capacity of link m
η_high	anticipation constant for a downstream density that is higher that the density in the actual segment (km^2/h)
η_low	anticipation constant for a downstream density that is lower that the density in the actual segment (km^2/h)
$\rho_d(k_\mathrm{f})$	downstream density scenario at destination d (veh/km/lane)
$q_\mathrm{r,min}$	minimum on-ramp flow (veh/h)
$J(k_\mathrm{c})$	objective function to be optimized
ξ_i	weights for the partial objective functions
N_p	prediction horizon length
N_c	control horizon length
T	time step size of the prediction model (h)
T_c	time step size of the MPC controller (h)
M	constant integer, equals T_c/T

Symbols related to MPC

k	discrete time index for the process model
k_c	discrete time index for the controller
$x(k)$	process (model) state
$\hat{\mathbf{x}}(k)$	$[\hat{x}(k+1\|k)\ \ldots\ \hat{x}(k+MN_\mathrm{p}-1\|k)]$, the predicted states for the simulation steps $\{k,\ldots,k+MN_\mathrm{p}-1\}$ based on knowledge at simulation step k
$d(k)$	disturbance vector at simulation time step k
$\mathbf{d}(k)$	$[d(k)\,d(k+1)\ \ldots\ d(k+MN_\mathrm{p}-1)]$, the disturbance signals for the simulation steps $\{k,\ldots,k+MN_\mathrm{p}-1\}$
$u(k)$	control vector
$\mathbf{u}(k_\mathrm{c})$	$[u(k_\mathrm{c}\|k_\mathrm{c})\,u(k_\mathrm{c}+1\|k_\mathrm{c})\ \ldots\ u(k_\mathrm{c}+N_\mathrm{p}-1\|k_\mathrm{c})]$, the control signal for the controller time steps $\{k_\mathrm{c},\ldots,k_\mathrm{c}+N_\mathrm{p}-1\}$ based on the knowledge at controller step k_c
$\mathbf{u}^*(k_\mathrm{c})$	$[u^*(k_\mathrm{c}\|k_\mathrm{c})\,u^*(k_\mathrm{c}+1\|k_\mathrm{c})\ \ldots\ u^*(k_\mathrm{c}+N_\mathrm{c}-1\|k_\mathrm{c})]$, the control signal that minimizes $J(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$ based on knowledge at controller step k_c
$J(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$	objective function
N_p	prediction horizon length
N_c	control horizon length
$f(x(k),u(k_\mathrm{c}))$	process (model) state update function
$g(x(k),u(k_\mathrm{c}))$	measurement function
$\phi(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$	equality constraint function
$\psi(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$	inequality constraint function
$\hat{\mathbf{y}}(k)$	$[\hat{y}(k+1\|k)\ \ldots\ \hat{y}(k+MN_\mathrm{p}-1\|k)]$, the predicted outputs for simulation time steps $\{k,\ldots,k+MN_\mathrm{p}-1\}$ based on knowledge at simulation time step k

Symbols related to the urban traffic model

s, n, u	intersection indexes (node)
T_u	time step used for the urban simulation (h)
k_u	urban time step counter
U_s	set of origins of intersection s
d	link index (when it is a destination)
O_s	set of leaving links of node (intersection) s
$x_{u,s,d}(k_\mathrm{u})$	queue length at time $t=k_\mathrm{u}T_\mathrm{u}$ (veh) at intersection s, for traffic that goes from origin u to link d

$l_{s,n}$	link connecting intersections s and n
$\beta_{u,s,d}(k_{\rm u})$	fraction of the traffic arriving from origin u at intersection s that wants to go to link d in the time interval $[k_{\rm u}T_{\rm u},(k_{\rm u}+1)T_{\rm u})$
$L_{s,n}$	length of link $l_{s,n}$ (veh)
$L_{{\rm km},s,n}$	length of link $l_{s,n}$ (km)
$L_{\rm vehicle}$	average length of the vehicles (km)
$S_{s,n}(k_{\rm u})$	available free space of link $l_{s,n}$ at time $t = k_{\rm u}T_{\rm u}$ (veh) (i.e., the buffer capacity $L_{s,n}$ minus the number of vehicles that are already present at time $t = k_{\rm u}T_{\rm u}$)
$m_{{\rm arr},u,s}(k_{\rm u})$	number of vehicles arriving at the tail of the queue in link $l_{u,s}$ during the time interval $[k_{\rm u}T_{\rm u},(k_{\rm u}+1)T_{\rm u})$
$m_{{\rm arr},u,s,d}(k_{\rm u})$	number of vehicles arriving at the tail of the queue with link d in link $l_{u,s}$ during the time interval $[k_{\rm u}T_{\rm u},(k_{\rm u}+1)T_{\rm u})$
$m_{{\rm dep},u,s,d}(k_{\rm u})$	number of vehicles departing from link $l_{u,s}$ toward link d in $[k_{\rm u}T_{\rm u},(k_{\rm u}+1)T_{\rm u})$
$m_{{\rm dep},s,d}(k_{\rm u})$	number of vehicles departing from intersection s towards link $l_{s,d}$ in $[k_{\rm u}T_{\rm u},(k_{\rm u}+1)T_{\rm u})$
$g_{u,s,d}(k_{\rm u})$	indicates whether the traffic sign at intersection s for the traffic going from u to d is green[2](1) or red (0) during $[k_{\rm u}T_{\rm u},(k_{\rm u}+1)T_{\rm u})$
$C_{u,s,d}(k_{\rm u})$	capacity of intersection s for traffic arriving from u and turning to d at time $t = k_{\rm u}T_{\rm u}$ (veh/h)
$v_{s,n}$	free-flow speed[3]for the urban traffic between the entrance of the link $l_{s,n}$ and the tail of the queue at intersection n (km/h)
$\delta_{s,n}(k_{\rm u})$	time required to reach the tail of the queue waiting in link $l_{s,n}$ at time $t = k_{\rm u}T_{\rm u}$ (units of urban time steps)
$w_{o,m}(k_{\rm u})$	queue length on on-ramp o (veh) coming from intersection s waiting to depart toward freeway link m at time $t = k_{\rm u}T_{\rm u}$.
$x_{u,s,d}(k_{\rm u})$	queue length link $l_{u,s}$ (veh) waiting to depart toward link d at time $t = k_{\rm u}T_{\rm u}$.
$\lambda n, s$	the number of lanes in urban link $l_{n,s}$

Acronyms and Abbreviations

DRIP	Dynamic Route Information Panel
MPC	Model Predictive Control
TTS	Total Time Spent
VMS	Variable Message Sign

Contents

Chapter 1

Introduction

1.1 Traffic problems

In this section we give a characterization of freeway traffic problems and briefly discuss the motivation for the traffic control problem statement in Section 1.2. A full policy analysis is out of the scope of this thesis but we refer the interested reader to [176, 178, 177] for more information on this topic.

Since the main focus of this thesis is on freeway traffic systems we will often refer to freeway traffic situations as examples for the argumentation in this section. However most of the arguments in this section are also applicable to urban traffic systems. Moreover, we will discuss the joint control of urban-freeway networks in Chapter 8.

1.1.1 The need for dynamic traffic management

As the number of vehicles and the need for transportation grows, cities around the world face serious traffic congestion problems: almost every weekday morning and evening during rush hours the capacity of many main roads is exceeded. Traffic jams do not only cause considerable costs due to unproductive time losses; they also augment the probability of accidents and have a negative impact on the environment (air pollution, lost fuel) and on the quality of life (health problems, noise, stress).

One solution to the ever growing traffic congestion problem is to extend the road network. Adding lanes and creating alternative new freeway connections is possible but rather expensive. Dynamic traffic management is an alternative that aims to increase the safety and efficiency of the existing traffic networks.

1.1.2 The need for network-oriented traffic control

The fact that the length, duration and the number of traffic jams is increasing has certain consequences for dynamic traffic control. When there are more congested locations, the available control measures have to solve more problems, which implies a higher complexity. Since nowadays the chances are higher that a vehicle encounters more than one traffic jam on its route, the traffic control measures influencing a vehicle in one traffic jam will also influence the other jam(s) that it encounters. Therefore, the spatial interrelations between traffic situations at different locations in the network get stronger, and consequently the interrelations between the traffic control measures at different locations in the network also get stronger. These interrelations may differ per situation (and depend on, e.g., network topology, traffic demand, etc.) and the control measures may be cooperative or counteract each other. Coordinative control strategies are required in these cases, to make sure that all available control measures serve the same objective.

Another development is that freeways are equipped with more and more traffic control measures. The increasing number of control measures increases the controllability of the freeways, but the number of possible combinations of control measures is also increasing drastically, which in its turn increases the complexity of the dynamic traffic management problem.

On modern freeways a large amount of data is available on-line and off-line that can serve as a basis for choices of appropriate control measures. However, the available data is not fully utilized neither by traffic control center operators whose actions are typically based on heuristic reasoning, nor by automatic control measures that mostly use only local data. Traffic data also contains information about the traffic system as a network (origin-destination (OD) relationships, route choice), and information about the current disturbances of the network (incidents, weather influences, unexpected demands). Automatic control systems can handle large amounts of data and benefit from the network-oriented information by selecting appropriate control measures for given OD patterns and disturbances.

Network-oriented traffic control has two main ingredients: *coordination* and *prediction*. Since in a dense network the effect of a local control measure could also influence the traffic flows in more distant parts of the network the control measures should be coordinated such that they serve the same objectives. Determining the effects of control measures on distant parts of the network also involves prediction, since the effect of the control measure has a delay that is at least the travel time between the two control measures in the downstream direction, and the propagation time of shock waves in the upstream direction.

Network-oriented traffic control has several advantages compared to local control. E.g, solving a local traffic jam only, can have as consequence that the vehicles run faster into another (downstream) jam, whereas still the same amount of vehicles have to pass the downstream bottleneck (with a given capacity). In such a case, the average travel time on the network level will still be the same. A global approach would take into account and, if possible, solve both jams.

Furthermore, if dynamic origin-destination (OD) data is available, control on the net-

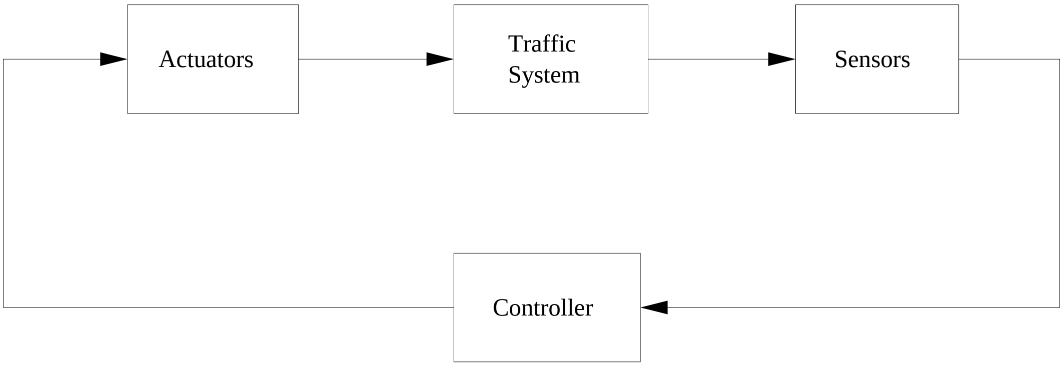


Figure 1.1: Schematic representation of the dynamic traffic management control loop. The controller determines the control signals sent to the actuators, based on the measurements provided by the sensors. Since the control loop is closed the deviations from the desired traffic system behavior are observed and appropriated control actions are taken.

work level can take advantage of the predictions of the flows in the network. Local controllers are not able to optimize the network performance even if the dynamic OD data is available, because the effect of the control actions on downstream area's is not taken into account. The flows in downstream area's may also be dependent on the actions of other local controllers. Since these controllers are not coordinated on the network level, actions may be taken that result in suboptimal performance of the downstream area's. E.g., on a freeway with several metered on-ramps (pro-active, coordinated) metering of the upstream ramps may be needed to prevent a jam at a downstream ramp caused by high ramp demands. Preventing such a jam can result in a better freeway performance. In other words, by anticipating on predictable future events a predictive control system can also *prevent* problems instead of only *reacting* to them.

Dynamic traffic management systems operated according to the control loop concept known from control systems theory (see Figure 1.1). The traffic sensors provide information about the current traffic state, such as speed, flow, density, or occupancy[1]. If the sensors do not provide all traffic states needed by the controller, data filtering or data estimation techniques may be used, such as Kalman filtering [174] or dynamic OD estimation [175]. The controller determines appropriate control signals that sent to the actuators. The reaction of the traffic system is measured by the sensors again, which closes the control loop. If new measurements show a deviation from the desired traffic system behavior

[1] The freeway traffic data monitoring systems in The Netherlands, Monica and Mare, provide speed and flow data. Also worth of mentioning here is the Regiolab [179] project where urban and freeway data is logged centrally.

(caused by unforeseen disturbances), the new control signals are adopted accordingly.

Another problem is when the parameters of the traffic system change, e.g., when an incident occurs, or the weather conditions significantly change the system behavior. In that case the parameters of the prediction model need to be adapted to the new situation. This is called *adaptivity*.

In this thesis we focus on the determination of the appropriate control signals and assume that all necessary traffic state variables are available to the controller, and that the process parameters are known and constant.

1.1.3 Objectives in traffic control

We will define what an ‘appropriate’ control signal is in terms of optimality. It is obvious that the formulation of optimality depends on the objectives. From network operator point of view typical objectives are:

- **Efficiency.** This objective is also shared by the individual drivers. However, situations may arise when minimizing, e.g., the total travel time in a network (network optimum) is different from minimizing individual travel times[2] (user optimum).

- **A sufficient level of safety.** In a certain sense the safety requirement is a boundary condition or constraint, because traffic control measure should never result in unsafe situations. However, there is also interaction between safety and efficiency, which consist of at least two processes. First, a safer traffic system results in less accidents and therefore more often in higher flows. Since a substantial part of the traffic jams is caused by accidents[3], this relationship is relevant. Second, less congestion (more efficiency due to control) increases safety. Third, lower speeds and densities positively influence safety. So, the objectives efficiency and safety may be non-conflicting or conflicting, depending on the case. In case they are conflicting the trade-off between safety and efficiency is a matter of policy.

- **Network reliability.** Even if not every traffic jam can be prevented, it is valuable for drivers when the travel time to their destinations is predictable. Predictable travel times and good arrival time estimations make departure time choices easier. Traffic control can aim at the realization of predicted travel times (or the reverse: predict realizable travel times, or both[4]). Furthermore, network reliability can be improved

[2] Note that this is a consequence of a non-cooperative multi-player game with a Nash equilibrium. See [9] for more information on game theory.

[3] In The Netherlands approximately 25 % of all traffic jams is caused by accidents.

[4] The prediction can (and should) take into account the control scenarios which influence the travel times in the considered route. In Chapter 7 we present an approach that integrates travel time prediction (in the form of route guidance) and ramp metering.

by synchronization of the traffic demand and the capacity supply of the network, and by the better distribution traffic flows over the network.

- **Low fuel consumptions, low air and noise pollution.** In urban areas the environmental effects of traffic may be considered more important than, e.g., efficiency, which can result in a different trade-off between the two objectives. An example of such a trade-off is between travel speed and air pollution [119, 5].

In this thesis only the first three objectives will be considered. Objectives that take into account fuel consumption, and air and noise pollution can be included in the controller in a similar way as the other objectives. Efficiency will be formulated as the total time spent (TTS) in the network by all vehicles. In addition, we will assume that the traffic demand is given[5]. Under this assumption, lower TTS means shorter travel times on the average. We will consider safety as a constraint for the speed limit control in Chapter 6, and formulate the minimization of the prediction error as a (sub)goal in Chapter 7.

1.1.4 Relation between outflow and the TTS: a reason for feedback

In this section we discuss the strong relation between TTS and the outflow of the network in congested situations (cf. [127]). It can be argued that because of this strong relationship a control method is desired that ‘has a great precision’. Even an improvement of the outflow (by control) of a few percents can significantly improve the TTS. We will argue that *feedback* is a structure that can improve the precision of the controller and is therefore desired for traffic control.

We explain the relationship of TTS and outflow by an example. Suppose a traffic network with an outflow that can be improved by 5 % due to traffic control. We compare two cases where the outflow of the network is 4000 veh/h (uncontrolled case) and 4200 veh/h (controlled case) respectively. Note that for the calculation of the TTS the network structure is irrelevant, the only variables that influence the TTS are the inflow and the outflow of the network. The demand at the entrances of the network is assumed to be fixed, but not constant: for a half an hour it exceeds the capacities of both cases, the controlled and the uncontrolled case, see Figure 1.2. In both cases the number of vehicles stored in the network (the ‘queues’ in Figure 1.3) is increasing, but there is a significant difference in the evolution of the queue length between the two cases. In the uncontrolled case the number of stored vehicles in the network increases faster, and decreases more slowly. The TTS is equal to the area below the queue length curves. In the uncontrolled case the TTS is 14 % higher than in the controlled case. Compare this to the 5 % difference in the outflow. The time that the queue is resolved is also significantly lower (half an hour) in the controlled case. The consequence of this relation between TTS and outflow

[5] This means that we assume that the traffic control measures will not affect mode choice or departure time

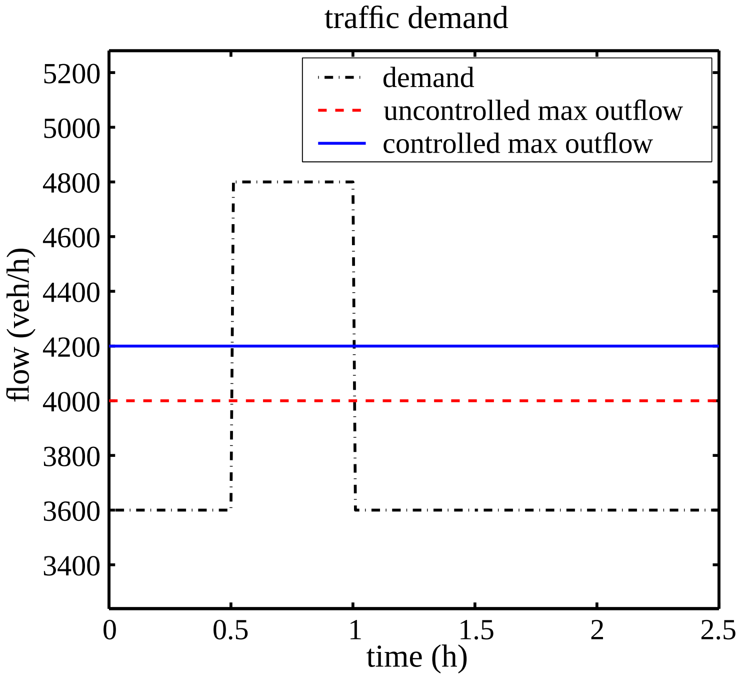


Figure 1.2: A simple illustration of the strong relationship between the total time spent and the outflow of a network. Uncontrolled and controlled networks are compared where the outflow of the controlled situation is 5% higher than the outflow of the uncontrolled situation. The demand exceeds both capacities for a half an hour.

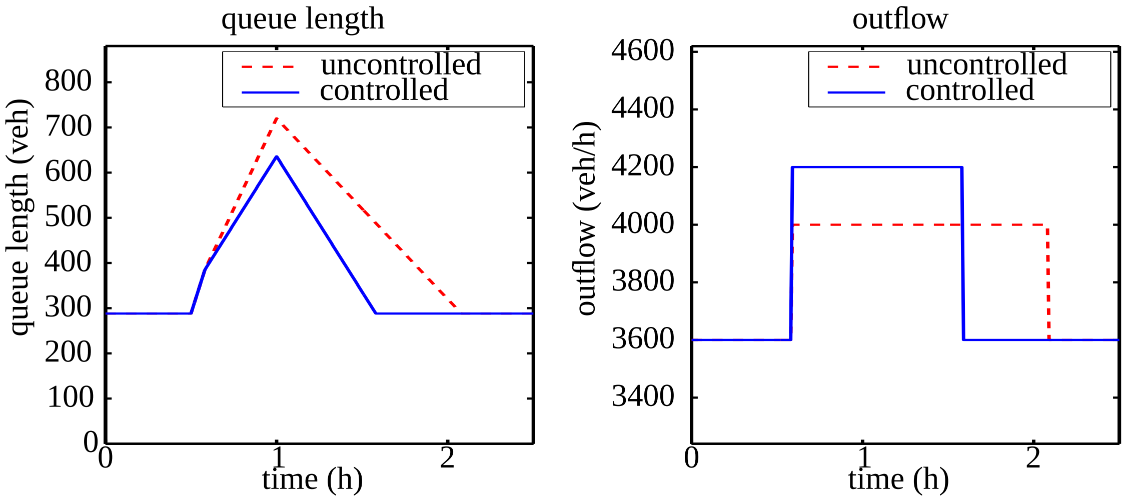


Figure 1.3: A simple illustration of the strong relationship between the total time spent and the outflow of a network. Uncontrolled and controlled networks are compared where the outflow of the controlled situation is 5% higher than the outflow of the uncontrolled situation. In the controlled case the total queue length increases more slowly and decreases faster, and the queue is resolved significantly faster. The difference in total time spent is 14 %.

is that traffic should be controlled with great precision. Any disturbance that reduces the outflow with a few percents, may significantly increase the TTS. In control engineering the effect of (unpredictable) disturbances is reduced by *feedback*. In control engineering the concept of feedback is important when there are unpredictable disturbances acting on the controlled process. Feedback is realized by regularly (or constantly) examining the state or the output of the system which gives information about the disturbances that are present. Given the disturbances an appropriate control signal is applied to the process.

1.2 Problem statement

Given the considerations above, the dynamic traffic control problem can be formulated as follows.

Dynamic traffic control problem

Given

– a network structure (possibly consisting of urban, freeway and secondary roads),

- the predictable disturbances: the traffic demand or the dynamic OD matrix in case of a network with multiple origins or destinations, incoming shock waves,
- the available traffic control measures,
- the constraints, such as minimum metering rates, forbidden speed limit combinations, etc.,
- a user definable control objective (which may consist of several sub-objectives),

find the control signals (traffic control measures) that optimize the given objective.

Based on the nature of the problem the controller should have the following properties:

- it can handle multiple-input multiple-output systems,
- it is predictive,
- it can optimize control inputs according to an objective function,
- it can handle constraints,
- it has a feedback structure,
- it is adaptive to process parameter variations.

Traffic control measures may have effect on drivers' route choice. When a traffic control strategy structurally creates travel time differences (or in general: cost difference) between alternative routes, drivers may adapt their routes in order to minimize their travel times. In this thesis we do not take these effects into account. We refer the interested reader to Taale [152] and Bellemans [10].

1.2.1 Approach: Model predictive control

To solve the dynamic traffic control problem we apply a model predictive control (MPC) framework [18, 39, 108]. The MPC framework fulfills all criteria listed in the problem statement in Section 1.2.

Gartner [40] introduced the concept of MPC to the field of urban demand-responsive traffic control. Another publication worth mentioning here is [122], where Papageorgiou applies the same control framework to sewer networks. Because of the similarities between traffic networks and sewer networks the approach and the findings in [122] are also relevant for traffic networks. In [122] MPC is found to be a control approach that results in a good performance even if the future disturbances are only partially known.

Bellemans [10] also considers MPC for traffic control. However, Bellemans considers ramp metering only, whereas we also include speed limits and route guidance. Furthermore, we use the extended version of the macroscopic traffic flow model METANET [156, 93, 126, 91], and develop a unified urban-freeway control framework that is suitable for MPC.

MPC is an optimal control method applied in a rolling horizon[6] framework. Optimal control is successfully applied by Kotsialos and Papageorgiou [93, 94, 91] to coordinate or integrate traffic control measures. Also Hoogendoorn has examined optimal control for route guidance [74]. Both optimal control and MPC have the advantage that the controller generates control signals that are optimal according to a user-supplied objective function. However, MPC has some important advantages over the traditional optimal control.

- Optimal control has an *open-loop* structure, which means that the disturbances (in our case: the traffic demands) have to be completely and exactly known before the simulation, and the traffic model has to be very accurate to ensure sufficient precision for the whole simulation. MPC operates in *closed-loop* which means that the traffic state and the current demands are regularly fed back to the controller, and the controller can take disturbances (here: demand prediction errors) into account and correct for prediction errors resulting from model mismatch.

- Adaptivity is easily implemented in MPC, because the prediction model can be changed or replaced during operation[7]. This may be necessary when traffic behavior significantly changes (e.g., in case of incidents, changing weather conditions, lane closures for maintenance).

- For MPC a shorter prediction horizon is usually sufficient, which reduces complexity, and make the real-time application of MPC feasible.

An essential part of the MPC controller is the model that is used to predict the effects of the control signals. This model needs to satisfy certain criteria:

- If the control is to be operated in real-time, the model needs to be fast when executed on a computer.

- The model should reproduce the dynamic traffic process with sufficient accuracy.

- The model should reproduce certain specific phenomena that are relevant to the controlled situation. In Chapter 5 we specify these phenomena, such as shock waves that remain existing for a long time, the capacity drop at on-ramps and at shock waves, and blocking.

[6] rolling horizon: also called receding horizon.

[7] While adaptivity is a property of MPC, in this thesis we will not examine this property explicitly.

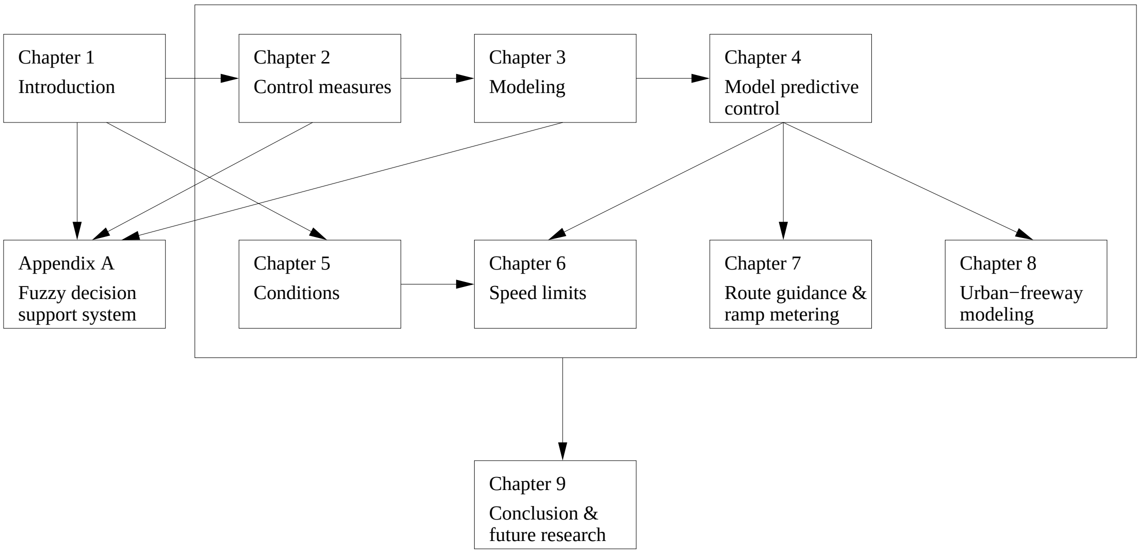


Figure 1.4: The relation between the chapters.

Although there may be other traffic models that satisfy these criteria, in this thesis we will use METANET as the prediction model in the controller. Since this model is deterministic, discrete-time discrete-space with relatively large time step and freeway segment length the execution of this model on a computer is very fast. Regarding the validation of the model we refer to [88, 38]. The capability to reproduce the relevant phenomena (shock waves, capacity drop at on-ramps and at shock, and blocking) is demonstrated in the experiments in in Chapter 6.

Note, however, that the MPC approach, which will be presented in Chapter 4 is generic so that we could also work with other traffic flow models.

1.3 Overview of the thesis

In this section an overview of the chapters in this thesis is given. The relations between the chapters is also illustrated in Figure 1.4.

As the main focus of this thesis is on freeway traffic control we describe the most frequently used freeway control measures (ramp metering, dynamic speed limits, route guidance, peak lanes and dedicated lanes, etc.) in more detail in Chapter 2. We present per control measure the control methods found in literature, field and simulation test results, and some practical considerations. Also some other control measures are described, that are less frequently used, but can potentially improve traffic flow.

In Chapter 3 we discuss the existing traffic flow models. The models are categorized according to several criteria: application area, level of detail, or process representation: deterministic versus stochastic, and continuous versus discrete. Next, we introduce the traffic flow model METANET. This model will be used throughout this thesis for the simulation of freeways and secondary roads. In Section 3.3 the METANET model is extended by the following items:

- We add an explicit model for dynamic speed limits.
- We add a model for main-stream metering.
- We add a model for main-stream origins which have different dynamics than on-ramps.
- We differentiate between the anticipation behavior at the head and the tail of shock waves.
- We add a formulation for the downstream boundary condition that can express scenarios where the downstream area is uncongested, except for some incoming shock waves.

These extensions and modifications will be used in the simulations in Chapters 6, 7, and 8.

In Chapter 4 we introduce the model predictive control (MPC) approach. After the mathematical description of MPC, the rules for tuning are discussed. Next, the advantages and the disadvantages (and possible solutions) of this method are presented. Furthermore, in Chapter 4 MPC is formulated in a traffic setting. It is shown that it is relatively easy to formulate the traffic control problem in an MPC framework: we discuss the formulation of some objective functions, boundary conditions, and the tuning of the controller for traffic systems.

It is unrealistic to expect that every traffic problem can be solved by traffic control. Therefore, it is important to describe the conditions under which we can expect improvement by applying certain control measures. In Chapter 5 we present necessary conditions for the effectiveness of ramp metering, and dynamic speed limits. These conditions are discussed with the assumption that the main goal of traffic control is to minimize TTS. The conditions include the specification of the traffic scenario, such as the network topology (locations of bottlenecks) and traffic demands. Since we use MPC, which includes an internal prediction model, we also pay attention to the phenomena that this model should be able to reproduce. A part of Chapter 5 has also been published in [61].

In Chapter 6 we demonstrate the MPC control framework with several traffic problems related to speed limits. We discuss the integrated control of ramp metering and the speed limits, where the speed limits can prevent a traffic breakdown when ramp metering only is insufficient. Since the main effect of the speed limits in this section is to limit the flow when necessary, this set-up is compared with a set-up where the speed limits are replaced by main-stream metering. A part of this work has also been published in [51, 52, 62].

We also consider in Chapter 6 another application of speed limits where the speed limits are used to reduce or eliminate shock waves on motorways. Parts of this work has also been published in [56, 55, 15, 57, 54, 58].

In Chapter 7 we apply the MPC approach to integrate ramp metering and dynamic route guidance. The main objective of the control is to minimize the TTS in the network by providing travel times shown on the dynamic route information panels (DRIPs) and by ramp metering. The second objective of the control is to keep the travel time predictions accurate. The addition of this goal is necessary, because there is a conflict between using DRIPs as an information source and using DRIPs as a control measure [90].

This conflict can be described as follows. Even if the display information is exactly what the driver will encounter on its route, the resulting route choice (splitting rates) will not be necessarily the optimal ones (from control point of view). As a consequence it may be necessary to display incorrect information for optimal route guidance, which is also undesirable, because the drivers' compliance depends on the correctness of the information. The material presented in Chapter 7 has also been published in [81].

Traffic problems frequently occur around the boundary between freeways and urban areas, e.g., when on-ramp queues block surface streets, or when off-ramp traffic cannot be accommodated by the urban network. In such situations both the urban and freeway networks can benefit from a coordinated control of urban and freeway control measures. In an MPC framework this means that a combined model is needed that enables us to predict the total effects of these measures. In Chapter 8 we develop such a combined urban-freeway traffic model. This work has also been published in [161, 160].

MPC is not the only possible approach for dynamic traffic control. As an alternative we present in Appendix A a prototype for a decision support tool for operators in traffic control centers. This tool aims at reducing the number of on-line real-time simulations that are necessary for a traffic operator to evaluate the alternative control scenarios. The decision support system uses case-based reasoning and fuzzy interpolation to evaluate the alternative control actions. A case base is made, based on off-line simulation, that contains typical combinations of traffic scenarios, control actions and performance measures. The system selects the cases from the case base, that are similar to the current traffic state, and predicts the performances of several combinations of control measures. The best control scenarios are shown to the operator who decides about the final choice. Parts of the material of Chapter 5 has also been published in [59, 60, 50].

1.4 Contributions to the state of the art

In this section we summarize the main contributions of this thesis:

- In Chapter 3 we extend the METANET model with the modeling of: dynamic speed limits, main-stream metering (as opposed to on-ramps), main-stream origin, differentiation between the anticipation behavior at the head and the tail of shock waves, a new formulation of the downstream boundary condition.

- In Chapter 4 we apply the MPC framework to traffic systems and present heuristic tuning rules for traffic control problems formulated in an MPC framework.

- In Chapter 5 we discuss the necessary conditions for successful traffic control in case of ramp metering, and dynamic speed limits.

- In Chapter 6 we examine several set-ups with speed limits and other control measures.

  Also in Chapter 6 we apply speed limits to suppress shock waves. The control concept is different from homogenization: it aims at resolving the high density region of the shock wave by flow limitation, and at restoring the dropped flow to the capacity flow. We also present a method to find discrete speed limit values, and introduce constraints that ensure the safe operation of speed limits.

- In Chapter 7 we introduce a new route guidance concept, that makes it possible to use DRIPs as a traffic control measure (instead of merely informing), while providing accurate travel time predictions. This concept is based on the fact that there is a conflict between informing drivers about travel times and controlling route choice of the drivers in order to maximize the network performance.

- In Chapter 8 we develop an urban traffic model and combine it with the freeway model METANET, such that the overall model is suitable for MPC. Special attention is paid to the development of the interface between the two models which operate at different sampling rates. We present a unified control framework for urban-freeway traffic control.

- In Appendix A we develop a prototype decision support tool for operators in traffic control centers, which is based on case-based reasoning and fuzzy interpolation.

Chapter 2

Traffic control measures

In this chapter we give an overview of control measures that are used or could be used to improve traffic flow. The list of control measures is not intended to include all possible traffic control measures, but we focus on control measures that are currently applied or could be applied in the near future, in particular ramp metering, speed limits and route guidance. For each control measure we present the control methods found in the literature, field and simulation test results.

We start in Section 2.1 with the review of ramp metering strategies and field and simulation studies about ramp metering. In Section 2.2 we discuss speed limit control systems. We distinguish between approaches that aim at homogenizing the traffic (which reduces the probability of a breakdown), and that limit the inflow to a traffic jam or shock wave (which can *resolve* an existing jam). In Section 2.3 we consider route guidance systems. These systems can serve to optimize network performance, or to help drivers to find the shortest route among the possible alternatives.

In Section 2.4 some other traffic control measures are listed that could also be used to improve the performance of traffic systems.

Next, we discuss the three main approaches to coordinated and integrated traffic control systems in Section 2.5: model-based optimal control, knowledge-based methods, and an approach with a relatively simple control law with parameters that are optimized for a large number of simulations such that the average behavior is optimal.

2.1 Ramp metering

Ramp metering (see Figure 2.1) is one of the most investigated and applied freeway traffic control measures. Ramp metering determines the flow rate at which vehicles can enter the freeway. The flow at the on-ramp is controlled by a traffic light and the flow rate is determined by selecting appropriate red, green and amber light timings. Ramp-metering can be used in two modes: the *traffic spreading mode* and the *traffic restricting mode*.

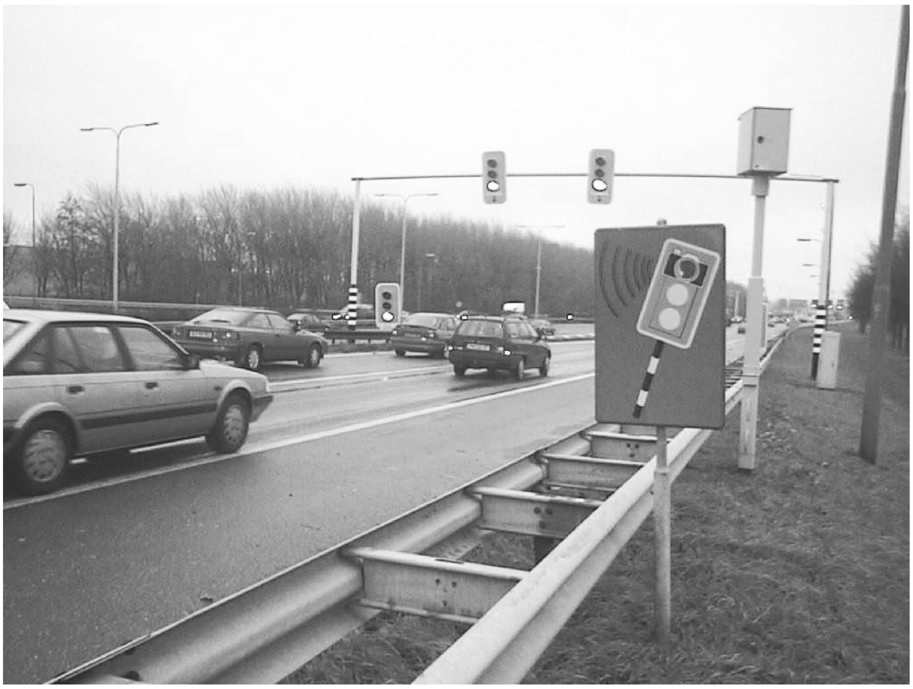


Figure 2.1: Ramp metering at the A13 at Delft in The Netherlands. One car may pass per green phase. To prevent red-light running the control is enforced.

In the traffic spreading mode the metering rate equals the average arrival rate of the vehicles at the on-ramp and its purpose is to spread the vehicles that enter the freeway. This is useful when, e.g., the traffic on the on-ramp arrives from a controlled intersection, because the vehicles arrive in platoons and could cause a serious disturbance when they enter the freeway simultaneously. By spreading the platoon the vehicles enter the freeway one-by-one and the probability of a disturbance that causes a traffic breakdown is reduced.

Restrictive ramp metering can be used for two different purposes.

- When traffic is dense, ramp metering can prevent a traffic breakdown on the freeway by adjusting the metering rate such that the density on the freeway remains below the critical value[1]. Preventing a traffic breakdown has not only the advantage of a higher flow downstream the on-ramp section (and thus shorter travel times), but also that it prevents the creation of a congestion that could block the off-ramp upstream the on-ramp (see Figure 2.2). These effects are studied in detail in [130].

- When drivers try to bypass congestion on a freeway by taking a local road (rat

[1] By the stochastic nature a traffic breakdown may occur even if the average density is below the critical density. To prevent such cases the controller could be tuned such that is aims at a density which is somewhat lower than the critical density. I such a way a ‘security’ margin is introduced. The choice of the magnitude of this margin represents a trade-off between efficiency and robustness.

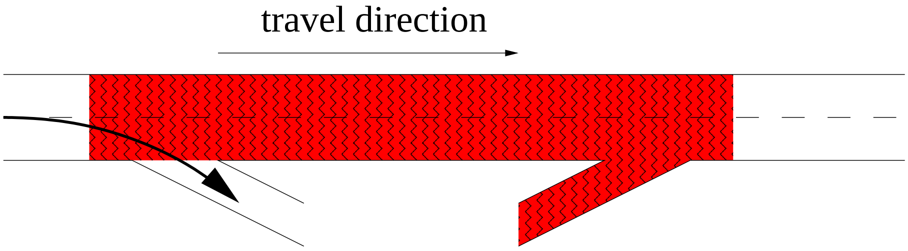


Figure 2.2: Congestion caused by excessive on-ramp demand blocks also the upstream off-ramp.

running), ramp metering can increase travel times and discourage the use of the bypass, see [110] for a synthetic study on the route-choice effects of ramp-metering.

2.1.1 Ramp metering strategies

Several ramp metering strategies have been developed for restrictive ramp metering and can be classified as static or dynamic, fixed-time or traffic-responsive, and local or coordinated.

Fixed-time strategies are determined off-line based on historical demands, and the demands and splitting rates at off-ramps are assumed to be constant in a given time slot, e.g., in the morning rush hour. This approach typically considers on-ramps and off-ramps along one freeway stretch, but is not difficult to extend to freeway networks. As control objective one may choose to maximize the number of served vehicles, to maximize the total traveled distance, or to balance ramp queues. These kind of ramp metering strategies result in linear-programming or quadratic-programming problems that can be solved by standard optimization methods. This approach was first suggested by Wattleworth [166], and is extended to a dynamic model by Papageorgiou [120]. The disadvantage of fixed-time strategies is that they do not take into account the traffic demand variations during a day or from day-to-day, which may result in underutilization of the freeway or inability to prevent congestion. Since traffic control requires precision as explained in Section 1.1.4, these disadvantages of fixed-time strategies may easily outweigh their advantages (their simplicity, and the fact that no traffic measurements are necessary).

Traffic-responsive strategies adjust on-line the metering rate as a function of the prevailing traffic conditions. These strategies typically aim at the same objectives as the fixed-time strategies, but also at preventing congestion. The traffic conditions are periodically fed into the controller to determine its control strategy. One of the best known strategies is the *demand-capacity* strategy:

$$q_{\text{ramp}}(k) = \begin{cases} q_{\text{cap}} - q_{\text{in}}(k-1) & \text{if } o_{\text{out}}(k) \leq o_{\text{cr}} \\ q_{\text{r,min}} & \text{otherwise} \end{cases}$$

where $q_{\text{ramp}}(k)$ is the admitted ramp flow, q_{cap} the freeway capacity, $q_{\text{in}}(k)$ the freeway flow measured upstream the on-ramp at sample step k, $o_{\text{out}}(k)$ the occupancy downstream the on-ramp at sample step k, o_{cr} is the critical occupancy (at which flow is maximal), and $q_{\text{r,min}}$ the flow according to the minimum metering rate. The critical *occupancy*[2] is needed to distinguish between free-flow and congested states, and the minimum metering rate is used to prevent completely blocked on-ramps. A similar strategy can be formulated based on upstream occupancy instead of upstream flow, where the upstream flow $q_{\text{in}}(o_{\text{upstream}}(k-1))$ is estimated based on a single upstream occupancy. However, both formulations have the disadvantage that they have an (open-loop) feed-forward structure, which is known to perform poorly under unknown disturbances.

A better approach is to use a (closed-loop) feedback structure, because it allows for controller formulations that can reject disturbances and have zero steady state error. ALINEA[3] [132] is the best known example of such a strategy and is formulated as

$$q_{\text{ramp}}(k) = q_{\text{ramp}}(k-1) + K[\hat{o} - o_{\text{out}}(k)]$$

where $K > 0$ is a control parameter and $\hat{o}$ a reference value for the occupancy downstream from the on-ramp.

The most advanced ramp metering strategies are the traffic-responsive coordinated strategies such as METALINE [126], FLOW [80], or methods that use optimal control [94] or model predictive control [10].

METALINE is a generalization and extension of ALINEA, that provides a control law for coordinated control of on-ramps:

$$\mathbf{q}_{\text{ramp}}(k) = \mathbf{q}_{\text{ramp}}(k-1) - \mathbf{K_1}[\mathbf{o}(k) - \mathbf{o}(k-1)] + \mathbf{K_2}\left[\hat{\mathbf{O}} - \mathbf{O}(k)\right]$$

where $\mathbf{q}_{\text{ramp}} = [q_{\text{r},1} \ldots q_{\text{r},m}]^T$ is the vector of the controlled ramp flows, and $\mathbf{o} = [o_1 \ldots o_n]^T$ the vector of the measured occupancies on the freeway, $\mathbf{O} = [O_1 \ldots O_m]^T$ is the sub-vector of $\mathbf{o}$ for which the reference values $\hat{\mathbf{O}} = [\hat{O}_1 \ldots \hat{O}_m]^T$ are specified, and the matrices $\mathbf{K_1}$ and $\mathbf{K_2}$ are the controller constants.

FLOW [80] is a heuristic ramp metering strategy where several traffic measurements

[2] *Occupancy is defined as the relative time (in percentages) that the induction loop (traffic sensor) is occupied by a vehicle. In practice this is often averaged over 1, 2 or 5 minutes.*

[3] ALINEA is the acronym for “Asservissement linéaire d’entrée autoroutière”, which could be translated as “Linear ramp metering control”.

are combined to determine the ramp metering rate. First, the local ramp metering rate is determined based on the occupancy level upstream of the metered ramp and a lookup table. Next, the metering rate based on the system capacity — called bottleneck metering rate — is determined based on the net inflow of a given freeway section downstream from the metered on-ramp, and the distance between the on-ramp and the given section. The bottleneck metering rate is only calculated when the net inflow of that section is positive and the occupancy at the downstream detector location of that freeway section exceeds a certain threshold. The final ramp metering rate — called system ramp metering rate — is the minimum of the two. In addition, when the ramp queue length exceeds a threshold w_1 the metering rate is increased, and when — despite the increased metering rate — the ramp queue exceeds a second threshold w_2, the metering rate is increased even more or the ramp metering is shut off.

The optimal control methods for the integration of several traffic control measures discussed in Section 2.5 can also be used to coordinate several ramp metering installation on several on-ramps.

2.1.2 Switching ramp metering on/off

An important aspect of ramp metering is that practical ramp metering algorithms also needs an (on/off) switching scheme. To the author's best knowledge the consequences of the choice of a certain switching scheme is not mentioned in any publication. There are several technical implementations of ramp metering to achieve a certain average desired ramp flow. In The Netherlands the typical implementation allows one car per green per lane (with a cycle length of few seconds). In other countries there exist implementations that allow two or more cars per green (with a longer cycle length of, e.g., 60 s).

The on/off switching scheme is especially important for one-per-green type of ramp metering, since the minimum red and amber times (typically respectively 2 s and 0.5 s) define the maximum flow achievable by ramp metering, which is around 3600 s.h^{-1} / 2.5 s.veh^{-1} = 1440 veh.h^{-1}, which is approximately 75 % of the capacity of a single lane[4]. In order to prevent unnecessary flow reduction ramp metering has to be switched off when the demand is so low that traffic is freely flowing, and has to be switched on when the demand is so high that the ramp flow has to be limited to less than 75 % of its capacity. The switching has to take place somewhere between the low and high demands. In practical systems often switching with hysteresis is used to prevent too frequent switching, but the effect of the thresholds on the performance is unknown. A possible way to circumvent the switching problem is to increase the number of lanes at the ramp metering device such

[4] The difference between the road capacity and the maximum flow when ramp metering is on is less articulated for other types of ramp metering because the relative red and amber times are smaller, but the capacity loss is still present.

that the on-ramp capacity can be reached even when ramp metering is switched on, but this is not feasible everywhere because of space limitations.

2.1.3 Field tests and simulation studies

Several field and simulation studies have shown the effectiveness of ramp metering. In Paris on the Boulevard Périphérique and in Amsterdam several ramp metering strategies have been tested [129, 128]. The demand-capacity, occupancy, and ALINEA strategies were applied in the field tests at a single ramp in Paris. It was found that ALINEA was clearly superior to the other two in all the performance measures (total time spent, total traveled distance, mean speed, mean congestion duration). Another comparison for a single on-ramp was performed in Amsterdam, where the Dutch RWS strategy (a variant of the demand-capacity strategy) was compared with ALINEA. Also in this case ALINEA proved to be superior, but the RWS strategy resulted in a more homogeneous traffic flow in the bottleneck. At the Boulevard Périphérique in Paris the multi-variable (coordinated) feedback strategy METALINE was also applied and was compared with the local feedback strategy ALINEA. Both strategies resulted in approximately the same performance improvement. In Amsterdam two local ramp metering strategies (RWS and ALINEA) were compared for metering four on-ramps simultaneously. Compared to the no-control case ALINEA achieved an improvement of the travel time losses, while the RWS strategy substantially increased the travel time losses.

Another field test was conducted in the Twin Cities metropolitan area of Minnesota [19]. In this area 430 ramp meters were shut down to evaluate their effectiveness. The results of comparing the situations with and without ramp metering can be summarized as follows.

- After the meters were turned off, there was an average traffic volume reduction on freeways of 9 %, and no significant volume change on parallel arterials.
- Without ramp metering the travel time increase was estimated at 25 121 hours of travel, which means that the decreased speeds on the freeways when metering is turned off outweigh the ramp delay when the metering is on.
- Without ramp metering travel time reliability was almost 50 % lower.
- The number of crashes in previously metered ramps and freeways increased by 26 %.
- Without ramp metering emissions were 1 160 tons/year higher.
- With ramp metering fuel consumption increases with 5.5 million gallons (20.8 million liters) on a yearly basis. This was the only criterion that was worsened by ramp metering.

- The benefit cost ratio indicated that the benefits are approximately 15 times greater than the cost of the ramp metering system.

A number of studies have simulated ramp metering for different transportation networks and traffic scenarios, with different control approaches, and with the use of microscopic and macroscopic traffic flow models [69, 94, 125, 123, 132, 49, 155]. Generally the total network travel time is considered as the performance measure and is improved by about 0.39 %–30 % when using ramp metering. Since the total time spent in the network is strongly dependent on the combination of the scenario (which determines the inflow or demand of the network) and on the control method (which determines the outflow of the network), these figures are encouraging but no guarantee for success in general.

For a further overview of field tests and simulation studies we refer to [49].

2.1.4 Main-stream metering

While ramp metering limits the flow at the entrances of the freeway, main-stream[5] metering limits the flow on the freeway itself. The technical implementation is similar: by choosing the relative green time in the red-green cycle the number of vehicles that may pass is controlled. Because of the similarity with on-ramp metering the same models are used for main-stream metering as for ramp metering. Simulation studies that include main-stream metering are presented in [94, 37].

2.2 Dynamic speed limits

Many modern freeways are equipped with variable speed limits signs (see Figure 2.3). Their main purpose currently is to increase safety by lowering the speed limits upstream of congested areas. However, attempts are also made to increase the traffic flow by more complex switching schemes [168, 139, 33].

2.2.1 Field tests and simulation studies

In the literature, basically two views on the use of dynamic speed limits can be found. The first emphasizes the homogenization effect (see [1, 2, 147, 149, 171, 97, 148, 72, 48]), whereas the second is more focused on preventing traffic breakdown by reducing the flow by means of speed limits (see [21, 100, 99]).

- The basic idea of homogenization is that speed limits can reduce the speed (and/or density) differences, by which a stabler (and safer) flow can be achieved. The homogenizing approach typically uses speed limits that are above the critical speed

[5] Main-stream metering is also called motorway-to-motorway control.

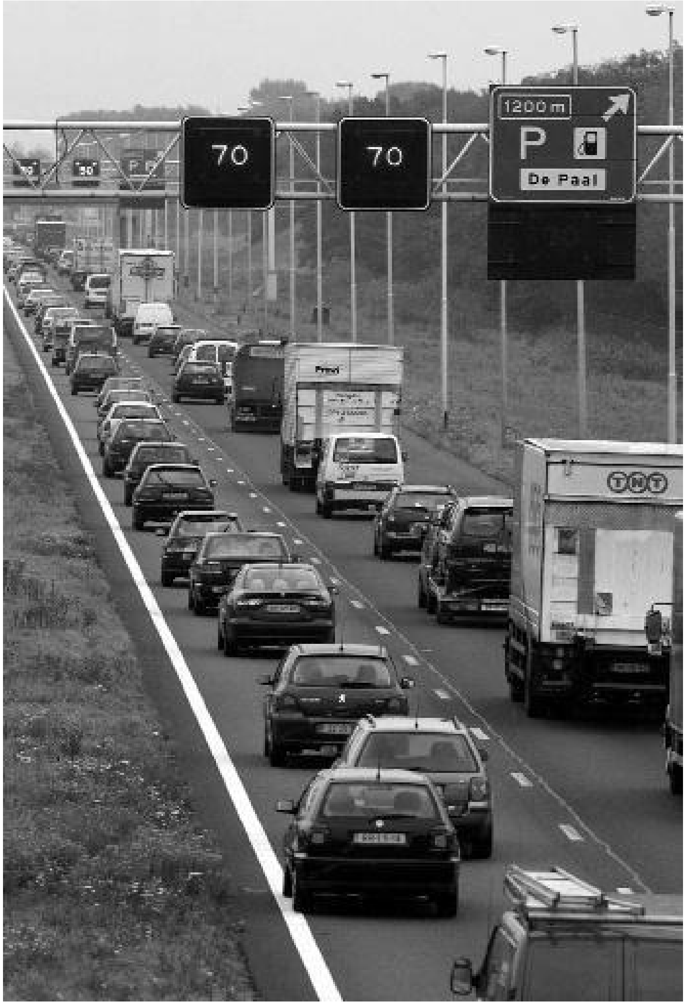


Figure 2.3: A variable speed limit gantry on the A1 freeway in The Netherlands.

(i.e., the speed that corresponds to the maximal flow; see Figure 2.4). So, these speed limits do not limit the traffic flow, but only slightly reduce the average speed (and slightly increase the density). In general, homogenization results in a more stable and safer traffic flow, but no significant improvement of traffic volume is expected nor measured [72]. In theory this approach can increase the time to breakdown [147], but it cannot suppress or resolve shock waves. An extended overview of speed limit systems that aim at reducing speed differentials is given by Wilkie [167]. It is interesting that while Wilkie recommends to place variable speed limit systems upstream of reduced-flow locations, in [72] it is concluded that speed control is not suitable to solve congestion at bottlenecks.

- The traffic breakdown prevention approach focuses more on preventing too high densities, and also allows speed limits that are lower than the critical speed in order to limit the inflow to these areas. By resolving the high density areas (bottlenecks) higher flow can be achieved in contrast to the homogenization approach.

Besides homogenization and the traffic breakdown prevention there may be other interesting applications of dynamic speed limits, such as dynamic speed limits at sharp curves to prevent sudden breaking and shock waves, or dynamic speed limits that harmonize the speeds of the incoming traffic streams at weaving and merging sections which may improve the traffic flow.

Several control methodologies are used in the literature to find a control law for speed control, such as multi-layer control [121, 103], sliding-mode control [100, 99], and optimal control [1, 2]. In [34] optimal control is approximated by a neural network in a rolling horizon framework. Other authors use (or simplify their control law to) a control logic where the switching between the speed limit values is based on traffic volume, speed or density [171, 97, 148, 72, 48, 99, 147]. In some cases the switching between the speed limit values is also based on special circumstances, such as weather and light conditions [171], or speed variance [97].

Several studies were made in the context of intelligent speed adaptation (ISA) in the field of advanced driver assistance systems (ADAS) (see [103, 22, 71]). For these systems a roadside controller sends the speed limit directly to an in-car device that executes the speed commands without driver intervention. By assuming no driver intervention, a wide range of speed and/or density profiles can be achieved, because it eliminates the drivers' reaction to the prevailing traffic conditions, such as relaxation and anticipation.

Some authors recognize the importance of anticipation in the speed control scheme. A pseudo-anticipative scheme is used in [99] by switching between speed limits based on the density of the neighboring downstream segment. Note that this anticipation does not involve a "real" prediction as it does not look ahead in time, only in space. Real predictions are used in [1, 2, 34] and this is the only approach that results in a significant flow improvement. The heuristic algorithm proposed in [167] also contains anticipation

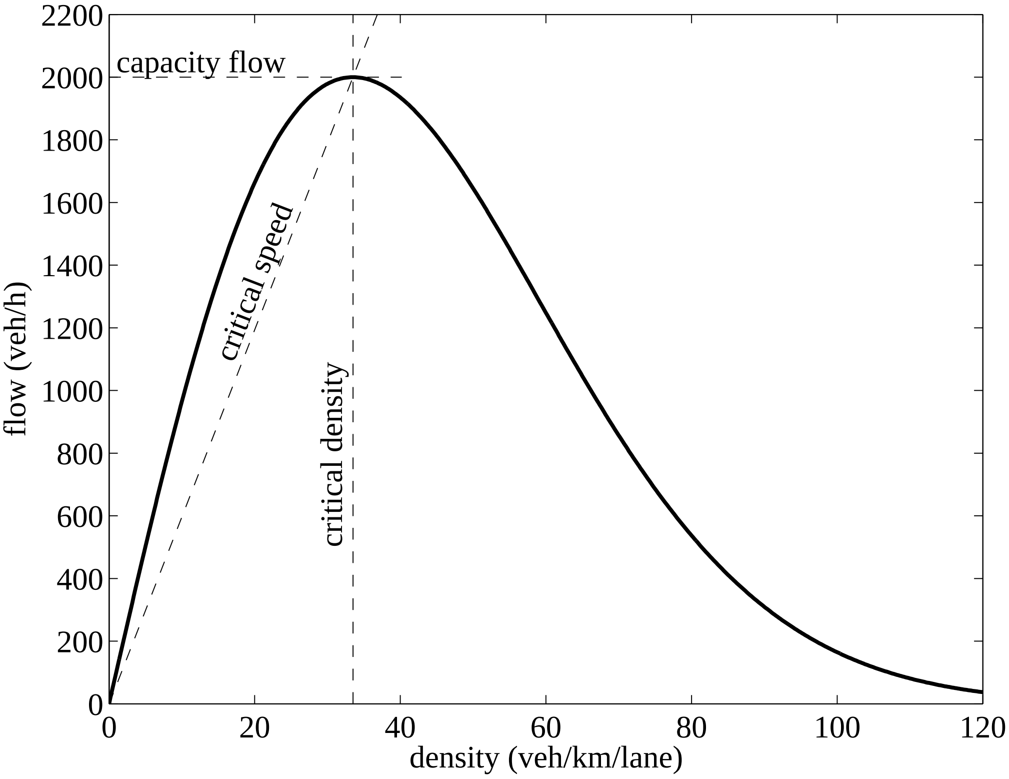


Figure 2.4: A typical example of the fundamental diagram. The fundamental diagram represents the traffic behavior on a homogeneous freeway (the spacial gradients of speed, flow and density equal zero). The meaning of the curve is the following. When the density is low, drivers travel at speeds close to the maximum allowed speed and the relationship between flow and density is approximately linear. When traffic gets more dense, drivers tend to reduce their speed until at a certain density, called the critical density, the capacity of the freeway is reached. When the density increases above the critical density, drivers tend to decrease their speed so strongly that the resulting flow is below capacity. The critical speed is the speed that corresponds to maximum flow. The slope of the line connecting the origin and a point on the fundamental diagram represents the speed corresponding to that point.

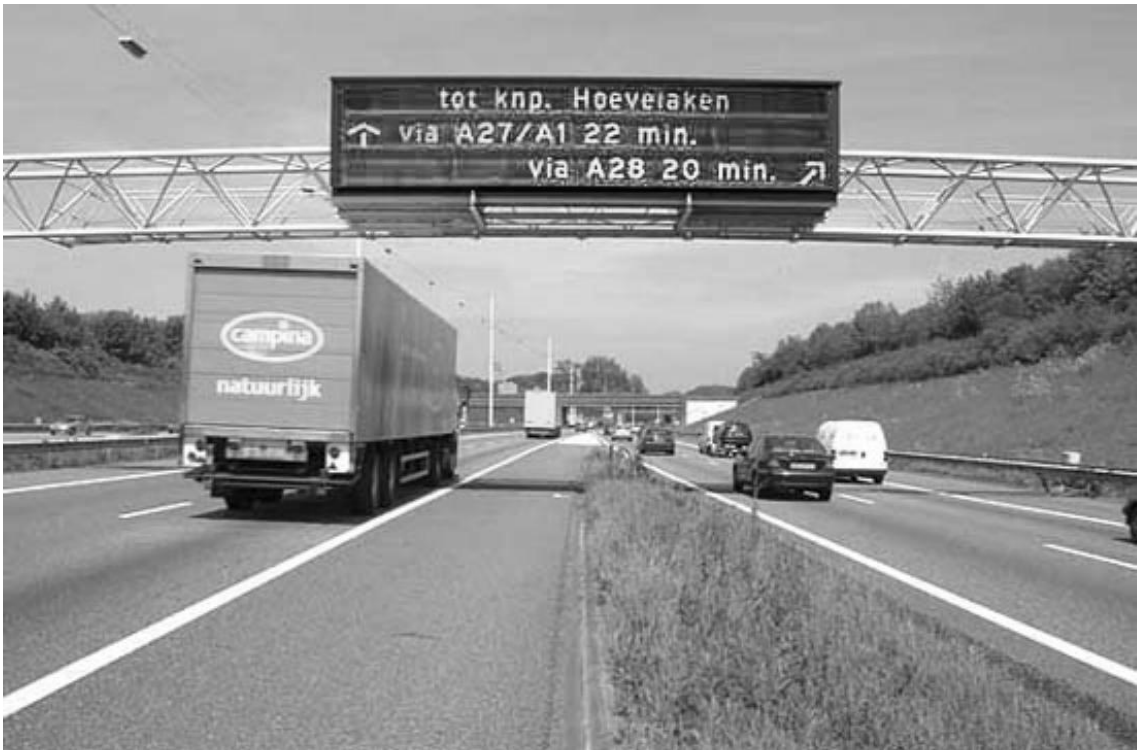


Figure 2.5: A route guidance system showing travel times for alternative routes to a common node.

to shock waves being formed.

Most application oriented studies [148, 72, 167, 150] enforce speed limits, except for [171, 97]. Enforcement is usually accepted by the drivers if the speed limit system leads to a more stable traffic flow.

As noted in [150] a common mistake in the argumentation for defining "optimal" speed levels, is based on misinterpretation of measurement results. Often it is (implicitly) assumed that stability or optimality observed at a certain speed can be reproduced by imposing that speed by speed limits. E.g., if the capacity flow is observed at 80 km/h this does not mean that if a speed limit is applied of 80 km/h then the flow will reach capacity, nor that 80 km/h is the optimal speed limit for any traffic situation.

For excellent overviews of practical speed limit systems see [167, 146].

2.3 Route guidance

Route guidance systems assist drivers in choosing their route when more alternative routes exist to their destination. The systems typically display traffic information on variable messages signs (see Figure 2.5) such as congestion length, travel time to the next common point on the alternative routes, or delay on the alternative routes. In the future possibly in-car systems could guide the driver individually to his destination taking into account the traffic situation on the alternative routes.

In route guidance the notions *system optimum* and *user equilibrium* (or *user optimum*) play an important role. The system optimum is achieved when the vehicles are guided such that the total costs of all drivers (typically the TTS) is minimized. However, the system optimum does not necessarily minimize the travel time for each individual driver. So, some drivers may have the choice for another route that has lower cost (shorter individual travel time). The traffic network is in user equilibrium when the costs on each utilized alternative route the cost is equal and minimal, and on routes that are not utilized the cost is higher that on the utilized routes. This means that no driver has the possibility to find another route that reduces his individual cost.

If the cost function is defined as the travel time it is typically defined as the *predicted travel time* or as the *instantaneous travel time* (or *reactive travel time*). The predicted travel time is the time that the driver will experience when he drives along the given route, while the instantaneous travel time is the travel time determined based on the current speeds on the route. In a dynamic setting these speeds may change when the driver travels over the route, and consequently the instantaneous travel time may be different from the predicted travel time.

Papageorgiou [123] and Papageorgiou and Messmer [131] have developed a theoretical framework for route guidance in traffic networks. Three different traffic control problems are formulated: an optimal control problem to achieve system optimum (minimize TTS), an optimal control problem to achieve user optimum (equalize travel times), and a feedback control problem to achieve user optimum (equalize travel times). The resulting controller strategies are demonstrated on a test network with six pairs of alternative routes. The feedback control strategy is tested with instantaneous travel times and results in a user equilibrium for most alternative routes, and the resulting TTS is very close to the system optimum.

Wang *et al.* [164, 165] combine the advantages of a feedback approach (relatively simple, robust, fast) and predicted travel times (necessary to achieve exact user equilibrium). The resulting *predictive feedback* controller is compared with optimal control and with a feedback controller based on instantaneous travel times. When the disturbances are known the results show that the predictive feedback results in nearly optimal splitting rates, and is clearly superior to the feedback based on instantaneous travel times. The robustness of the feedback approach is shown for several cases: incorrectly predicted demand, an (unpredictable) incident, and an incorrect compliance rate.

The studies [123, 131, 164, 165] assume that the turning rates can be manipulated by appropriate traffic control measures. In the case of in-car systems it is plausible that by giving direct route advice to individual drivers the splitting rates can be influenced sufficiently. However, in the case of route guidance by VMSs or DRIPs the displayed message does not directly determine the splitting rate: the drivers make their own decisions. Therefore, empirical studies about drivers' reaction to DRIP messages, and the effectiveness of route guidance can provide useful information.

Kraan *et al.* [95] present an extensive evaluation of the impact on network performance of VMSs on the freeway network around Amsterdam. Several performance indicators are compared before and after the installation of 14 new VMSs (of which 7 are used as incident management signs and 7 as dynamic route information signs). The performance indicators used for comparison are:

- *Total traveled distance (veh.km)* by all vehicles in the network during the peak period.
- *Total congestion length and duration (km.min)* occurring in the network during the peak period, where congestion is defined as traffic traveling at speed of 35 km/h or lower.
- *Instantaneous travel time delay (veh.h)* the delay for all drivers during the peak period, based on instantaneous travel time calculations.

The performance indicators are compared for alternative routes and for most locations a small but statistically significant improvement is found. The day-to-day standard deviation of these indicators decreased after the installation of the VMSs, which indicates that the travel times have become more reliable.

In the paper [95] the user response to VMSs messages (showing congestion lengths) is also analyzed. It is found that for each additional kilometer of queue length displayed for a route leads to a reduction of between 0.8 % and 1.6 %.

2.4 Other control measures

Besides ramp metering, dynamic speed limits, and route guidance, there are also other dynamic traffic control measures that can potentially improve the traffic performance. In this section we describe such measures, and describe in what situation they are useful (cf. [111]).

- **Peak lanes.** During peak hours the hard shoulder lane of a freeway (which is normally used only by vehicles in emergency) is opened for traffic. Whether the lane is opened or closed is communicated by VMSs showing a green arrow or a red cross. Due to the extra lane the capacity of the road is increased which could prevent congestion. The disadvantage of using the emergency lane as a normal lane is that the safety is reduced. This traffic control measure is useful where the additional capacity prevents congestion and the downstream infrastructure can accommodate the increased traffic flow.

- **Dedicated lanes.** In congestion the shoulder lane may be opened for dedicated vehicles, such as public transport or freight transport. This reduces the hindrance that congestion causes to these vehicles. Furthermore, public transport can be made more reliable and thus more attractive by this measure. A dedicated freight transport lane increases the stability and homogeneity of the traffic flow.

- **Bi-directional lanes.** A bi-directional lane is a freeway lane that can be used in both directions. Depending on the direction of the highest traffic demand the direction of operation is determined. The direction is communicated by a VMS showing a red cross or a green arrow. This traffic control measure is useful when the traffic demand is typically not high in both directions simultaneously.

- **The “keep your lane” directive.** When the “keep your lane” directive is displayed, the drivers are not allowed (not recommended) to change lanes. This results in less disturbances in the freeway traffic flow, which may prevent congestion. This traffic control measure is useful when the traffic flow is nearly unstable (close to the critical density) and may be a good alternative to the *homogenizing* speed limits.

2.5 Coordinated and integrated traffic control systems

In the literature basically three approaches exist for coordinating traffic control measures: model-based optimal control methods, knowledge-based methods, and methods that use simple feedback or switching logic.

2.5.1 Model-based optimal control methods

Model-based optimal traffic control techniques use a model for predicting the future behavior of the traffic system based on

1. the current traffic state,

2. the expected traffic demand on the network level, possibly including origin-destination relationships, and other possible external influences, such as weather conditions.

3. the planned traffic control measures.

Since the first two items cannot be influenced on the short term[6], the future performance of the traffic system is optimized by selecting an appropriate scenario for the traffic control

[6] Except for the possibility that people postpone or cancel their planned trip based on real-time congestion information.

measures. Methods that use optimal control or model predictive control take the complex nonlinear nature of traffic explicitly into account. E.g., they take into account the fact that the effect of ramp metering on distant on-ramps will be delayed by the (time-varying) travel time between the two on-ramps. In general, the other existing methods do not take this kind of delay into account. Furthermore, other advantages of these methods are that traffic demand predictions can be utilized, constraints on the ramp metering rate and the ramp queues can be included easily, and a user-supplied objective function can be optimized. In addition, the rolling horizon framework (used in model predictive control, see Chapter 4) has the advantage that it can handle demand prediction errors, disturbances (incidents); it can be made adaptive by updating the prediction model on-line, and it is computationally more efficient due to the shorter prediction and control horizon. The details will be explained in Chapter 4.

In [93, 91] a small network is studied with the possibility of route guidance and ramp metering. A feasible-direction optimization algorithm is applied to find the control signals that minimize a weighted sum of the TTS, the control signal variation, and a term that penalizes too long ramp queues. In the no-control situation the performance degradation is mainly caused by back-propagating congestion caused by an on-ramp that blocks also another traffic stream with a route that does pass the on-ramp. The TTS is improved by 16 % when only ramp metering is applied, by 29 % when only route guidance is applied, and by 30.5 % when both ramp metering and route guidance is used. In [94] integrated control of ramp metering and freeway-to-freeway control is applied to the ring road of Amsterdam, The Netherlands. Depending on the admissible queue length on the on-ramps and the freeway-to-freeway links, the improvement of the TTS was 31.7 % or 37.8 %. When only the on-ramps were controlled, the improvement of the TTS was 22 %. In [34] similar techniques are used to coordinate ramp metering and dynamic speed limits on a freeway stretch. The controller aims at minimizing a weighted sum of the TTS and several penalty terms for too high or too low speeds, densities, or queue lengths. The control law resulting from optimal control is approximated by a neural network, and it demonstrated by simulation that the difference in performance (TTS) between the two is 0.25 % for the deterministic situation and 0.7 % for the stochastic situation.

2.5.2 Knowledge-based methods

Knowledge-based traffic control methods typically describe the knowledge about the traffic system in combination with the control system in terms that are comprehensible for humans. Via reasoning mechanisms the knowledge-based system generates a solution (control measure) given the current traffic situation. A typical motivation for these systems is to help traffic control center operators to find good (not necessarily the best) combinations of control measures. The operators often suffer from cognitive overload by the large number of possible actions (control measures) or by time pressure in case of incidents. The possibility for the operators to track the reasoning path of the knowledge-based system makes these systems attractive.

Such a system is the TRYS system [70, 26, 113], which stores the knowledge in different knowledge bases:

- Physical structure of the network: the sensor data is processed here (data abstraction).
- Traffic problems: knowledge about the detection and diagnosis of the presence of incidents or congestion. The severity is estimated and the cause of the problem is determined.
- Control actions: knowledge about the definition of control strategies adequate to solve the different problems.

In the TRYS system the controlled traffic network is divided in several partially overlapping areas, each of them having its own set of knowledge bases. On top of these knowledge bases a coordinator unit removes the incompatible control actions (such as different messages for one variable message sign proposed by the agents of different areas, or semantically conflicting messages), using a rule-base.

The TRYS system has been installed in traffic control centers in Madrid and Barcelona.

The freeway incident management system [45] developed in Massachusetts assists in the management of non-recurrent congestion. The system contains a knowledge base and a reasoning mechanism to lead the traffic operators through the appropriate questions to manage incidents. Besides incident detection and verification the system assists in notifying the necessary agencies (e.g., ambulance, clean up forces, towing company) and in applying the appropriate diversion measures. The potential benefits (reduced travel times by appropriate diversion) are illustrated by a case-study on the Massachusetts Turnpike. The knowledge-based expert system called freeway real-time expert-system demonstration [142, 173, 141] has similar functionality and is illustrated by applying it to a section of the Riverside Freeway (SR-91) in Orange County, California.

2.5.3 Control parameter optimization

Allessandri and Di Febbraro [1, 2] follow another approach: a relatively simple control law is used for speed limit control and ramp metering, and the parameters of the control law are found by simulating a large number of scenarios and optimizing the average performance. In [2] a dynamic speed limit switching scheme is developed. The speed limits switch between approximately 70 km/h and 90 km/h, and the switching is based on the density of the segment to be controlled and two thresholds (to switch up and to switch down). The switching scheme uses a hysteresis loop to prevent too frequent switching.

Optimizing the thresholds with Powell’s method for several objectives resulted in an average increase of the throughput of 0.6 %, a decrease of squared densities (which can be considered as a measure of inhomogeneity) of 3.8 %, and a decrease of the TTS of 1 %. In [1] ramp metering is applied in addition to the speed limit control. The ramp metering scheme is the demand-capacity algorithm, which has only one parameter: the main-stream capacity. This capacity together with the density thresholds are optimized according to several objectives. The average result for several simulations was an increase of throughput with 9 %, a decrease of squared densities of 28 %, and a decrease of TTS of 18 %.

2.6 Summary

In this chapter an overview is given of traffic control measures that could be useful to improve the performance of a traffic network. In particular we have considered ramp metering, dynamic speed limits, and route guidance in detail, since we will use these control measures in the simulation studies in this thesis. However the MPC approach developed in this thesis is also suitable for other control measures. Therefore, we have also discussed shortly some other possible traffic control measures, such as, peak lanes, dedicated lanes, bi-directional lanes, and the “keep your lane” directive.

The existing ramp metering strategies are categorized according to several properties. Ramp metering can be used in two different modes: the spreading mode and the restricting mode. The spreading mode aims at reducing the probability of a breakdown caused by a platoon of vehicles arriving from the on-ramp. The restricting mode aims at redirecting drivers to other routes, or at preventing demands that exceed freeway capacity. For this last aim several ramp metering strategies have been developed, which can be classified as static or dynamic, fixed-time or traffic-responsive, and local or coordinated.

For speed limits systems a distinction is made between approaches that aim at homogenizing and approaches that aim at the resolution of shock waves or jams. While in theory the homogenizing approach is promising, field studies show that the achievable improvement is negligible. The main disadvantage is that these systems cannot resolve congestion after breakdown has occurred. The approaches that aim at resolution of shock waves or jams use speed limits that are low enough to limit the inflow of the congested area while the homogenizing approach uses speed limits that are above the critical speed.

Route guidance systems may aim at the realization of user optimum or system optimum. To achieve system optimal route choice behavior the only approach developed is optimal control. User optimum can be reached by optimal control or approximated by a feedback approach. For the feedback approach the use of predicted times instead of instantaneous travel times give better results.

For coordinated and integrated systems three approaches are presented: model-based optimal control, knowledge-based methods, and methods with simple control laws with parameters optimized for a large number of simulations.

Chapter 3

Traffic flow modeling

In this chapter we consider traffic flow modeling issues. In Section 3.1 we give an overview of traffic flow models. The traffic flow models are classified according to *application area, level of detail,* and *deterministic versus stochastic*, or *continuous or discrete* process representation. In Section 3.2 we present the macroscopic traffic flow model METANET model [156, 93, 126, 91], which we extend in Section 3.3. The extended METANET model results in a good trade-off between efficiency and accuracy and will be used for the simulations in the subsequent chapters.

The main contribution of this chapter to the state of the art can be found in Section 3.3, where we extend the METANET model with:

- an explicit model for dynamic speed limits,
- a model for main-stream metering,
- a model for the main-stream origin which has different dynamics than on-ramps,
- different anticipation behavior at the head and the tail of shock waves,
- a formulation for the downstream boundary condition that can express scenarios where the downstream area is uncongested, except for some incoming shock waves.

3.1 Modeling overview

Several traffic flow models have been developed for different application areas [77], e.g.:

- **Assessment of traffic control strategies** with a simulation model instead of a field operation test has several advantages. Above all, simulation is cheaper and faster, but it also provides an environment where the unpredictable disturbances of a field test, such as weather influences, traffic demand variations, and incidents, can be excluded, or if necessary simulations can be repeated under exactly the same disturbance scenario.

- **Model-based traffic control** makes use of an internal prediction model in order to find the best traffic control measures to be applied to the real traffic process. Since these models are operated in real-time, and are often used to evaluate several control scenarios, they need to be fast when executed on a computer.

- **Design of new transportation facilities,** e.g., geometric design of infrastructure can benefit from simulations that confirm that the design meets the specifications.

- **Training of traffic operators** in traffic control centers is supported by simulations that instantly give feedback about the consequences of the actions of the traffic operators in a certain situation.

Since none of the available traffic models perfectly describes the real traffic behavior one has to keep in mind the intended application, when making the choice between the available traffic flow models. As Papageorgiou [124] argues for macroscopic traffic flow models (but this can be generalized to any traffic flow model) an important criterion is that the model should have sufficient descriptive power to reproduce all important phenomena for the intended application. In Chapter 5 we will describe the necessary phenomena that the traffic flow model should be able to reproduce in order to be useful for model-based on-line control. Another important criterion for traffic flow models is the execution speed of a simulation. For off-line traffic control strategy assessment the trade-off between speed and accuracy can result in a choice for an accurate but slower model, while for model-based traffic control the model should be considerably faster than real-time.

Besides the intended application there are also other possible classifications for traffic flow models:

Level of detail. Traffic models may distinguished according to the level of detail at which they describe the traffic process: microscopic, mesoscopic or macroscopic.

- **Microscopic** models describe the behavior of vehicles individually. Important aspects of microscopic models are the so-called *car-following* and *lane-changing* behavior. Car-following and lane-changing behavior is generally described as a function of the distance to and (relative) speed of the surrounding vehicles, and the desired speed by the actual vehicle. Since the vehicles are modeled individually in microscopic traffic models, it is easy to assign different characteristics to each vehicle. These characteristics can be related to the driving style of the driver (aggressive, patient), vehicle type (car, truck), its destination, and chosen route. Examples of well-known microscopic simulation packages are AIMSUN2 [7], Vissim [136], Paramics [137], and FLEXSYT-II- [112]. See [3] for an extensive comparison of microscopic simulation models.

A special type of microscopic traffic models are the *cellular-automaton* models [114, 115, 169] where the freeway is partitioned into cells of about 7.5 m length. Each cell can contain one vehicle or can be empty. Depending on the speed of the vehicle it can hop one or more cells forward in each simulation step. The speed evolution of the vehicle is described by two deterministic processes: *acceleration* toward the desired speed, and *deceleration*, which means breaking to avoid collision with the vehicle in front. The third process in this model is the probabilistic *random deceleration* which expresses the inability of the drivers to maintain a constant speed without automatic cruise control. This extremely simple and computationally efficient model can reproduce shock waves and metastability in traffic flow. However, it is unknown how to calibrate the cellular-automaton models with real traffic data.

- **Mesoscopic** models do not track individual vehicles, but describe the behavior of individual vehicles in probabilistic terms. Examples of mesoscopic models are: headway distribution models [14, 16], cluster models [13] and gas-kinetic models [133, 65, 73].

- **Macroscopic** models use a high level of aggregation without distinguishing between individual vehicle actions such as a lane change. Instead traffic is described in aggregate terms as average speed, average flow, and average density. These models can further be classified according to the number of independent state variables.

  We may say that macroscopic traffic flow modeling started when Lighthill and Whitham [104] presented in 1955 a model based on the analogy between traffic flow and flow in rivers. Independently of Lighthill and Whitham one year later Richards [140] published a similar model. Therefore, this model is usually referred to as the Lighthill-Whitham-Richards (LWR) model. This model is simple and predictions can be calculated easily. However, it has also some disadvantages (cf. [30] and [124]):

  1. The LWR is a continuous model where the speed $v(x,t)$ at location x and time t is a function of the density $\rho(x,t)$, according to the relationship $v(x,t) = V(\rho(x,t))$, where $V(\rho(x,t))$ is the speed that drivers assume when they experience density $\rho(x,t)$. As a consequence, the speed instantaneously adapts to the speed prescribed by $V(\rho)$. This is quite unrealistic where the density changes are large, such as at the head and the tail of shock waves or traffic jams.
  2. According the LWR model the low-density tail of a shock wave will have a higher speed than the high-density body and therefore the tail will catch

up with the body and cause a sharp rear end of the shock wave, which is unrealistic.

3. In the LWR model the outflow of a shock wave or an (on-ramp) traffic jam is allowed to be equal to the capacity or the road, which is unrealistic. This property makes this model unsuitable for freeway traffic control where the capacity drop is one of the main reasons of the performance degradation. However, on urban streets the traffic flows are primarily governed by the signal timings and the prevention of the capacity drop does not play a significant role.
4. The LWR model does not predict instabilities of the stop-and-go type, which can occur at bottlenecks.

The first three disadvantages are solved by the Payne-type, where a second equation is introduced to describe the dynamics of the speed (see [134, 123]). These models describe the speed dynamics as a result of three processes:

- *Relaxation:* The drivers' tendency to accelerate or decelerate toward their desired speed $V(\rho)$.
- *Convection:* The speed evolution at a certain location x is also determined by the speed of vehicles upstream from x traveling toward x.
- *Anticipation:* The vehicles adjust their speed according to the traffic state immediately downstream from their location. Vehicles slow down when the density is increasing in downstream direction or accelerate when the density is decreasing.

However, the way how the Payne model solves the deficiencies of the LWR model is criticized by Daganzo on the following points [30]:

- Models with a diffusion (anticipation) term may result in negative flows in situations where the spacial derivative of the density is large. This is unrealistic.
- Wave characteristics that result form Payne's model may be faster than the mean speed of traffic. This seems to be unrealistic because it implies that information is traveling faster than the vehicles in the traffic flow.

As a reply to these critics Papageorgiou [124] proposed to simply assume zero speed when the model would predict negative speeds. He also argues that the fast characteristics are not necessarily unrealistic because the Payne model describes mean speeds, and since there may be vehicles that travel faster than the mean speed the fast characteristics could represent those vehicles.

Note that other improvements of the LWR model have been made in [29, 106, 102].

Finally, we note that high-order models are also capable of predicting more complex phenomena, such as stop-and-go waves at bottlenecks (see [75, 159, 67, 65]). Since the reproduction of these phenomena is not necessary for our approach, we will not discuss these models in detail here.

Deterministic versus stochastic representation of the process. Deterministic models define a relationship between model inputs, variables and outputs that typically describes the average behavior of traffic. Stochastic models describe traffic behavior in terms of relationships between random variables, e.g., drivers' random reaction time, randomness in equilibrium speed-density (or car-following) relationships, route choice, etc. These stochastic effects can reproduce phenomena such as the creation of traffic jams by random fluctuations in traffic flows (see [147, 149]).

Continuous versus discrete representation of the process. Many traffic models are formulated in continuous time and space [75, 65, 77, 134, 104]. Since most of these models (except for the LWR model) are too complex to solve analytically, the models are usually solved numerically by discretization and simulation. Other models, such as the cell-transmission model [27, 28] and cellular-automaton model [169, 68] are formulated in discrete time and space.

We refer the interested reader to [77, 102, 106, 172, 64, 66, 63, 76, 75, 84, 27, 28] for a detailed overview of traffic flow models, and to [3] for an extensive comparison of microscopic simulation models.

3.2 The basic METANET model

In the experiments in Chapters 6, 7, 8 we use the METANET traffic flow model. This model was chosen because it provides a good trade-off between simulation speed and accuracy [88, 38, 118]. The fact that this model is deterministic, discrete-time, discrete-space, and macroscopic makes it very suitable for model-based traffic control. Since the simulation time step and the segment length of the discretized freeway are relatively large, the execution of the model simulation can be very fast. Note, however, that the MPC approach, which will be presented in Chapter 4 is generic so that we could also work with other traffic flow models.

Regarding the validation of the model we refer to [125, 38, 118]. The reported validation results are in general satisfactory, except for the results in [38], which can be explained by the fact that the model in [38] was not calibrated before validation. Furthermore, the small number of parameters makes it easy to calibrate. An important property for model-based traffic control is that this model (including extensions) can reproduce capacity drop at on-ramps and in shock waves. In the subsequent sections we describe the METANET model and the extensions we have made to this model.

The METANET model can operate in two modes: the destination-independent and the destination-dependent mode. The destination-dependent mode is useful for networks

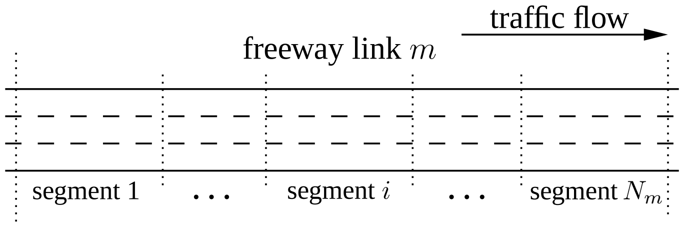


Figure 3.1: In the METANET model a freeway link is divided into segments.

that have multiple destinations and the possibility for route choice. We first present the equations for the destination-independent mode, and next we give the extensions necessary to represent destination dependent traffic. For the full description of METANET we refer to the METANET manual [156] and to the literature [93, 126, 91].

3.2.1 Link equations

The METANET model represents a network as a directed graph with the links (indicated by the index m) corresponding to freeway stretches. Each freeway link has uniform characteristics, i.e., no on-ramps or off-ramps and no major changes in geometry. Where major changes occur in the characteristics of the link or in the road geometry (e.g., on-ramp or an off-ramp), a node is placed. Each link m is divided into N_m segments (indicated by the index i) of length L_m (typically 500-1000 m, see also Figure 3.1). Each segment i of link m is characterized by three quantities:

- *traffic density* $\rho_{m,i}(k)$ (veh/km/lane),
- *mean speed* $v_{m,i}(k)$ (km/h),
- *traffic volume* or *outflow* $q_{m,i}(k)$ (veh/h),

where k indicates the time instant $t = kT$, and T is the time step used for the simulation of the traffic flow (typically $T = 10\,\mathrm{s}$). For stability, the segment length and the simulation time step should satisfy for every link m

$$L_m > v_{\mathrm{free},m} T \quad ,$$

where $v_{\mathrm{free},m}$ is the average speed that drivers assume if traffic is freely flowing.

The outflow of each segment is equal to the density multiplied by the mean speed and the number of lanes on that segment (denoted by λ_m):

$$q_{m,i}(k) = \rho_{m,i}(k)\, v_{m,i}(k)\, \lambda_m \quad . \tag{3.1}$$

The density of a segment equals the previous density plus the inflow from the upstream segment, minus the outflow of the segment itself (conservation of vehicles):

$$\rho_{m,i}(k+1) = \rho_{m,i}(k) + \frac{T}{L_m \lambda_m}\big(q_{m,i-1}(k) - q_{m,i}(k)\big) \ . \tag{3.2}$$

While equations (3.1) and (3.2) are based on physical principles and are exact, the equations that describe the speed dynamics and the relation between density and the desired speed are heuristic. In the METANET model the mean speed at the simulation step $k+1$ is taken to be the mean speed at time instant k plus a *relaxation term* that expresses that the drivers try to achieve a desired speed $V(\rho)$, a *convection term* that expresses the speed increase (or decrease) caused by the inflow of vehicles, and an *anticipation term* that expresses the speed decrease (increase) as drivers experience a density increase (decrease) downstream:

$$\begin{aligned} v_{m,i}(k+1) = v_{m,i}(k) + \frac{T}{\tau}\Big(V\big(\rho_{m,i}(k)\big) - v_{m,i}(k)\Big) + \\ \frac{T}{L_m} v_{m,i}(k)\big(v_{m,i-1}(k) - v_{m,i}(k)\big) - \\ \frac{\eta T}{\tau L_m} \frac{\rho_{m,i+1}(k) - \rho_{m,i}(k)}{\rho_{m,i}(k) + \kappa} \ , \end{aligned} \tag{3.3}$$

where τ, η[1] and κ are model parameters, and with

$$V\big(\rho_{m,i}(k)\big) = v_{\text{free},m} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_{\text{crit},m}}\right)^{a_m}\right] \ , \tag{3.4}$$

with a_m a model parameter, and where the free-flow speed $v_{\text{free},m}$ is the average speed that drivers assume if traffic is freely flowing, and the critical density $\rho_{\text{crit},m}$ is the density at which the traffic flow is maximal on a homogeneous freeway. We present in Figure 3.2 for an example of the speed-density relationship $V(\rho)$, also called the fundamental diagram.

Origins are modeled with a simple queue model. The length $w_o(k)$ of the queue at origin o equals the previous queue length plus the demand[2] $d_o(k)$, minus the outflow $q_o(k)$:

$$w_o(k+1) = w_o(k) + T\big(d_o(k) - q_o(k)\big) \ .$$

[1] In the original METANET model this parameter is denoted by ν (nu), but because of the small typographical difference with v (speed) we prefer to use η.

[2] Just as in [93, 94, 125], we assume that the demand is independent of any control actions taken in the network.

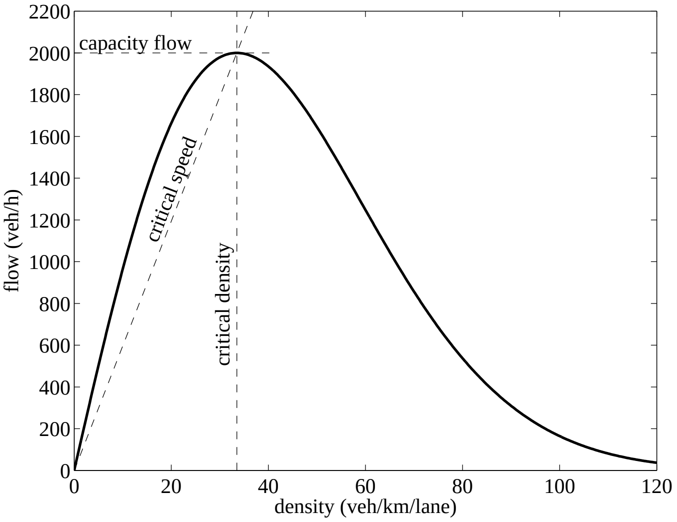


Figure 3.2: An example of the speed-flow relationship (3.4), with $a_m = 1.867$*,* $v_{\text{free},m} = 102$ *km/h,* $\rho_{\text{crit},m} = 33.5$ *veh/km/lane.*

The outflow of origin o depends on the traffic conditions on the main-stream and, for the metered on-ramp, on the ramp metering rate[3] $r_o(k)$, where $r_o(k) \in [0,1]$. The ramp flow $q_o(k)$ is the minimum of three quantities:

- the available traffic at simulation step k (queue plus demand),
- the maximal flow allowed by the metering rate,
- and the maximal flow that could enter the freeway because of the main-stream conditions.

So,

$$q_o(k) = \min\left[d_o(k) + \frac{w_o(k)}{T},\ C_o r_o(k),\ C_o\left(\frac{\rho_{\max} - \rho_{m,1}(k)}{\rho_{\max} - \rho_{\text{crit},m}}\right)\right], \tag{3.5}$$

where C_o is the on-ramp capacity (veh/h) under free-flow conditions, the global parameter $\rho_{\max}$ (veh/km/lane) is the maximum density of a segment (also called jam density), and m is the index of the link to which the on-ramp is connected.

Remark 3.2.1 In the literature two slightly different formulations of ramp metering are published for the METANET model. The variant above can be found in [92, 93]. In the second variant [156, 94] the ramp flow $q_o(k)$ equals the ramp metering rate times

[3] For an unmetered on-ramp we also can use (3.5) by setting $r_o(k) \equiv 1$.

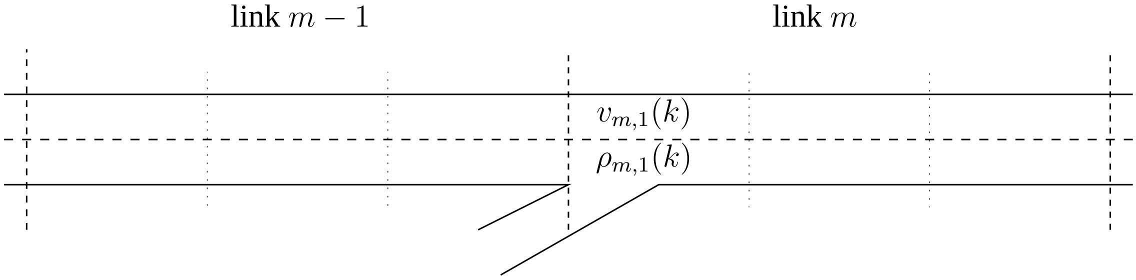


Figure 3.3: When there is an on-ramp connected to the freeway the speed $v_{m,1}(k)$ in the first segment of link m is reduced by merging phenomena according to (3.7).

minimum of the available traffic at simulation step k (queue plus demand), the capacity of the on-ramp, and the maximal flow that can enter the freeway because of the mainstream conditions:

$$q_o(k) = \tilde{r}_o(k) \min \left[d_o(k) + \frac{w_o(k)}{T},\ C_o,\ C_o \left(\frac{\rho_{\max,m} - \rho_{m,1}(k)}{\rho_{\max,m} - \rho_{\mathrm{crit},m}} \right) \right] , \tag{3.6}$$

where $\tilde{r}_o(k) \in [0,1]$ is the ramp metering rate. In the experiments in Chapters 6, 7, 8 we prefer the first formulation, because there a constant ramp metering rate corresponds to a constant maximum flow that is allowed to enter the freeway. This is an advantage when the ramp metering rate is bounded and the lower and upper bounds should correspond to certain flows (in practice determined by the bounds on the cycle time or the red and green times). The advantage of the second formulation is that metering rates $r_o(k) < 1$ correspond to situations where ramp metering is the limiting factor on the ramp flow (and not the on-ramp capacity or the traffic conditions on the freeway), which may be an advantage when interpreting plots of the ramp metering signal. □

In order to account for the speed drop caused by merging phenomena, if there is an on-ramp, then the term

$$-\frac{\delta T q_o(k) v_{m,1}(k)}{L_m \lambda_m (\rho_{m,1}(k) + \kappa)} \tag{3.7}$$

is added to (3.3), where $v_{m,1}(k)$ and $\rho_{m,1}(k)$ are the speed and density of the segment that the on-ramp is connected to, as shown in Figure 3.3, and δ is a model parameter.

When there is a lane drop as shown in Figure 3.4, the speed reduction due to weaving phenomena,

$$-\frac{\phi T \Delta\lambda_m \rho_{m,N_m}(k) v_{m,N_m}^2(k)}{L_m \lambda_m \rho_{\mathrm{crit},m}} , \tag{3.8}$$

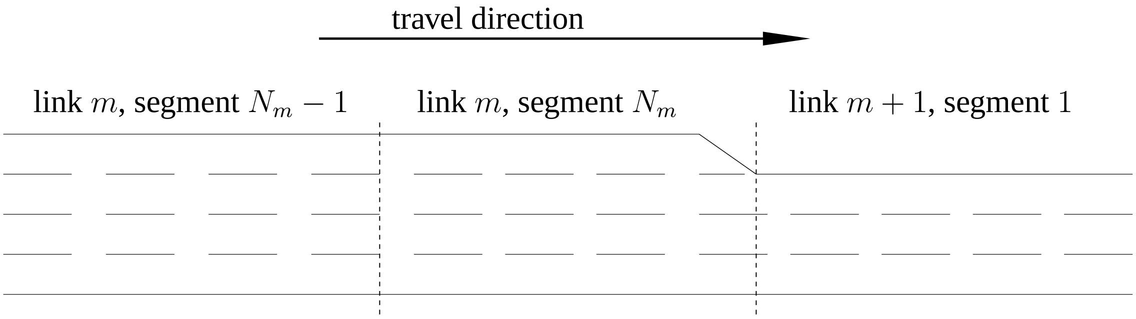


Figure 3.4: When there is a lane drop the speed $v_{m,N_m}(k)$ in the last segment of link m is reduced by merging phenomena according to (3.8).

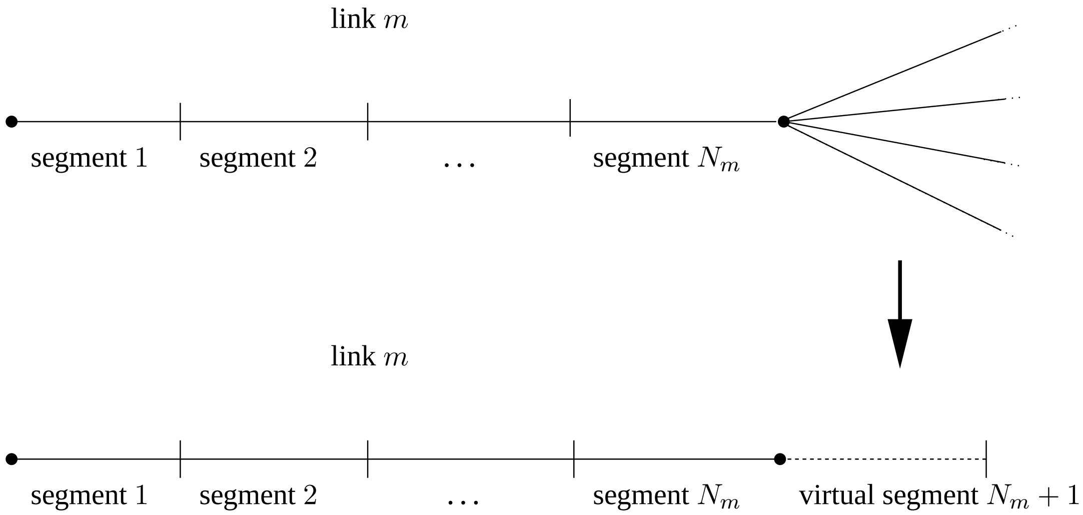


Figure 3.5: A node with one entering link m and several leaving links. The densities in the first segments of the leaving links are aggregated in the virtual downstream density $\rho_{m,N_m+1}(k)$ according to (3.9).

is added to (3.3), where $\Delta\lambda_m = \lambda_m - \lambda_{m+1}$ is the number of lanes being dropped, and ϕ is a model parameter.

3.2.2 Node equations

The coupling equations to connect links are as follows. Every time there is a major change in the link parameters or there is a junction or a bifurcation, a node is placed between the links. This node provides the incoming links with a downstream density (or a virtual downstream density when there are more leaving links), and the leaving links with an inflow and an upstream speed (or a virtual upstream speed when there are more entering links). The flow that enters node n is distributed among the leaving links according to

$$Q_n(k) = \sum_{\mu \in I_n} q_{\mu,N_\mu}(k) \ ,$$
$$q_{m,0}(k) = \beta_{n,m}(k) Q_n(k) \ ,$$

where $Q_n(k)$ is the total flow that enters the node at simulation step k, I_n is the set of links that enter node n, $\beta_{n,m}(k)$ are the turning rates (the fraction of the total flow through node n that leaves via link m), and $q_{m,0}(k)$ is the flow that leaves node n via link m, where link m is one of the links leaving node n.

When node n has more than one leaving link as shown in Figure 3.5, the virtual downstream density $\rho_{m,N_m+1}(k)$ of entering link m is given by

$$\rho_{m,N_m+1}(k) = \frac{\sum\limits_{\mu \in O_n} \rho_{\mu,1}^2(k)}{\sum\limits_{\mu \in O_n} \rho_{\mu,1}(k)} \ , \tag{3.9}$$

where O_n is the set of links leaving node n.

When node n has more than one entering link as shown in Figure 3.6, the virtual upstream speed $v_{m,0}(k)$ of leaving link m is given by

$$v_{m,0}(k) = \frac{\sum\limits_{\mu \in I_n} v_{\mu,N_\mu}(k) q_{\mu,N_\mu}(k)}{\sum\limits_{\mu \in I_n} q_{\mu,N_\mu}(k)} \ . \tag{3.10}$$

3.2.3 Boundary conditions

Boundary conditions need to be defined for the entries and exits of the traffic network. As in METANET the state of a segment also depends on the upstream speed, the outflow of the upstream node, and the downstream density, we need to define the upstream speed and the inflow for the entries of the network, and the downstream density for the exits of the network. These boundary conditions can be user-specified or a default value can be assumed. We already presented the boundary conditions for the traffic demand at on-ramps, now we present the boundary conditions for the upstream speed and the downstream density.

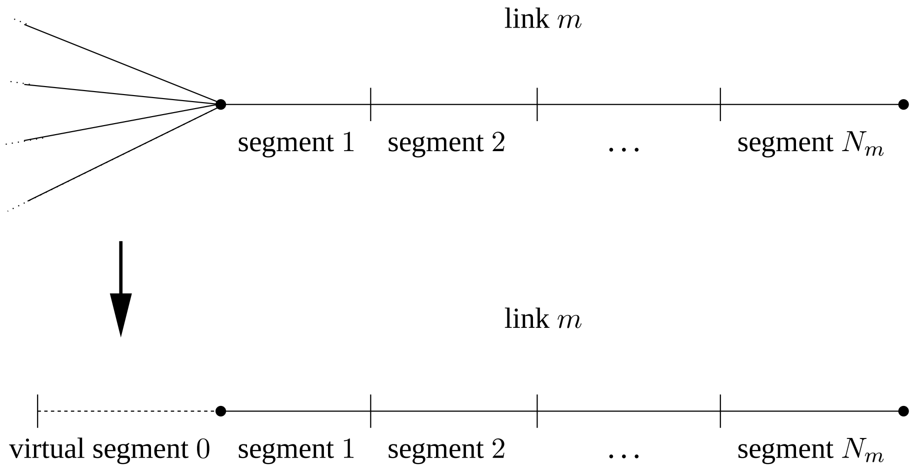


Figure 3.6: A node with one leaving link m and several entering links. The speeds in the last segments of the entering links are aggregated in the virtual upstream speed $v_{m,0}(k)$ according to (3.10).

Upstream speed

When there is a main-stream origin entering node n the virtual speed $v_\mu(k)$ of the origin can be user-specified, where μ is the index of the origin. If $v_\mu(k)$ is not specified then it equals the speed of the first segment of the leaving link

$$v_\mu(k) = v_{m,1}(k) \ .$$

If there is a second incoming link besides the origin then the speed of the last segment of the second incoming link is taken for $v_{m,0}(k)$.

Downstream density

Similarly, the virtual downstream $\rho_{m,N_m+1}(k)$ density for the entering link at a node that is connected to a destination, is calculated as follows. First, the user can specify a destination density scenario $\rho_\mu(k)$ where μ is the index of the destination link. Alternatively, a flow limitation $q_{\text{bound},\mu}(k)$ can be defined, and the virtual downstream density is calculated according to

$$\rho_\mu(k+1) = \begin{cases} \rho_{\text{upstream},n}(k) \\ \qquad \text{if } q_\mu < q_{\text{bound},\mu}(k) \text{ and } \rho_{\text{upstream},n}(k) < \rho_{\text{crit},\mu} \\ \rho_\mu(k) + C_\mu\big(q_\mu(k) - q_{\text{bound},\mu}(k)\big) \\ \qquad \text{else.} \end{cases}$$

where $\rho_{\mathrm{upstream},n}(k)$ is the density of the upstream link, and C_μ is a parameter. If $\rho_\mu(k)$ nor $q_{\mathrm{bound},\mu}(k)$ is defined then

$$\rho_\mu(k) = \rho_{m,N_m}(k) \ ,$$

or if there is a second leaving link at node n then the density of the first segment of that link is taken for $\rho_{m,i+1}$.

3.2.4 The destination-dependent mode

For the destination-dependent mode the variable $\gamma_{m,i,j}(k_{\mathrm{f}})$ is introduced to express the fraction of traffic on link m, segment i that has destination j. The total density $\rho_{m,i}(k)$ is now decomposed into partial densities $\rho_{m,i,j}(k)$ for each destination j:

$$\rho_{m,i,j}(k) = \gamma_{m,i,j}(k)\rho_{m,i}(k) \ .$$

The conservation equation also becomes destination dependent:

$$\rho_{m,i,j}(k+1) = \rho_{m,i,j}(k) + \frac{T}{L_m\lambda_m}\big(\gamma_{m,i-1,j}(k)q_{m,i-1}(k) - \gamma_{m,i,j}(k)q_{m,i}(k)\big) \ ,$$

Origins are modeled with a destination-dependent queue model. The evolution of the partial queue length $w_{o,j}(k)$ at origin o with destination j is described by:

$$w_{o,j}(k+1) = w_{o,j}(k) + T\big(\gamma_{o,j}(k)d_o(k) - \gamma_{o,j}(k)q_o(k)\big),$$

where $d_o(k)$ is the traffic demand at the origin, $\gamma_{o,j}(k)$ the fraction of the demand traveling to destination j, and $q_o(k)$ the outflow of the origin, and

$$w_o(k) = \sum_{j\in J_o} w_{o,j}(k) \, ,$$

where J_o is the set of destinations reachable from origin o.

The flow that enters node n is distributed among the leaving links according to

$$\begin{aligned}
Q_{n,j}(k) &= \sum_{\mu\in I_n} q_{\mu,N_\mu}(k)\gamma_{\mu,N_\mu,j}(k) \\
q_{m,0}(k) &= \sum_{j\in J_m} Q_{n,j}(k)\beta_{n,m,j}(k), \\
\gamma_{m,0,j}(k) &= \frac{\beta_{n,m,j}(k)Q_{n,j}(k)}{q_{m,0}(k)} \, ,
\end{aligned}$$

where $Q_{n,j}(k)$ is the total flow that enters the node at simulation step k, N_μ the index of the last segment of link m, I_n is the set of links that enter node n, $\beta_{n,m,j}(k)$ are the splitting rates (the fraction of the total flow through node n with destination j that leaves via link m), J_m is the set of destinations reachable from link m, and $q_{m,0}(k)$ is the flow that leaves node n via link m.

3.3 The extended METANET model

In this section we present four extensions to the METANET model. The first two extensions are regarding the traffic control measures: dynamic speed limits and main-stream metering. The third and fourth extension are regarding the modeling of the main-stream origin (as opposed to on-ramps) which have different dynamics; and regarding the different anticipation behavior at the head and the tail of shock waves. Finally, we present a slight modification of the boundary conditions at destinations. All these extensions and modifications are used in the simulations in Chapters 6, 7, and 8.

3.3.1 Dynamic speed limits

Cremer's models

To the author's best knowledge all macroscopic speed limit models in the literature originate from one of the two models proposed by Cremer [24] based on observations made by Zackor [171]. The first model is given by

$$V(\rho, u) = v_{\text{free}} u \exp\left(-\frac{(1+u)}{4}\left(\frac{\rho}{\rho_{\text{crit}}}\right)^2\right) ,$$

and the second by

$$V(\rho, u) = v_{\text{free}} u \left(1 - \left(\frac{\rho}{\rho_{\max}}\right)^{(l-1)(3-2u)}\right)^{\frac{1}{1-m}}$$

where v_{free} is the free flow speed, ρ_{crit} is the critical density, $\rho_{\max}$ the jam density, $V(\rho, u)$ is the speed on a homogeneous freeway given the density ρ, and l and m are fitting parameters (with $0 < m < 1, l > 1$). The speed control input is u, $0.6 \leq u \leq 1$, and the corresponding speed limits are given in Table 3.1. The speed-density and flow-density plots are shown in Figure 3.7 for model 1 and in Figure 3.8 for model 2.

Both models have certain disadvantages that can be summarized as follows:

- In model 1 certain speed limits may result in traffic speeds that are considerably lower than the displayed speed limit and the speed that traffic would assume without

u	speed limit (km/h)
1	without limitation
0.87	105
0.73	90
0.6	75

Table 3.1: The relation between control variable u and the imposed speed limit defined by Cremer [24]. The free flow speed is assumed to be 125 km/h.

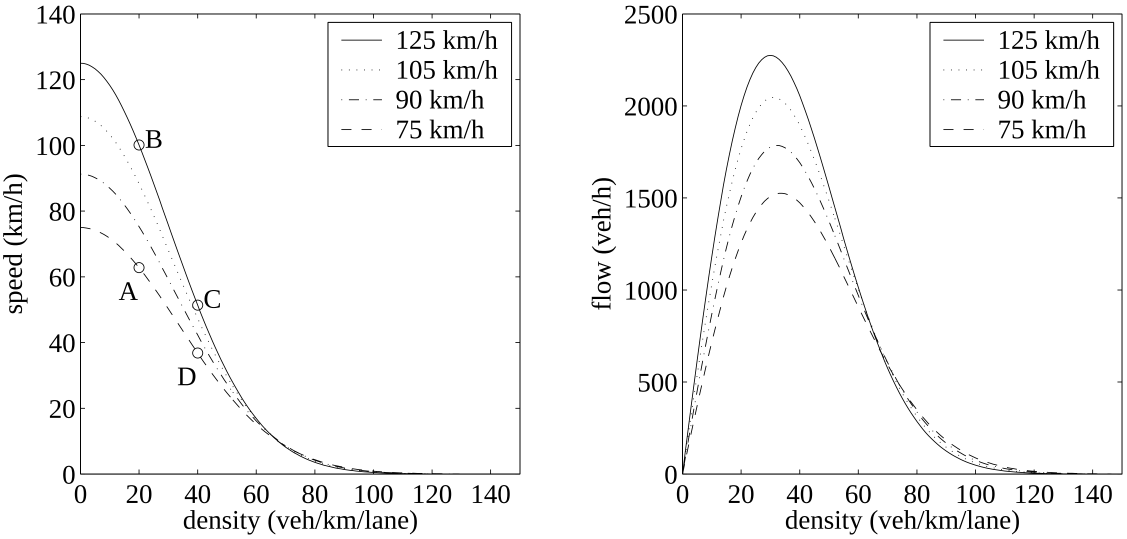


Figure 3.7: The speed-density and flow-density relationships for model 1. The speed limit of 125 km/h corresponds to the unlimited case.

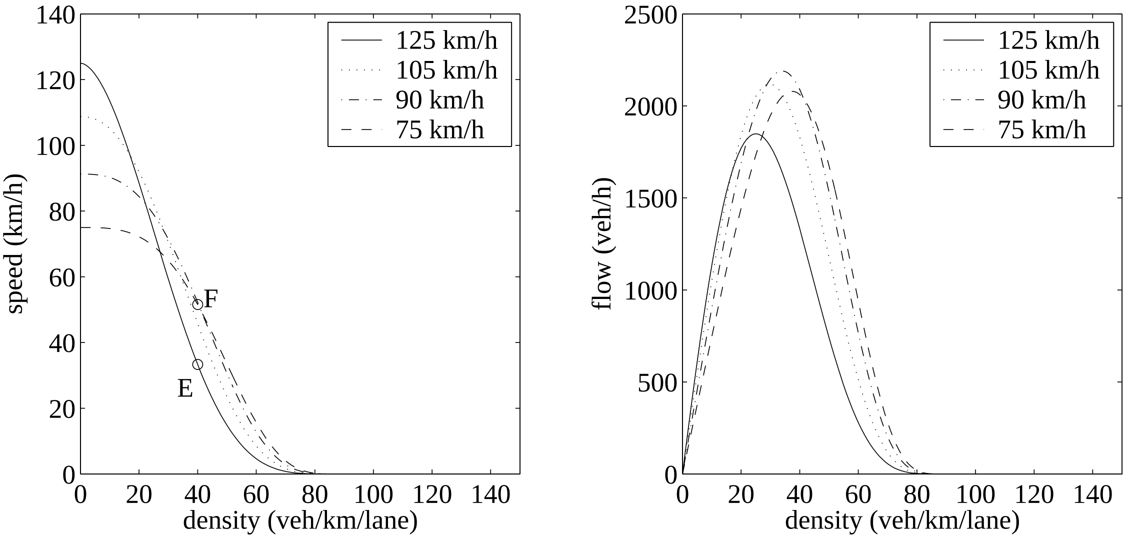


Figure 3.8: The speed-density and flow-density relationships for model 2. The speed limit of 125 km/h corresponds to the unlimited case.

speed limit. This is best explained by an example. In model 1 for a speed limit of 75 km/h and a density of 20 veh/km/lane the traffic speed is around 60 km/h (point A). However in the unlimited case the speed would be around 100 km/h (point B) for this density. It is rather unrealistic to assume that drivers would drive 15 km/h slower (from 75 km/h to 60 km/h) than allowed (recommended) speed when they would drive (and “feel safe”) around 100 km/h in the unlimited case.

- Furthermore, in model 1 the speed-density and flow-density characteristics are (approximately) scaled by u. This means that the densities that result in the uncontrolled case in speeds that are already lower than the speed limit are still affected. For example, if a speed limit of 75 km/h is imposed then the speed for the density of 40 veh/km/lane would drop from 52 km/h to 37 km/h (from point C to point D). This is also unrealistic.

- Model 2 describes a different behavior for the medium densities (around the critical density) and high densities (i.e., jam). For high densities the equilibrium speed actually *increases* when a *lower* speed limit is applied. See, e.g., the change from point E to point F. Also for medium densities the top of the flow-density curve shifts to the right and increases (up to a speed limit of 90 km/h). This is explained by Cremer by the variation of the individual speeds that is higher for the unlimited case than for the limited case. As lower speed variations result in a stabler flow, and a stabler flow allows for higher speeds at the same density, a higher traffic flow can be achieved. While in theory this is a possible explanation [147], some field experiments show that this effect is negligible [148, 72]. Cremer's model is based on the observation made in [170, 171] where the author also notes that the increased speeds were probably caused by drivers who thought the speed limits were advisory speeds instead of maximum speeds.

Despite the disadvantages of these models, all models that can be found in the literature are based on model 1 [100, 99] or model 2 [1, 2, 34]. In the next section we introduce a speed limit model that does not have these disadvantages and additionally can be tuned to represent enforced or unenforced situations.

A new speed limit model

Although METANET allows for speed limit modeling through changing the parameters $\rho_{\text{crit},m}$, $v_{\text{free},m}$, and a_m (as suggested in the METANET manual [156]) we believe that the following model is more consistent with the considerations in the previous section. We introduce in this section a model that only influences the speeds that are (would be) higher than the speed limit.

Similarly to the approach of Hoogendoorn [75] we introduce the speed limit in the speed equation is introduced such that the desired speed is the minimum of the following two quantities: the target speed based on the experienced traffic conditions (density), and the target speed caused by the displayed speed limit (Figure 3.9):

$$V\big(\rho_{m,i}(k)\big) = \min\left(v_{\text{free},m} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_{\text{crit},m}}\right)^{a_m}\right], (1+\alpha)v_{\text{control},m,i}(k)\right) , \tag{3.11}$$

where $v_{\text{control},m,i}(k)$ is the speed limit imposed on segment i, link m, at simulation step k, and $(1+\alpha)$ is the non-compliance factor. If $(1+\alpha) > 1$ it expresses that the drivers' target speed is higher that the displayed speed limit, if $(1+\alpha) < 1$ then the drivers' target speed is lower than the speed limit. In other words, the value of α is chosen such that if the average drivers' target speed is higher than the speed limit then α is positive, otherwise α is negative. Therefore, the factor $(1+\alpha)$ allows for modeling, e.g., enforced and unenforced speed limits.

Enforcement plays a dominant role in compliance as observed in data from the A1 freeway in The Netherlands near Deventer. On this freeway the Dutch government conducted a speed limit experiment[4] where the speed limits were enforced for several months. We have compared speeds for the enforced and unenforced situations (see Table 3.2).[5]

[4] This experiment was not intended to study compliance. but to study a speed limit switching scheme.

[5] The speed limit of 50 km/h in this table is not representative because an automatic incident detection system is installed on this freeway which automatically displays 50 km/h when the speed drops below approximately — in the unenforced case — 35 km/h.

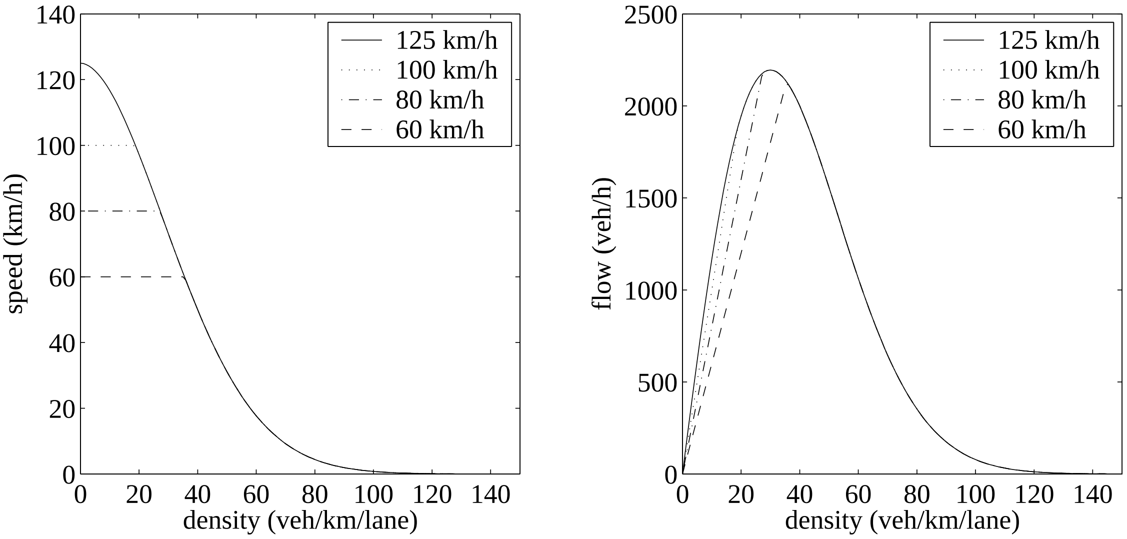


Figure 3.9: The speed-density and flow-density relationships of the new speed limit model proposed in this thesis, (with $v_{\mathrm{free},m} = 125$ km/h, $a_m = 1.867$, $\rho_{\mathrm{crit},m} = 30$ veh/km/lane, and $\alpha = 0$.)

speed limit (km/h)	average speed (km/h)		compliance (%)	
	unenforced	enforced	unenforced	enforced
50	34	37	82	81
70	78	64	32	66
80	–	73	–	91
90	99	–	11	–
100	–	90	–	93

Table 3.2: Average speeds under enforced and unenforced conditions on the A1 in The Netherlands. The dashes indicate that no data was available.

From the table we can conclude in general that when the speed limits are not enforced the average speed is approximately 10 % higher than what is displayed ($\alpha = 0.1$), and when they are enforced the average speed is approximately 10 % lower than what is displayed ($\alpha = -0.1$).

3.3.2 The modeling of main-stream metering

Main-stream metering restricts the flow on the freeway. The implementation of main-stream metering is often similar to ramp metering: a traffic light is placed on the freeway and one, two or multiple vehicles may pass when the light turns green. Note that the modeling tool METACOR also has this feature [90].

Main-stream metering on segment i of link m is modeled as a restriction on the outflow of the segment as follows:

$$q_{m,i}(k) = \min(r_{\text{msm}}(k)C_m,\ q_{\text{orig},m,i}(k))$$

where C_m is the nominal capacity[6] of link m, $r_{\text{msm}}(k)$ is the main-stream metering signal where $r_{\text{msm}}(k) \in [0,1]$, and $q_{\text{orig},m,i}(k)$ is the outflow of segment i of link m that follows from (3.1).

If $q_{m,i}(k) < q_{\text{orig},m,i}(k)$ the speed of the segment has to be adapted accordingly:

$$v_{m,i}(k) = \begin{cases} v_{\text{orig},m,i}(k)\dfrac{q_{m,i}(k)}{q_{\text{orig},m,i}(k)} & \text{if} \quad q_{m,i}(k) < q_{\text{orig},m,i}(k) \ , \\ v_{\text{orig},m,i}(k) & \text{otherwise,} \end{cases}$$

where $v_{\text{orig},m,i}(k)$ is the speed that follows from (3.3). Note that if we would adapt the density, then vehicles would be lost. However, the density will adapt to the new situation in the next simulation step as a consequence of the reduced outflow.

3.3.3 The modeling of main-stream origins

To express the different nature of a main-stream origin link o compared to a regular on-ramp (the queue at a main-stream origin is in fact an abstraction of the sections upstream of the origin of part of the freeway network that we are modeling), we use a modified version of (3.5) with another flow constraint, because the inflow of a segment (and thus the outflow of the main-stream origin) can be limited by an active speed limit or by the actual speed on the first segment (when either of them is lower than the speed at critical density). Hence, we assume that the maximal inflow equals the flow that follows from the speed-flow relationship from (3.1) and (3.4) with the speed equal to the speed limit or the actual speed on the first segment, whichever is smaller. So if o is the origin of link μ, then we have

$$q_o(k) = \min\left[d_o(k) + \frac{w_o(k)}{T},\ q_{\text{lim},\mu,1}(k)\right] \ ,$$

[6] We prefer not to use $q_{\text{cap},m} = \lambda_m V(\rho_{\text{crit},m})\rho_{\text{crit},m}$, which is the capacity following from the fundamental diagram, because equations (3.1), (3.2), (3.3), (3.4) also allow higher flows than $q_{\text{cap},m}$. Based on simulations, we chose C_m to be 5 % higher than $q_{\text{cap},m}$.

where $q_{\mathrm{lim},\mu,1}(k)$ is the maximal inflow determined by the limiting speed in the first segment of link μ:

$$q_{\mathrm{lim},\mu,1}(k) = \begin{cases} q_{\mathrm{speed},\mu}(k) & \text{if } v_{\mathrm{lim},\mu,1}(k) < V(\rho_{\mathrm{crit},\mu}) \\ q_{\mathrm{cap},\mu} & \text{if } v_{\mathrm{lim},\mu,1}(k) \geq V(\rho_{\mathrm{crit},\mu}), \end{cases}$$

where

$$v_{\mathrm{lim},\mu,1}(k) = \min(v_{\mathrm{control},\mu,1}(k), v_{\mu,1}(k))$$

is the speed that limits the flow, $q_{\mathrm{speed},\mu}(k)$ is the flow that corresponds to the limiting speed $v_{\mathrm{lim},\mu,1}(k)$ in congested branch of the fundamental diagram, and is given by

$$q_{\mathrm{speed},\mu}(k) = \lambda_\mu v_{\mathrm{lim},\mu,1}(k) \rho_{\mathrm{crit},\mu} \left[-a_\mu \ln \left(\frac{v_{\mathrm{lim},\mu,1}(k)}{v_{\mathrm{free},\mu}} \right) \right]^{\frac{1}{a_\mu}} ,$$

and

$$q_{\mathrm{cap},\mu} = \lambda_\mu V(\rho_{\mathrm{crit},\mu}) \rho_{\mathrm{crit},\mu}$$

is the capacity flow.

3.3.4 Anticipation constant

The anticipation term in (3.3) is significant when the density difference in consecutive segments is large. This typically occurs at the head and the tail of traffic jams or shock waves. At the head of a congested area the density decreases, so the anticipation term (including the minus sign) is positive, and the value of η will influence how fast congestion in the considered segment will resolve. Therefore, η influences the outflow of the congestion, which directly determines whether or not there is a capacity drop, and the speed of the back-propagation of the head of the congestion.

On the other hand, at the tail of the congestion the anticipation term is negative and η influences the process how fast the average speed drops when there is a congestion ahead. So, at the tail of the congestion η influences how high the density gets in the congested area (if the speed drops faster, the inflow to the congested area will drop faster and less vehicles will accumulate). Therefore, the value of η influences the back-propagation speed of the tail of the congestion.

In the original METANET model it is implicitly assumed that the two processes depend on the same underlying parameter η. There are also theoretical results that suggest that anticipation behavior is different for different traffic states, or different traffic densities. Barlovic *et al.* [8] showed that the introduction of the so called *slow-to-start* rule into the cellular-automaton model results in the reproduction of hysteresis effects (such as the capacity drop and shock waves). The slow-to-start rule describes that vehicles exiting high-density areas have the tendency to accelerate slower than for other densities.

To describe a similar tendency in the METANET model we distinguish between the two situations where the density is spatially increasing or decreasing. The distinction is made by defining two different anticipation constants for the two situations. Simulation experiments confirmed that splitting the anticipation constant results in better reproduction of shock waves and the capacity drop. For these reasons we use the model described by equation (3.12) below in the simulation experiments in Chapters 6 and 7 where shock waves are relevant.

In (3.3) η is a global parameter and has the same value for all segments. However, based on the arguments presented above, here we take different values for $\eta_{m,i}(k)$, depending on whether the downstream density is higher or lower than the density in the actual segment:

$$\eta_{m,i}(k) = \begin{cases} \eta_{\text{high}} & \text{if } \rho_{m,i+1}(k) \geq \rho_{m,i}(k) \\ \eta_{\text{low}} & \text{otherwise.} \end{cases} \tag{3.12}$$

3.3.5 Downstream boundary conditions

For the downstream boundary conditions we mostly assume in this thesis the destination to be congestion-free. We approximate this by defining the virtual downstream density $\rho_{\mu,N_\mu+1}(k)$ of link μ to be always smaller or equal to $\rho_{\text{crit},\mu}$. If the last segment of the incoming link is congested, the virtual downstream density is taken to be $\rho_{\text{crit},\mu}$, if it is uncongested the virtual downstream density is taken to be equal to the density of the last segment of the incoming link:

$$\rho_{\mu,N_\mu+1}(k) = \min\left(\rho_{\mu,N_\mu}(k),\ \rho_{\text{crit},\mu}\right). \tag{3.13}$$

If, in addition, a downstream density scenario $\rho_d(k)$ for destination d is defined we distinguish between the situation when the downstream density is restricting or non-restricting. This distinction is made because in free flow the information propagates downstream (in the travel direction) and in congested flow information propagates upstream (opposite to the travel direction). So, in free flow (i.e., the non-restricting case) the downstream density should not influence the outflow of segment N_μ of link μ, while when the tail of a jam or shock wave is propagating upstream (i.e., the restricting case) the outflow should be restricted by the downstream density scenario.

So, when the downstream density $\rho_d(k)$ is less than the virtual downstream density that would follow from (3.13) then it is considered to be non-restricting and the scenario has no effect. Otherwise the downstream density scenario is restricting and the virtual downstream boundary condition is taken to be equal to the density defined by the scenario:

$$\rho_{\mu,N_\mu+1}(k) = \max\left(\rho_d(k), \min\left(\rho_{\mu,N_\mu}(k),\ \rho_{\text{crit},\mu}\right)\right).$$

3.4 Model calibration

Before a traffic model can be used to predict the evolution of the traffic situation, the model needs to be calibrated and validated. In this section we present an approach to calibrate a traffic flow model, such as METANET.

For the calibration (parameter identification) we construct a nonlinear least-squares minimization problem in which we (numerically) minimize the following function by selecting the model parameters appropriately:

$$\sum_{l=0}^{N_{\text{samp}}} \sum_{(m,i)\in I_{\text{all}}} \big((\hat{q}_{m,i}(l) - \tilde{q}_{m,i}(l)\big)^2 + \xi\big((\hat{v}_{m,i}(l) - \tilde{v}_{m,i}(l)\big)^2 \ ,$$

with N_{samp} is the number of data samples, I_{all} is the set of indexes of all pairs of links and segments, $\tilde{q}_{m,i}(l)$ and $\tilde{v}_{m,i}(l)$ denote the measured flow and speed data, ξ is a tuning weight, and l corresponds to the time instant $t = lT_{\text{samp}}$ where T_{samp} is the sampling time. We choose T_{samp} and T such that $T_{\text{samp}}/T = K$, with $K \in \mathbb{N}$, and we compute the simulated values $\hat{q}_{m,i}(l)$ and $\hat{v}_{m,i}(l)$ as

$$\hat{q}_{m,i}(l) = \frac{1}{K} \sum_{k=Kl}^{K(l+1)-1} q_{m,i}(k)$$

$$\hat{v}_{m,i}(l) = \frac{1}{K} \sum_{k=Kl}^{K(l+1)-1} v_{m,i}(k) \ .$$

In this approach we assume that the measurement error has a Gaussian distribution, which implies that the least square estimate equals the maximum likelihood estimate. We refer the reader for the results of calibration with this method to [118].

After calibration the model should be validated. The calibrated model should reasonably reproduce traffic scenarios that were not used for the calibration.

3.5 Conclusions

In this chapter an overview of existing traffic flow models is given, where the models were classified according to *application area*, *level of detail*, and *deterministic versus stochastic* or *continuous or discrete* process representation. The choice for a traffic model should be based on the efficiency and accuracy of the model and on the typical phenomena that the traffic flow model can or cannot represent.

We have introduced the METANET model, and made the following extensions:

- We have formulated an explicit model for dynamic speed limits, such that the speed limit influences the traffic only if the speed in the unlimited case would be higher than the displayed speed limit.

- We have formulated a model for main-stream metering, which is similar to ramp metering, except that the traffic dynamics upstream the main-stream metering device is not formulated in terms of a queue, but in terms of speed, flow and density.
- We have formulated a model for the main-stream origin which has different dynamics than on-ramps. The main difference is that a main-stream origin has a different lay-out which allows less inflow in congested situation.
- We have separated the anticipation constant into two constants that represent the anticipation behavior at the head and the tail of shock waves. This gives a better reproduction of shock waves and the capacity drop.
- We have added a new formulation for the downstream boundary condition, which expresses free flow downstream conditions except for some upstream propagating shock waves.

The extended METANET model will be used in the simulations in the subsequent chapters.

Chapter 4

Model predictive control and traffic related issues

In this chapter we introduce the control methodology called *model predictive control* (MPC) that is used in this thesis to solve the dynamic traffic control problem. The same methodology is also known under different names such as *model-based predictive control*, and as *receding, rolling,* or *moving horizon control.*

MPC is most suited for systems where prediction is essential. Predictive control has two advantages over non-predictive (sometimes also called "myopic") control:

- The output of the system can be forced to follow a *trajectory* instead of controlling to a set-point. This is useful since for many processes it is not only desired to reach a set-point but also the trajectory along which the set-point is reached is relevant.

- A trade-off can be made between immediate performance and future performance. In some processes (as we will see in Chapter 6, also in traffic) optimal performance can be achieved by a control signal that may reduce the performance now, but offers more gains in the future.

We will start in Section 4.1 with the introduction of the MPC framework. Next, we discuss the guidelines for tuning an MPC controller, and we present the main advantages and disadvantages of the MPC approach. In Section 4.2 we describe the elements of an MPC controller in a traffic setting and we discuss issues specific for applying MPC to traffic systems.

The main contribution of this chapter is the application of the MPC framework to traffic systems, and the discussion of the properties of an MPC controller that makes it suitable for traffic control problems.

This chapter is not meant as a full introduction on MPC, but provides a basic knowledge that is necessary to understand Chapters 6, 7, 8. For more information on MPC we refer the interested reader to [18, 108, 4, 39] and the references therein.

4.1 Model predictive control

4.1.1 Basic principle

In this section we explain the basic principles of general MPC. We will start with an intuitive description of the MPC process, and in the following sections we introduce the necessary notation and give the formal description of MPC.

The MPC process has the following elements:

1. **Prediction.** The future behavior of the system is predicted for a certain time horizon. The prediction is based on:
   - the current state of the system,
   - the expected disturbances[1],
   - the planned control signal.
2. **Performance evaluation.** The performance is evaluated according to a user-specified objective function. This objective function is typically based on:
   - the (evolution of the) states and outputs of the system during the prediction period,
   - planned control signal, since some signals may be more desirable than others (e.g., signals with frequent variations or higher cost may be less desirable).
3. **Optimization.** The controller finds the control signal that optimizes the objective function.
4. **Control action.** The first sample of the optimal control signals is applied to the process. The remaining part of the control signal is recalculated in the *rolling horizon* scheme. In this scheme the optimal control signal is recalculated every controller sample step to take into account the unpredictable disturbances and the prediction error. The controller sampling time is typically many times smaller than the prediction horizon. For each recalculation the start and the end of the prediction horizon is shifted. This is called rolling horizon. The rolling horizon structure has certain advantages that will be explained later in this chapter.

4.1.2 Notation and formal description

In this section we again describe the general MPC process, but first we introduce the necessary notation. The meaning of the used symbols is summarized in Table 4.1.

[1] By disturbances we mean external influences to the process that cannot be controlled or are not controlled. These disturbances can be known or unknown. E.g., the weather is a known disturbance to the traffic process, a phone call that distracts the driver's attention is an unknown disturbance.

symbol	meaning
k	discrete time index for the process model
k_c	discrete time index for the controller
$x(k)$	process (model) state
$\hat{\mathbf{x}}(k)$	$[\hat{x}(k+1\|k)\ \ldots\ \hat{x}(k+MN_\mathrm{p}-1\|k)]$, the predicted states for the simulation steps $\{k,\ldots,k+MN_\mathrm{p}-1\}$ based on knowledge at simulation step k
$d(k)$	disturbance vector at simulation time step k
$\mathbf{d}(k)$	$[d(k)\, d(k+1)\ \ldots\ d(k+MN_\mathrm{p}-1)]$, the disturbance signals for the simulation steps $\{k,\ldots,k+MN_\mathrm{p}-1\}$
$u(k)$	control vector
$\mathbf{u}(k_\mathrm{c})$	$[u(k_\mathrm{c}\|k_\mathrm{c})\, u(k_\mathrm{c}+1\|k_\mathrm{c})\ \ldots\ u(k_\mathrm{c}+N_\mathrm{p}-1\|k_\mathrm{c})]$, the control signal for the controller time steps $\{k_\mathrm{c},\ldots,k_\mathrm{c}+N_\mathrm{p}-1\}$ based on the knowledge at controller step k_c
$\mathbf{u}^*(k_\mathrm{c})$	$[u^*(k_\mathrm{c}\|k_\mathrm{c})\, u^*(k_\mathrm{c}+1\|k_\mathrm{c})\ \ldots\ u^*(k_\mathrm{c}+N_\mathrm{c}-1\|k_\mathrm{c})]$, the control signal that minimizes $J(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$ based on knowledge at controller step k_c
$J(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$	objective function
N_p	prediction horizon length
N_c	control horizon length
$f(x(k),u(k_\mathrm{c}))$	process (model) state update function
$g(x(k),u(k_\mathrm{c}))$	measurement function
$\phi(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$	equality constraint function
$\psi(\hat{\mathbf{x}}(k),\mathbf{u}(k_\mathrm{c}))$	inequality constraint function
$\hat{\mathbf{y}}(k)$	$[\hat{y}(k+1\|k)\ \ldots\ \hat{y}(k+MN_\mathrm{p}-1\|k)]$, the predicted outputs for simulation time steps $\{k,\ldots,k+MN_\mathrm{p}-1\}$ based on knowledge at simulation time step k

Table 4.1: The symbols used for the MPC problem formulation.

As we will explicitly make a difference between the simulation time step T for the process model and for the controller time step length T_c, we will also use two different counters for the process model (k), and the controller (k_c) denoting the time kT and $k_\mathrm{c}T_\mathrm{c}$ respectively. For the sake of simplicity, we assume that T is an integer divisor of T_c:

$$T_\mathrm{c} = MT \ \ ,$$

with M an integer. Note that the model time indexes in the following description are denoted in terms of k, and the control signal time indexes are denoted in terms of k_c. The MPC process is described by:

1. **Prediction.** In Figure 4.1 the predictive part of MPC is shown. The prediction is made at controller time step k_c (for which the corresponding model time index is $k = Mk_\mathrm{c}$). We assume that the process is described by the discrete-time system f:

   $$x(k+1) = f(x(k), u(k_\mathrm{c}), d(k))\,,\ \text{with}\ \ Mk_\mathrm{c} \leq k < (k_\mathrm{c}+1)M \tag{4.1}$$

   where $x(k)$ is the state vector of the system at simulation step k, $u(k_\mathrm{c})$ is the vector of control inputs at controller time step k_c, and $d(k)$ is the disturbance vector at simulation step k.

   The prediction is made by repeatedly applying (4.1) for the simulation time steps $\{k, \ldots, MN_\mathrm{p} - 1\}$, where N_p is the length of the horizon (in units of controller time steps) for which the process behavior is predicted (we assume that the future disturbances are known or can be predicted[2]),

   The inputs for the model-based prediction are the current state of the process $x(k)$, the expected disturbances

   $$\mathbf{d}(k) = [d(k)\, d(k+1)\, \ldots\, d(k+MN_\mathrm{p}-1)]\,,$$

   and the control signal matrix

   $$\mathbf{u}(k_\mathrm{c}) = [u(k_\mathrm{c}|k_\mathrm{c})\, u(k_\mathrm{c}+1|k_\mathrm{c})\, \ldots\, u(k+N_\mathrm{p}-1|k_\mathrm{c})],$$

   where $u(l_\mathrm{c}|k_\mathrm{c})$ denotes the computed control signal for controller time step l_c based on information available at controller time step k_c. Based on $x(k)$, $\mathbf{d}(k)$, and $\mathbf{u}(k_\mathrm{c})$ the future evolution of the process is predicted and is denoted by the matrix

   $$\hat{\mathbf{x}}(k) = [\hat{x}(k+1|k)\, \ldots\, \hat{x}(k+MN_\mathrm{p}-1)].$$

[2] This assumption is realistic if the traffic state can be measured on-line on a sufficiently large area upstream and downstream from the controlled network. In addition historical data may be used.

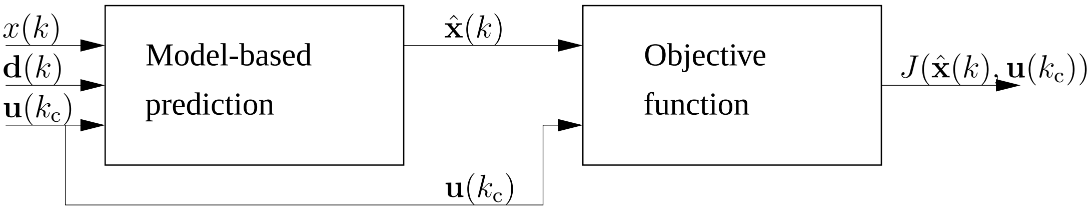

$x(k)$
$\mathbf{d}(k)$
$\mathbf{u}(k_\mathrm{c})$
Model-based prediction
$\hat{\mathbf{x}}(k)$
Objective function
$J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$
$\mathbf{u}(k_\mathrm{c})$

Figure 4.1: Model-based prediction and objective function evaluation.

2. **Performance evaluation.** The objective function $J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$ is evaluated based on the prediction $\hat{\mathbf{x}}(k)$ and the control signal $\mathbf{u}(k_\mathrm{c})$. This function is chosen such that it expresses the performance of the process as a real scalar: the smaller $J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$, the better the performance.

3. **Optimization** The goal of the controller is to find the control signal $\mathbf{u}(k_\mathrm{c})$ that minimizes $J(\hat{\mathbf{x}}(k), \hat{\mathbf{u}}(k_\mathrm{c}))$ for a given initial state $x(k)$ and disturbance $\mathbf{d}(k)$. The control inputs are only optimized for the so called control horizon $N_\mathrm{c}(\leq N_\mathrm{p})$, which is also expressed in controller time steps. The control signal after $k_\mathrm{c} + N_\mathrm{c} - 1$ is assumed to be constant:

$$u(k_\mathrm{c} + l_\mathrm{c}|k_\mathrm{c}) = u(k_\mathrm{c} + N_\mathrm{c} - 1|k_\mathrm{c}) \text{ for } l_\mathrm{c} \in \{N_\mathrm{c}, \ldots, N_\mathrm{p} - 1\}\,. \tag{4.2}$$

   In Figure 4.2 the model-based prediction and objective function evaluation block are connected with the optimization block that finds that control signal

$$\mathbf{u}^*(k_\mathrm{c}) = [u^*(k_\mathrm{c}|k_\mathrm{c})\, u^*(k_\mathrm{c}+1|k_\mathrm{c}) \,\ldots\, u^*(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c})]$$

   that minimizes $J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$ for

$$[u(k_\mathrm{c}|k_\mathrm{c})\, u(k_\mathrm{c}+1|k_\mathrm{c}) \,\ldots\, u(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c})] = u^*(k_\mathrm{c}|k_\mathrm{c})$$

   and constant values of $u(l_\mathrm{c}|k_\mathrm{c})$ for $l_\mathrm{c} \in \{N_\mathrm{c}, \ldots, N_\mathrm{p} - 1\}$ as defined in (4.2).

   The optimization block contains a suitable optimization algorithm, that depends on the process model and objective function.

4. **Control action.** Only the first sample (first column) $u^*(k_\mathrm{c}|k_\mathrm{c})$ of the optimal control signal $\mathbf{u}^*(k_\mathrm{c})$ is applied to the process. In the rolling horizon scheme the procedure from prediction to control action is repeated at controller time step $k_\mathrm{c} + 1$ with the prediction horizon shifted one time step ahead, and so on (see Figure 4.3 for the full MPC scheme).

Most systems cannot (physically) accept any control input, or cannot be in any state. Hence, in the practice of industrial systems specific signals must not violate specified

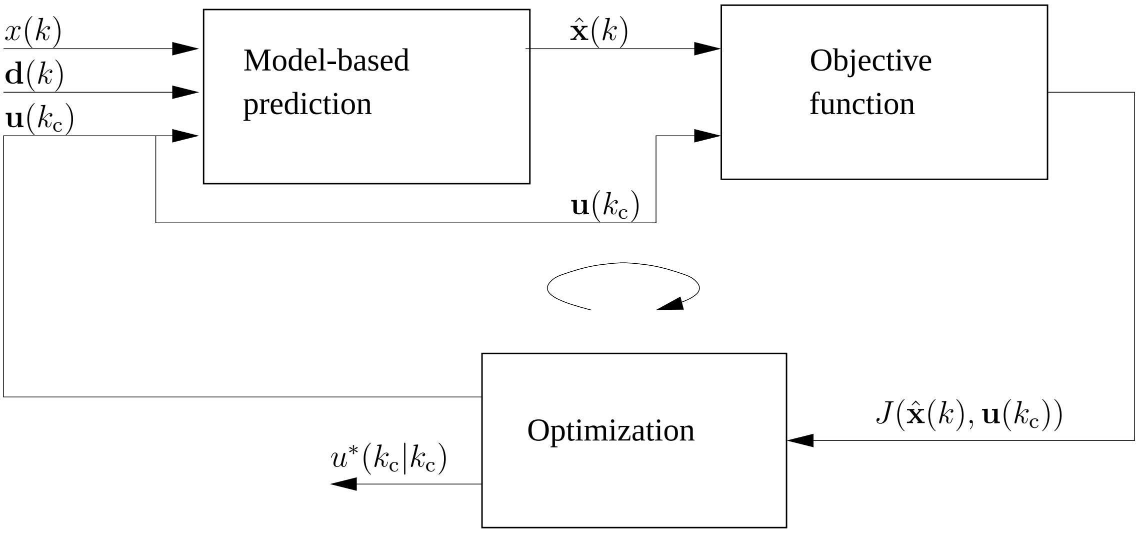

$x(k)$
$\mathbf{d}(k)$
$\mathbf{u}(k_\mathrm{c})$
Model-based prediction
$\hat{\mathbf{x}}(k)$
Objective function
$\mathbf{u}(k_\mathrm{c})$
Optimization
$J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$
$u^*(k_\mathrm{c}|k_\mathrm{c})$

Figure 4.2: Optimization of the control signals $\mathbf{u}(k_\mathrm{c})$.

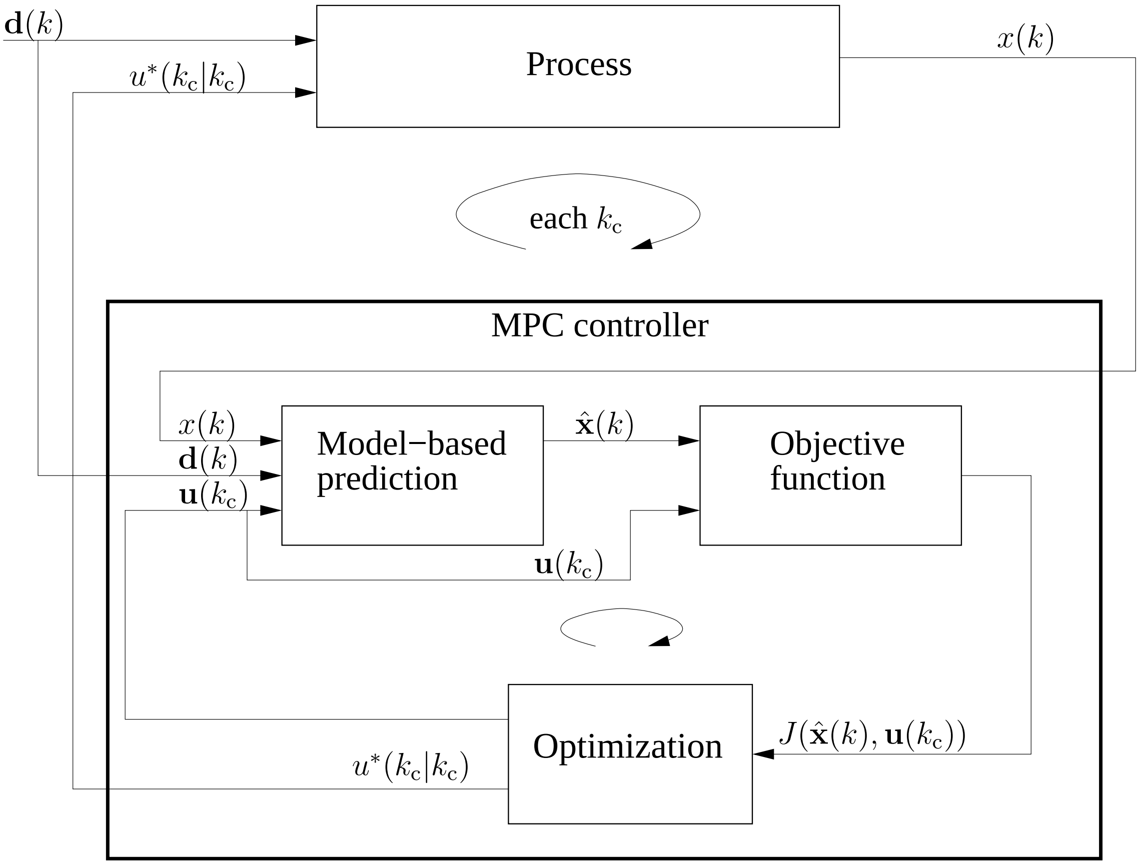

$\mathbf{d}(k)$
$u^*(k_\mathrm{c}|k_\mathrm{c})$
Process
$x(k)$
each k_c
MPC controller
$x(k)$
$\mathbf{d}(k)$
$\mathbf{u}(k_\mathrm{c})$
Model−based prediction
$\hat{\mathbf{x}}(k)$
Objective function
$\mathbf{u}(k_\mathrm{c})$
Optimization
$J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$
$u^*(k_\mathrm{c}|k_\mathrm{c})$

Figure 4.3: The full MPC scheme.

bounds due to safety limitations, environmental regulations, consumer specifications and physical restrictions such as minimum and/or maximum temperature, pressure, level limits in reactor tanks. These limitations determine the admissible states and admissible control inputs of the process. The admissible states and control inputs are described by the equality constraint vector $\phi(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c})) = 0$ and the inequality constraint vector $\psi(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c})) \leq 0$. We can now formulate the MPC control problem as follows.

MPC problem:
Given the discrete-time system f with

$$x(k+1) = f(x(k), u(k_\mathrm{c}), d(k)) \,, \text{ with } \; Mk_\mathrm{c} \leq k < (k_\mathrm{c}+1)M$$

the known initial state $x(k^*)$, the known disturbance $\mathbf{d}(k^*)$, and $k_\mathrm{c}^* = \lfloor k^*/M \rfloor$, (where $\lfloor z \rfloor$ means the largest integer that is smaller or equal to z), and the objective function

$$J(\hat{\mathbf{x}}(k^*), \mathbf{u}(k_\mathrm{c}^*))$$

with the equality constraint

$$\phi(\hat{\mathbf{x}}(k^*), \mathbf{u}(k_\mathrm{c}^*)) = 0 \,,$$

and the inequality constraint

$$\psi(\hat{\mathbf{x}}(k^*), \mathbf{u}(k_\mathrm{c}^*)) \leq 0 \,,$$

with the meaning of the symbols as given in Table 4.1, find for each k_c^* the control signal $\mathbf{u}^*(k_\mathrm{c})$ that minimizes

$$J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$$

for

$$[u(k_\mathrm{c}|k_\mathrm{c})\, u(k_\mathrm{c}+1|k_\mathrm{c}) \; \ldots \; u(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c})] = \mathbf{u}^*(k_\mathrm{c})$$

and

$$u(l|k_\mathrm{c}) = u^*(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c}) \text{ for } l \in \{N_\mathrm{c}, \ldots, N_\mathrm{p}-1\}.$$

Remark 4.1.1 We assume that the disturbance vector is fully known. If this is not the case, in some the disturbance signal can be split as a sum of a known part $w(k)$ and an unknown part $v(k)$, thus $d(k) = w(k) + v(k)$. The knowledge of the know part $v(k)$ can be fully utilized in the same way as above. □

Remark 4.1.2 We also assume that the process state is fully measurable. If this is not the case, an observer is needed to estimate the state of the process (assuming that the state is observable). This process state is needed for each k_c^* as the initial state of the prediction. □

4.1.3 Tuning

In the MPC problem formulation above, there are basically three items to tune: the prediction horizon N_p, the control horizon N_c, and the parameters (e.g., weights of subobjectives, as given in (4.7).) of the objective function $J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$. In conventional MPC heuristic tuning rules have been developed to select appropriate values for N_p and N_c (for linear systems, see [108] and [151]). However, these rules cannot be straightforwardly applied to a general non-linear framework presented above. In any case, the prediction horizon N_p should be long enough to include all important process dynamics. However if N_p is chosen too long, it may increase computational complexity unnecessarily. For N_c a trade-off has to be made between low computational complexity (i.e., short control horizon) and more control freedom (i.e., long control horizon). Too little control freedom may result in insufficient control actions, but too much control freedom may result in a "wilder" control signal, which is undesired in general.

A possible drawback of a short control horizon in combination with a long prediction horizon is that the controller may react too early. In the period after the control horizon has passed, the control signal is generally assumed to be constant (cf. (4.2)). This means that the last value of $\mathbf{u}^*$ determines a relatively large part of the performance function. When a disturbance occurs at the end of the prediction horizon, this may determine the value of the vector $u^*(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c})$ and in combination with a control variation penalty term (or a boundary on the rate of change of the control signal) this may even influence the control value $u^*(k_\mathrm{c}|k_\mathrm{c})$ that is applied to the process. In other words the controller reacts immediately to a disturbance at the end of the prediction horizon, which may be too early and therefore undesired.

Tuning the parameters of the objective function is a matter of describing the desired behavior of the controller (and attaching weights to subgoals if present), such as reference tracking, penalizing control effort, or control variations.

4.1.4 Advantages and disadvantages

In this section we present the main advantages and disadvantages of MPC.

Advantages

1. **Feedback.** Feedback (as a part of the rolling horizon scheme) significantly reduces the adverse effects of unpredictable disturbances. For every k_c the "real"[3] state of the process is taken as the initial state for the prediction. In this way the real effects of previous control actions are fed back into the controller. This is useful when the process model is not exactly equal to the real process (i.e., when we have a model mismatch) or when the disturbance is not fully known. In both cases the prediction of the next process state will be imprecise and is corrected by updating the state from the real process. Therefore feedback reduces the sensitivity to model mismatch and to prediction errors.

2. **Easy tuning.** There are only a few parameters that need tuning: the prediction horizon N_p, the control horizon N_c, and the parameters (weights) of the objective function $J(\hat{\mathbf{x}}(k), \mathbf{u}(k_\mathrm{c}))$. In general, finding appropriate values for N_p and N_c is straightforward. The selection of the parameter (weights) of the obejctive function may recquire a trial-and-error procedure.

3. **Prediction.** As mentioned in the introduction of this chapter, a predictive controller can use the dynamics of the process, and the (partial) knowledge of the future disturbances.

4. **Multiple inputs.** Multiple control inputs are handled by MPC exactly the same way as single inputs, no theoretical extension is necessary.

5. **Easy constraint formulation.** In practical control problems there are often constraints present for the maximum or minimum value of the control signals, for the control signal rate of change, or for the state of the process. All these constraints are easily formulated as constraints for the optimization problem.

6. **Modularity.** The controller has a modular structure: the process model, the objective function, the constraints, and the optimization algorithm can be replaced independently.

7. **Adaptivity.** Due to the modular structure it is even possible to update the model for each iteration. Updating the model may be necessary when the process behavior changes significantly and the controller is required to adapt to the changes. Typically, the parameter updating process is slower than the feedback process. We refer to [10] for an analysis of an adaptive ramp metering system.

[3] See also remark 4.1.2. If the states are not directly measurable an observer may be necessary. For traffic systems this is typically the case.

Disadvantages and solutions

1. **Complexity.** As for any optimal control methodology the computational complexity can become too high if the control horizon is long or if the number of inputs is large. We mention here some techniques to reduce complexity [108]:

   - **Reduce N_c.** As mentioned in Section 4.1.3 a shorter N_c reduces the computational complexity, but may result in insufficient control actions.
   - **Increase the controller sampling time.** If M is increased a shorter control horizon will be sufficient to cover the same time period. A shorter control horizon means less control variables (to optimize) but also less control freedom, which may worsen the performance. Another, but similar option is to use irregular sampling, where the optimization problem is re-parametrized into a problem with less optimization variables.
   - **Choose a good starting point.** Using a good first guess for the optimization problem can also reduce the computation time to find an optimum. When the disturbance signal does not vary too much, a good guess for the new control signal $\mathbf{u}^*(k_\mathrm{c}+1)$ is the shifted version — shifted by one sample — of that part of the control signal $\mathbf{u}^*(k_\mathrm{c})$ that was not applied to the process. So, let $[u(k_\mathrm{c}+1|k_\mathrm{c}+1),\ldots,u(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c}+1)] = [u^*(k_\mathrm{c}+1|k_\mathrm{c}),\ldots,u^*(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c})]$. For the last column $u(k_\mathrm{c}+N_\mathrm{c}|k_\mathrm{c}+1)$ a new initial guess has to be chosen, e.g., $u(k_\mathrm{c}+N_\mathrm{c}|k_\mathrm{c}+1) = u(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c}+1)$ or a random value.
     What should be avoided is the selection of a starting point for which the gradient of the objective function is zero, while it is known that the resulting performance is not optimal. E.g., choosing an initial ramp metering rate that is higher than the traffic demand will not influence the traffic nor the TTS. At this starting point it is difficult for the optimization algorithm to find the direction (higher or lower metering rate) that brings the objective function closer to an optimum. In this case a better choice is to take the minimum metering rate as an initial guess.

2. **Precise model and precise disturbance prediction are necessary.** Since the control action is based on the predicted consequences of the planned control action, the prediction must be as accurate as possible. This accuracy is influenced by the accuracy of the prediction model and the predicted disturbances. The accuracy of the prediction model and the disturbance signal determines the length of the prediction horizon for which the prediction is meaningful. As mentioned above, feedback (as a part of the rolling horizon scheme) may significantly reduce the adverse effects of unpredictable disturbances and model mismatch.

3. **Stability is difficult to prove.** While it is often easy to achieve stability in practice by tuning, it is difficult to prove mathematically that the closed-loop system will be

stable.

4. **Optimality is not guaranteed.** For an arbitrary nonlinear system and nonlinear objective function, the optimization problem can become non-convex. This means that no guarantee can be given that the solution $\mathbf{u}^*(k_c)$ globally minimizes $J(\hat{\mathbf{x}}(k), \mathbf{u}(k_c))$.

   Since most numerical optimization algorithms take some initial guess as a starting point, we mention here two methods that make use of multiple initial guesses and increase the probability that the global optimum is found:

   - **Multi-start optimization.** We could generate multiple initial guesses for the optimization algorithm. If these result in different solutions, we should preserve the one with the best performance.
   - **Best M of N starting points.** We could generate M (many) initial guesses, and evaluate the performance in these points (without optimization). Next, we select the N (with $N << M$) best points — these are considered to be the most promising candidates — and we run a full optimization with these starting points.

   Both methods make use of initial guesses, which may be random or non-random. The non-random guesses may be a shifted version of the previous optimization step.

   Both methods further increase complexity, thus a trade-off between optimality and computation time has to be made.

4.1.5 Sequential quadratic programming

In this section we present the sequential quadratic programming (SQP) optimization method that can be used to solve optimization problems as formulated in the MPC problem. SQP is a powerful method for solving nonlinearly constrained, continuous optimization problems, with a smooth objective function. We will use SQP in combination with a multi-start approach to solve the MPC problem.

MPC has been used to solve a large set of important practical problems [12]. The name SQP does not refer to a single algorithm, but rather to a conceptual method from which numerous specific algorithms have evolved.

Kotsialos *et al.* [92, 91, 93, 94] use a nonlinear optimal control approach to solve similar problems. The optimality conditions that have to be fulfilled for this approach are given in [91]. A similarity between the SQP approach and the optimal control approach is that for none of the methods it can be guaranteed that it reaches the global optimum. However, Kotsialos *et al.* mention that for repeated experiments similar performances are achieved. The same conclusion can be drawn from the case-studies in Chapters 6–8, where SQP will be used. Further comparison of the methods is a topic for future research.

In this section we continue with presenting the SQP framework without going into the details of a specific implementation. For more details we refer the interested reader to [117, 12, 138].

Main problem

The SQP framework aims at solving the nonlinear programming problem

$$\begin{aligned} \min_{\tilde{u}} \;& J(\tilde{u}) \\ \text{subject to} \;& \tilde{\phi}(\tilde{u}) = 0\,, \\ & \tilde{\psi}(\tilde{u}) \leq 0\,, \end{aligned} \tag{4.3}$$

where $\tilde{u}$ is the vector of optimization variables with $J(\tilde{u})$ the objective function, $\tilde{\phi}(\tilde{u})$ the vector of nonlinear equality constraints, and $\tilde{\psi}(\tilde{u})$ the vector of nonlinear inequality constraints.

The MPC problem formulated in this form reads: Solve

$$\begin{aligned} \min_{[u(k_\mathrm{c}|k_\mathrm{c})\, u(k_\mathrm{c}+1|k_\mathrm{c}) \ldots u(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c})]} \;& J(\hat{\mathbf{x}}(k^*), \mathbf{u}(k_\mathrm{c}^*)) \\ \text{subject to} \;& \phi(\hat{\mathbf{x}}(k^*), \mathbf{u}(k_\mathrm{c}^*)) = 0\,, \\ & \psi(\hat{\mathbf{x}}(k^*), \mathbf{u}(k_\mathrm{c}^*)) \leq 0\,, \end{aligned}$$

given the system equation

$$x(k+1) = f(x(k), u(k_\mathrm{c}), d(k))\,,\ \text{with}\ \ Mk_\mathrm{c} \leq k < (k_\mathrm{c}+1)M$$

and the end point constraint

$$u(l|k_\mathrm{c}) = u^*(k_\mathrm{c} + N_\mathrm{c} - 1|k_\mathrm{c}) \text{ for } l \in \{N_\mathrm{c}, \ldots, N_\mathrm{p} - 1\},$$

and given the known initial state $x(k^*)$, the known disturbance $\mathbf{d}(k^*)$, and $k_\mathrm{c}^* = \lfloor k^*/M \rfloor$. In other words, the dynamic MPC problem is formulated as a static optimization problem in the parameters

$$[u(k_\mathrm{c}|k_\mathrm{c})\, u(k_\mathrm{c}+1|k_\mathrm{c})\ \ldots\ u(k_\mathrm{c}+N_\mathrm{c}-1|k_\mathrm{c})].$$

We now continue with the notation of 4.3 for the sake of readability.

In SQP this optimization problem is solved by translating it to a sequence of quadratic subproblems with linear equality and inequality constraints, for which efficient solution algorithms exist. The sequence of quadratic subproblems is solved iteratively, and for each iteration a new subproblem is formulated.

Quadratic subproblem

Before we present the subproblem we introduce the notation for the *Lagrangian* of the main problem and the *Hessian* of the Lagrangian. The Lagrangian $\mathcal{L}(\tilde{u}, \tilde{v}, \tilde{w})$ of the main problem is defined as

$$\mathcal{L}(\tilde{u}, \tilde{v}, \tilde{w}) = J(\tilde{u}) + \tilde{v}^T \tilde{\phi}(\tilde{u}) + \tilde{w}^T \tilde{\psi}(\tilde{u}) \,,$$

where $\tilde{v}$ and $\tilde{w}$ are the *multiplier vectors* with appropriate lengths. We denote the Hessian of the Lagrangian by the matrix $W(\tilde{u}, \tilde{v}, \tilde{w})$:

$$W(\tilde{u}, \tilde{v}, \tilde{w}) = \nabla^2_{\tilde{u}\tilde{u}} \mathcal{L}(\tilde{u}, \tilde{v}, \tilde{w}) \,,$$

where $\nabla^2_{\tilde{u}\tilde{u}} \mathcal{L}(\tilde{u}, \tilde{v}, \tilde{w})$ denotes the matrix of second derivatives with respect to the vector $\tilde{u}$.

The subproblem to be solved in each iteration step ℓ is defined in terms of the auxiliary variable p_ℓ is an auxiliary variable in the ℓ-th subproblem:

$$\min_p \frac{1}{2} p_\ell^T W_\ell p_\ell + \nabla J_\ell^T p_\ell \tag{4.4}$$

$$\text{subject to } \nabla \tilde{\phi}(\tilde{u}_\ell)^T p_\ell + \tilde{\phi}(\tilde{u}_\ell) = 0 \,, \tag{4.5}$$

$$\nabla \tilde{\psi}(\tilde{u}_\ell)^T p_\ell + \tilde{\psi}(\tilde{u}_\ell) \leq 0 \,, \tag{4.6}$$

where the subscript ℓ denotes the variables in the ℓ-th iteration, i.e., $\tilde{u}_\ell$ is the solution of the main problem after ℓ iterations, J_ℓ is a shorthand notation for $J(\tilde{u}_\ell)$, and W_ℓ is the shorthand notation for $W(\tilde{u}_\ell, \tilde{v}_\ell, \tilde{w}_\ell)$.

Basic SQP algorithm

The basic SQP algorithm can now be formulated as follows:

```
Choose an initial tuple ($\tilde{u}_0$, $\tilde{v}_0$, $\tilde{w}_0$);

for    $\ell$ = 0, 1, 2, ...
       Evaluate $\nabla J_\ell$ and $W_\ell$;
       Solve the quadratic subproblem (4.4)–(4.6) to obtain $p_\ell$ and
           the multipliers $\tilde{u}_{\ell+1}$, $\tilde{v}_{\ell+1}$, $\tilde{w}_{\ell+1}$;
       Update the optimization variable $\tilde{u}_{\ell+1} = \tilde{u}_\ell + p_\ell$;
       Stop if converged;
end
```

4.2 Traffic related issues

In this section we consider issues that are related to the application of MPC to traffic systems. First, we describe the MPC controller in a traffic setting. The main reason to use MPC for traffic control is that it possesses all necessary properties listed in Chapter 1:

- it coordinates the control measures,
- it predicts the effect of the control measures,
- it has a feedback structure,
- it optimizes the performance,
- it can handle constraints.

In the following sections we will discuss these properties in more detail.

4.2.1 MPC control for traffic

In Figure 4.4 the MPC scheme for a traffic setting is shown. The disturbances, i.e., the uncontrollable inputs of the traffic system, are among others:

- the traffic demand at the origins of the network, if present, including origin destination information. The traffic demand is a disturbance in the sense that the controller has to react to (future) changes in the demand and to their consequences for the traffic network.
- the shock waves or congestion entering at the destinations of the network. When a destination is congested, the information propagates upstream into the network.
- the weather conditions that can influence the traffic anywhere in the network.

Depending on the model that is used, the state of the traffic process is represented by speed, flow and density in the various links and segments of the network, or by some other variables. Finally, the control signal consists of the control values of the traffic control measures, such as ramp metering rates, main-stream metering rates, speed limits, or route guidance signals.

4.2.2 Coordination and prediction

As already mentioned in the introduction of this thesis two ingredients that are necessary for successful network control are coordination and prediction. In MPC both ingredients are present, since prediction is an inherent property of MPC and coordination is equivalent to using multiple control input signals.

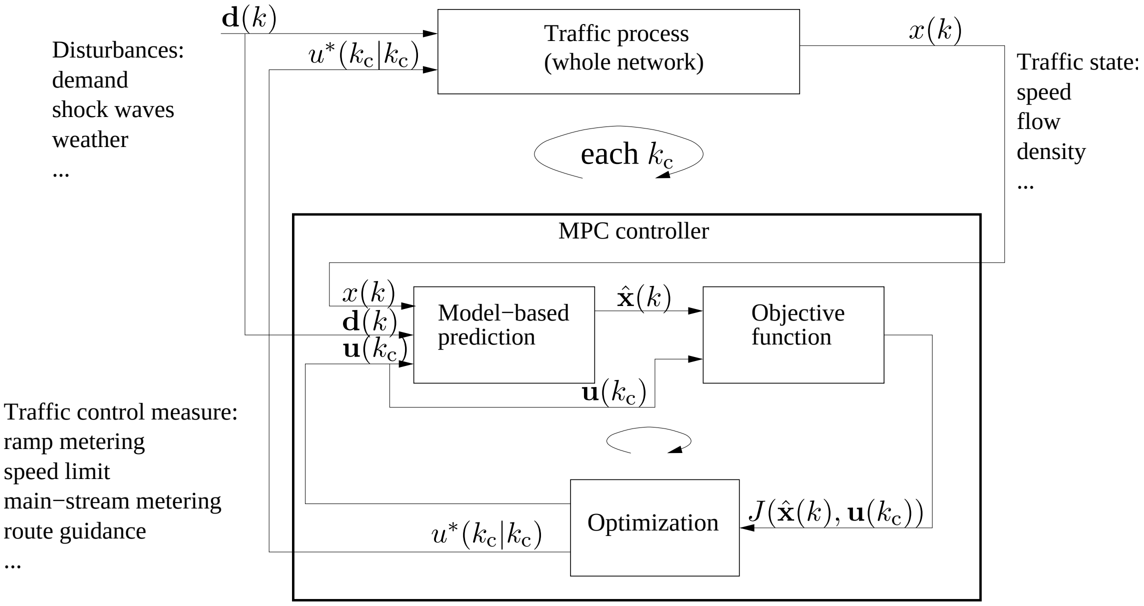


Figure 4.4: The MPC scheme for traffic control.

The first reason why prediction is necessary is that control measures also have effect on more distant parts of the network, and this effect takes place with a certain time delay. Another reason why prediction is necessary is that in traffic systems to eventually improve the flow, the flow often has to be limited first. That means that initially the performance may decrease to achieve a better performance later. To be able to find the control signal that improves the performance later, but possibly reduces the performance now, it is necessary to predict until at least the moment that the improved performance is achieved. A typical example of this is restricting a traffic flow temporarily in order to prevent a gridlock that could block traffic for a long time. Other examples are given in Chapter 6 where congestion is resolved by limiting the inflow to the congested area. After the congestion has been resolved the capacity flow can be reached again (as opposed to the outflow of a congested area, which is lower than capacity). Note that an examination of the practical applicability of predictive control can be found in [158].

4.2.3 Feedback

In general, feedback is used to suppress uncertainty (e.g., model mismatch or unpredictable disturbances) and hence to improve the behavior of the controlled system. Feedback in (predictive) traffic control systems is necessary, because even small unpredictable disturbances (such as a demand misprediction or model mismatch) can have serious impact on the TTS. As we know from Chapter 1 the TTS has a high sensitivity to outflow variations, and by feedback the adverse effect of the disturbances is minimized.

4.2.4 Problem formulation

Traffic control problems are in general easily formulated in the MPC framework, since most elements follow naturally from the traffic problem. In this section we consider the formulation of the elements of the MPC controller for traffic control problems.

Model

Most computer implementable traffic models are discrete time[4], therefore they are directly usable in the MPC framework.

Objective function

The main objective of traffic control is improving network performance. However, network performance can be interpreted in many ways, and for every interpretation a different objective function can be formulated. The most frequently used objective is to minimize the total time that all vehicles spend in the network (i.e., the TTS), because it is directly related to the average travel time for all vehicles and minimizing travel time common goal. Another advantage of the TTS is that it can easily be calculated for macroscopic models. The value of the TTS can be expressed between the controller steps k_c and $k_\mathrm{c} + N_\mathrm{p}$ as:

$$J_\mathrm{TTS}(k_\mathrm{c}) = \sum_{k=Mk_\mathrm{c}}^{M(k_\mathrm{c}+N_\mathrm{p})-1} T \left\{ \sum_{(m,i)\in I_\mathrm{all}} \rho_{m,i}(k) L_m \lambda_m + \sum_{o\in O_\mathrm{all}} w_o(k) \right\},$$

where T is the length of the simulation time step, $\rho_{m,i}(k)$ is the density in segment i of link m at simulation time step k, L_m the length of a segment in link m, λ_m the number of lanes in link m, $w_o(k)$ is the number of vehicles queuing at on-ramp o, I_all is the set of pairs of indexes (m,i) of all links and segments in the network, and O_all is the set of indexes of all on-ramps.

Other objective functions have been proposed for a freeway stretch without ramp queues in [2] (in an optimal control setting), such as maximizing the sum of the traffic flows going through all sections:

$$J_\mathrm{flow}(k_\mathrm{c}) = \sum_{k=Mk_\mathrm{c}}^{M(k_\mathrm{c}+N_\mathrm{p})-1} \sum_{(m,i)\in I_\mathrm{all}} \rho_{m,i}(k) v_{m,i}(k),$$

where $v_{m,i}(k)$ is the speed of link m segment i at simulation step k. Furthermore, by noting that less congestion means lower density, in [2] a cost function is proposed that

[4] Even if the original model is continuous, for computer simulation in practice it is always discretized by some numerical scheme.

equalizes density over time and the segments:

$$J_{\text{density}}(k_{\text{c}}) = \sum_{k=Mk_{\text{c}}}^{M(k_{\text{c}}+N_{\text{p}})-1} \sum_{(m,i)\in I_{\text{all}}} \rho^2_{m,i}(k)\,,$$

where a synthetic freeway is considered without on-ramp and off-ramps.

Another term that is often used in the objective function (in industry as well as in traffic) is a term that equalizes and penalizes control signal variations,

$$J_{\text{contr\_var}}(k_{\text{c}}) = \sum_{\ell=k_{\text{c}}+1}^{k_{\text{c}}+N_{\text{c}}-1} \|u(\ell|k_{\text{c}}) - u(\ell-1|k_{\text{c}})\|^2 + \|u(k_{\text{c}}|k_{\text{c}}) - u(k_{\text{c}}-1|k_{\text{c}}-1)\|^2 \ ,$$

where the first term penalizes the variations in the control signal that is being optimized, and the second term penalizes the difference of the next control vector with the last applied control vector. These terms may be necessary to prevent that the controller produces completely different control signal at each controller time step. The second term also introduces memory in the control system since, the last applied control vector is also taken into account in the optimization. If necessary, more elements of $\mathbf{u}^*(k_{\text{c}}-1|k_{\text{c}}-1)$ can be include in the objective function to, e.g., penalize deviations from the previously planned control signals.

In Chapter 7 we coordinate route guidance and ramp metering. To ensure a certain reliability of the travel time shown on the route guidance information panel we include a term that expresses the difference between the displayed travel times and the realized travel times. It is important to keep the difference small because otherwise drivers may lose confidence in the route guidance system and the control effect of the route guidance may be lost. The objective function that we use is of the form

$$J_{\text{pred\_err}}(k_{\text{c}}) = \sum_{\zeta\in\mathcal{V}(k_{\text{c}})} \sum_{\omega\in\mathcal{D}(\zeta)} (\vartheta_{\text{pred}}(\zeta,\omega)) - \vartheta_{\text{real}}(\zeta,\omega))^2$$

where $\mathcal{V}(k_{\text{c}})$ is the set of indexes of all vehicles that left the network in the period $[k_{\text{c}}, k_{\text{c}}+N_{\text{p}}-1)$, $\mathcal{D}(\zeta)$ is the set of indexes of DRIPs that vehicle ζ has encountered, $\vartheta_{\text{pred}}(\zeta,\omega)$ is the travel time shown on the DRIPs ω for vehicle ζ, and $\vartheta_{\text{real}}(\zeta,\omega)$ is the actually realized travel time for vehicle ζ from DRIP ω to its destination.

Furthermore, other options for terms in the objective function are terms that express the cost of environmental effects of traffic, such as air or noise pollution, or terms that express safety (cf. Section 1.1.3).

Combining objectives

For a traffic problem there are often several objectives present, that may be conflicting. E.g., minimizing noise pollution and maximizing flow is typically conflicting. In such a case a trade-off has to be made between the partial objective functions $J_i(k_\mathrm{c})$, which can be expressed by combining the objective functions into one objective function

$$J_\mathrm{total}(k_\mathrm{c}) = \sum_i \xi_i J_i(k_\mathrm{c}) \ , \tag{4.7}$$

where ξ_i are appropriately chosen weights to express the trade-off between the several partial objective functions.

Constraints

Constraints on the control signal are often present in traffic. In all cases the control signal is bounded by a minimum and a maximum value, such as minimum/maximum green times, metering rates or speed limits. But also other constraints may be present such as the constraint that ensures safe operation of the speed limits by bounding the maximum speed drop that drivers can encounter.

Also constraints on the traffic state can occur, e.g., the constraint that limits the on-ramp queue (which can be used to prevent queue spill-back to urban intersections).

It must be noted that putting constraints on the traffic state could result in infeasible problems. In these cases alternatively an extra term can be included in the objective function that penalizes the violation of the constraint (i.e., constraint relaxation). A disadvantage of this second approach is that the weight of this extra term also has to be tuned, and that the constraints may be violated by the resulting controller.

4.2.5 Tuning

The tuning rules to select appropriate values for N_p and N_c that have been developed for conventional MPC cannot be straightforwardly applied to the traffic flow control framework presented above. However, based on heuristic reasoning we can determine an appropriate initial guess for these parameters.

The main rule for tuning N_p is that the prediction interval should be long enough to include all important process dynamics. The following reasoning is based on the assumption that the objective is to minimize TTS. Since TTS is strongly related with the outflow of the network, N_p should be larger than the typical travel time from the controlled segments to the exit of the network, because if we take the prediction horizon N_p shorter than the typical travel time in the network, the effect of the vehicles that are influenced by the current control measure and — as a consequence — have an effect on the network performance before they exit the network, will not be taken into account. Furthermore, a control action may affect the network state (by improved flows, etc.) even when the actually affected vehicles have already exited the network. On the other hand, N_p should not be too large because of the computational complexity of the MPC optimization problem.

For the control horizon N_c we will select a value that represents a trade-off between the computational effort and the performance. In Chapter 6 we will present benchmark problems where we also discuss the relation between efficiency and the choice of N_p and N_c.

The tuning of ξ_i is easy when the trade-off between the different terms in the objective function can be made explicit. E.g., for some industrial processes the trade-off between performance (e.g., profit) and control signal variations (e.g., energy costs) can be made on a financial basis. In traffic it is also often required not to have “too large” control signal variations but the meaning of “too large” is not defined explicitly. In this case tuning of the corresponding weight is performed by trial-and-error.

Remark 4.2.1 The tuning of ξ_i may be dependent on the choice of N_p and N_c, or on the choice of the controlled network and control measures. E.g., if in the traffic network more uncontrolled links and segments are included, a term expressing the TTS will increase, but a term expressing the control signal variations will remain the same, and therefore the second term will have relatively less weight in the objective function. □

4.3 Conclusions

We have presented an intuitive and a formal description of the MPC framework. Guidelines for tuning the prediction and the control horizon were discussed. In particular, the length of the prediction horizon is a trade-off between complexity and the requirement that it should be long enough to reproduce all important process dynamics. The length of the control horizon is a trade-off between complexity and performance.

We have also discussed the advantages (feedback, easy tuning, prediction, multiple inputs, easy constraint formulation, modularity, and adaptivity) and the disadvantages (complexity, precise model and disturbance prediction are necessary, stability is difficult to prove, optimality is not guaranteed) and some solutions to these disadvantages.

The main reason why we have considered the MPC framework is that it is very well suited to solve traffic control problems. The MPC controller has all important properties that are necessary for traffic control:

- coordination,
- prediction,
- feedback structure,
- optimality according to a user-definable performance function,

- constraint handling.

In the following chapters we apply the MPC methodology presented in this chapter, to several benchmark problems for dynamic traffic control. We will consider several combinations of ramp metering, speed limits, and route guidance in the benchmark problems, and demonstrate the effectiveness of this approach.

Chapter 5

Necessary conditions for successful traffic control

Traffic control is not a panacea for traffic problems. Therefore, the conditions for successful traffic control should be identified carefully. These conditions include the choice of the *boundaries* of the controlled network, the typical *traffic scenarios* for which control is expected to improve the performance, the phenomena that should be present in traffic and that should be represented by the prediction model in case of model-based control.

In this chapter first we present general conditions applicable to all kinds of freeway traffic control in Section 5.1. Next, we present some specific conditions in more detail for speed limit control in Section 5.2 and for ramp metering in Section 5.3.

For all the argumentation in this chapter we assume that the control objective is to minimize travel times, or more precisely, to minimize the total time spent (TTS). Furthermore, the focus in this chapter is on freeways. The analysis of urban networks is a topic for future research (cf. Section 9.4).

5.1 General conditions

The assumption that the control objective is to minimize the TTS means that if the outflow of the network is maximal under a given traffic demand[1].

The main sources of network performance degradation are

- the capacity drop,

- blocking of traffic that have routes that do not pass through the bottleneck location,

[1] Here we assume that the demand is given and is not influenced by the traffic control. If the effect of traffic control on the traffic demand (e.g., departure time choice) is taken into account, the TTS can also be influenced by demand spreading [163].

- capacity reducing events, such as incidents, road works, or weather conditions.

At least one of the three phenomena should occur in the controlled network, otherwise the TTS cannot be improved. In the following sections we consider only the case of capacity drop and blocking.

The general conditions that we state in this section serve two goals: first, to verify whether or not the capacity drop or blocking is present in the network, and second, to verify other conditions that determine if the traffic control can be applied successfully or not. In this respect we will consider the following topics:

- The modeling of the relevant phenomena. If a model-based approach is applied, the prediction model should be able to reproduce the relevant phenomena.
- The verification of the capacity drop.
- The verification of blocking.
- The availability of control measures that can limit the flow sufficiently in order to resolve congestion.
- The choice of the network boundaries and the set of traffic scenarios for which control can be useful.

5.1.1 Modeling relevant phenomena

If the control method applied to the traffic problem is model-based, the controller model should be able to reproduce the *capacity drop* or *blocking* phenomenon. Otherwise, the controller will not find the control signals that can eliminate the capacity drop.

Note that calibrating the controller model with real data through the minimization of some criterion (such as the mean quadratic difference of the model states/outputs and the data) may or may not result in a model that reproduces capacity drop. Therefore, the model should be tested if it reproduces the capacity drop after calibration. The reproduction of blocking effects is usually formulated more explicitly in traffic models and is not a problem.

5.1.2 Capacity drop

Capacity drop is the phenomenon that the outflow of a traffic jam is significantly lower than the maximum achievable flow at the same location[2]. As we know from Section 1.1.4 even a small drop in the outflow can have a big effect on the TTS in congested situations.

[2] The outflow of a moving bottleneck should be measured sufficiently far (downstream) away from the bottleneck, such that the flow is stationary, otherwise one may measure a transient (acceleration) state.

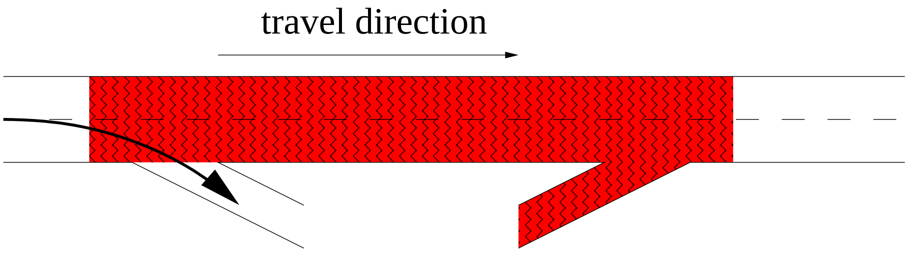


Figure 5.1: Congestion caused by excessive on-ramp demand blocks also the upstream off-ramp.

Note that the capacity drop resulting from a shock wave on a freeway stretch is different from a capacity drop resulting from a fixed bottleneck, such as an on-ramp. Kerner [84], Kerner and Rehborn [83], and Treiber *et al.* [159] distinguish between phenomena (i.e., traffic states) that occur at fixed bottlenecks, and phenomena that are moving, such as shock waves (or in their terms: moving localized clusters and wide moving jams), which may emerge from fixed bottlenecks or metastable traffic streams. The traffic states occurring at fixed bottlenecks include a stable congested state and one or more oscillatory states.

From a control point of view all these states are undesirable and should be prevented or eliminated. Capacity drop at fixed bottlenecks has been reported in field studies [20, 47], and the decrease in flow ranges from 0–15 %. Since the capacity drop is not observed in all cases, traffic data from the bottleneck that is to be controlled has to be studied carefully. A capacity drop of around 30 % resulting from shock waves has been reported by Kerner and Rehborn [85, 83]. In Section 5.2 we will find a similar value.

5.1.3 Blocking

Blocking occurs when a traffic jam spills back to an off-ramp, to an urban intersection or to a freeway diverge, and blocks traffic that has route other than via the effective bottleneck (see Figure 5.1). When blocking occurs, the performance can be improved by resolving the blockage, given that the downstream network on the route(s) of the blocked traffic can accommodate the increased flow. In this perspective the considerations of Section 5.1.5 about the choice of the boundaries of the network and scenarios also apply.

5.1.4 Sufficient flow limitation

A good traffic control strategy not only *prevents* or *reduces the chances of* congestion, but also *eliminates* congestion in the case that it could not be prevented by traffic control. A precondition for elimination congestion is that the net outflow of the congested area should be positive. In other words, the traffic control measures should be able to limit the inflow to the congested area to a level that is less than the outflow of the area.

5.1.5 Network and traffic scenario

When integrating multiple control measures one has to choose the network boundaries. For both the entry and exits of the network we present some guidelines to make the choice.

The roads downstream the exits of the network should be able to accommodate the traffic exiting the network. The goal of minimizing the TTS on a network is that drivers should — on the average — reach their destination as fast as possible. Ideally all destinations of all drivers are included. However, for some drivers the distance between their origin and destination is very large and it is not always possible to include all destinations in the controlled network. Since it is not useful to maximize the outflow of the network if there is a bottleneck (e.g., congestion) downstream one or more of the exits of the network[3], it is necessary to verify[4] whether the roads downstream the network exit can accommodate the increased flows achieved by traffic control [17]. This also means that if there are several bottlenecks on a route, the 'most downstream' bottleneck has to be solved first.

The entrances of the network should be chosen such that all delays caused by control occur in the network. Traffic control may temporarily delay traffic, and these delayed vehicles should be inside the network in order to be taken into account when optimizing TTS.

Also the upstream traffic demand scenario can strongly influence the improvement that can be achieved by traffic control. Traffic control is only effective for certain traffic demands and it should be verified whether these demands occur frequently enough to justify the control. E.g., if there is a bottleneck upstream the entrance of the network (such as a tunnel or a congested on-ramp) the capacity of certain roads may not be achieved just because of the limited inflow.

5.2 Speed limit control against shock waves

In this section we focus on conditions necessary for successful speed limit control. We examine the general conditions of capacity drop and sufficient flow limitation in relation with speed limits. Furthermore, two other conditions are discussed that are necessary to

[3] This could even make the situation worse, because the improved traffic flow runs faster in the downstream congestion. This may cause a longer traffic jam and/or increase the time necessary to resolve the congestion.

[4] This verification can be accomplished by estimating capacities based on traffic data or based on geometrical considerations, such as given in the Highway Capacity Manual, or in case of urban areas based on traffic signal timings (since traffic signals at off-ramps often form a bottleneck).

completely eliminate shock waves: *metastability*, which means that a short unstable shock wave can be converted into a longer but stable disturbance; and *the minimum length of the speed limit controlled freeway,* which makes this conversion possible.

Now we discuss these points in more detail:

- **Capacity drop.** To assess the achievable improvement the capacity drop has to be estimated. The capacity drop is estimated by comparing outflow of a shock wave with the maximum flow of freely flowing traffic. The time and location for the outflow measurement of the shock wave have to be such that there are no on-ramps or off-ramps between the shock wave and the measurement point, otherwise the entering or exiting traffic could bias the estimation. Furthermore, the traffic should be in homogeneous free flow, to be sure that the flow drop is not caused by a downstream bottleneck, and that we are not measuring a transient state. For other capacity estimation techniques we refer to [107, 101].

  We explain the capacity drop estimation by an example. In Figure 5.3 a typical traffic scenario is shown for the A1 close to Deventer in The Netherlands. The location of the on-ramp and off-ramps is shown in Figure 5.2.On this stretch typically on-ramp jams are created at location 88.7 km and 96.2 km. From these jams shock waves emerge that propagate upstream through the freeway. In Figure 5.4 (which is a zoom-in of Figure 5.3) one shock wave is shown, which starts at point A and is caused by excessive on-ramp traffic just before 96.2 km. There is an off-ramp just after 92.7 km, and on-ramps before respectively 89.9 km and 88.7 km, which can be recognized by the sudden increase of the flow at the detector locations after the on-ramps. Between 89.9 km and 92.7 km there are no on-ramps or off-ramps and this is the area that we consider in this example for capacity drop estimation[5]. The area where the traffic is in free flow again is around point B in Figure 5.4: the speed is approximately 100 km/h and the flow around 3000 veh/h. An example of the capacity flow is around point C, where the flow is approximately 4200 veh/h. We can conclude that there is a capacity drop of roughly 30 %, which is close to the value found in [85].

- **Sufficient flow limitation.** To effectively eliminate the shock wave the minimum value of the speed limit should result in a flow that is lower than the outflow of a congested area, otherwise the density will not decrease even when the speed limit is set to its minimum value. In The Netherlands the lowest dynamic speed limit is 50 km/h, and at the A1 freeway near to Deventer the average flow at this speed is

[5] It is interesting to note in Figure 5.4 that after the shock wave has passed the on-ramp upstream of 88.7 km (which connects another freeway) the flow suddenly increases because of the additional vehicles from the on-ramp.

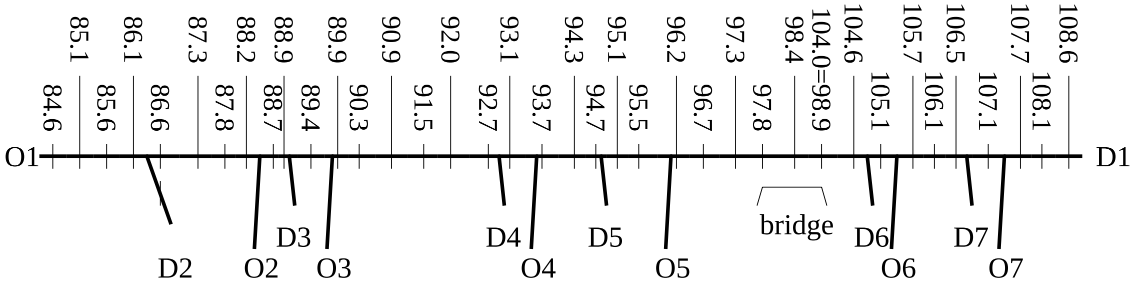


Figure 5.2: A schematic view of the considered stretch of the A1. The detector locations are in units of km's, and the on/off-ramps are indicated by respectively O1–O7 and D1–D7.

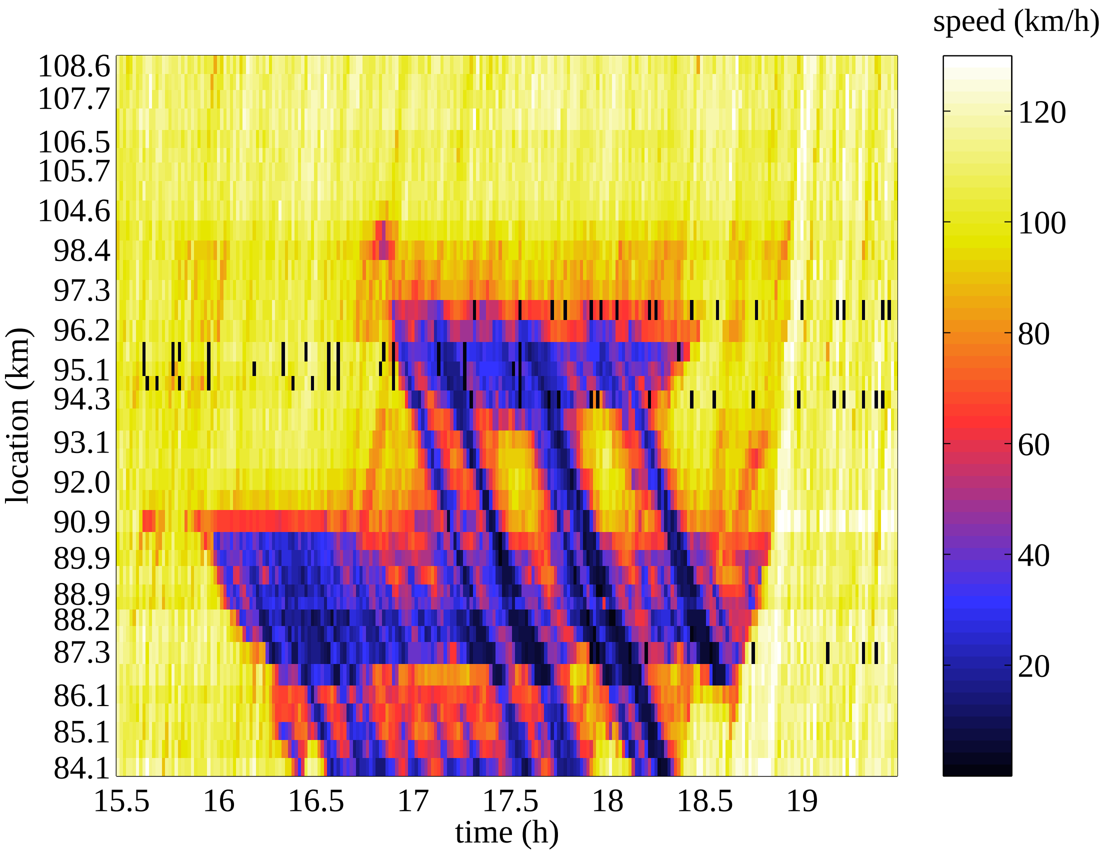


Figure 5.3: A typical traffic scenario on the A1 freeway in The Netherlands. The shock waves start at on-ramps and propagate backward through the link.

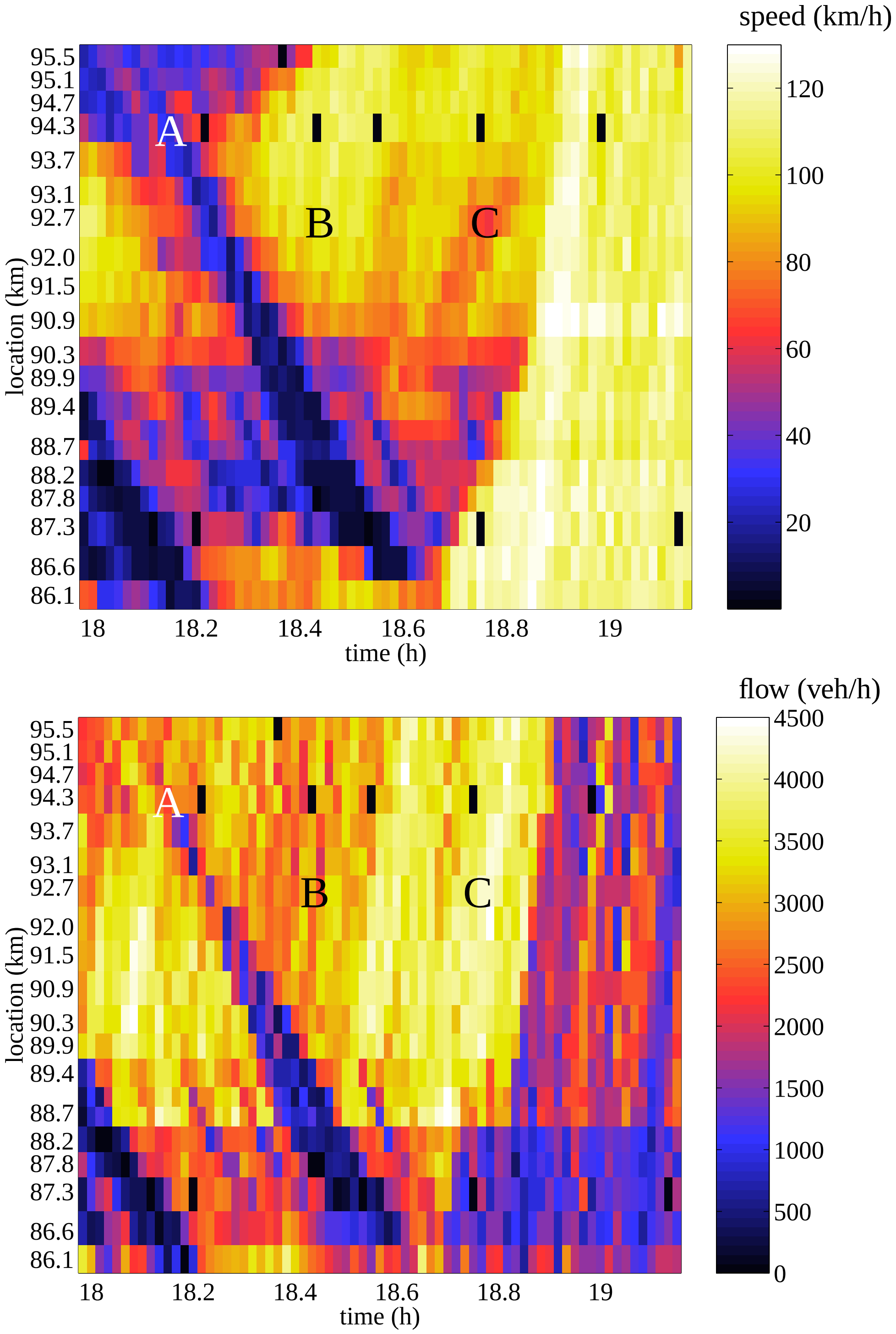


Figure 5.4: Example of the capacity drop resulting from a shock wave.

2900 veh/h which is somewhat less than the estimated outflow of the shock wave (3000 veh/h).

- **Metastability.** Traffic flow should be in metastable state, by which we mean that the traffic demand is such that two stable states may be possible: a stable freely flowing traffic stream, or a stable upstream propagating traffic jam (or: shock waves). Metastability also implies that starting from a freely flowing situation small disturbances in the traffic flow will disappear, but large disturbances cause a breakdown and a upstream propagating traffic jam. The metastability condition is necessary for speed control because the speed limits should have the possibility to 'convert' the shock wave into a wider, less intense and stable wave. It seems plausible to say that this condition is satisfied if the traffic demand is between the outflow (which is now reduced) of a shock wave and the capacity of the freeway. If the demand is lower than the outflow of a shock wave, the high density region will automatically disappear so the traffic stream is stable. A demand close to capacity is marginally stable since even a very small disturbance can cause a breakdown. For demands between the outflow of the shock wave and the capacity of the freeway it seems reasonable to expect that the closer the demand is to capacity, the smaller the disturbance needed to cause a breakdown, which means that we have metastability.

- **Sufficient length of the speed controlled freeway.** There should be enough speed limits (enough length) to suppress a typical shock wave without causing a new shock wave. When a speed limit becomes active to limit the inflow of a downstream shock wave, it will cause an increasing density in the upstream segment. To prevent that this density becomes too high and causes instability, a second speed limit should become active that limits the inflow to this segment, and so on. This process continues until the shock wave is resolved and the speed limits can be released gradually. In this way the unstable shock wave is redistributed in a longer but smaller (in density) and stable wave. The necessary length of the speed controlled area depends on the number of excess vehicles (compared to capacity flow) in the shock wave.

5.3 Ramp metering against on-ramp jams

In this section we focus on conditions necessary for successful ramp metering against on-ramp jams. We examine the general conditions of capacity drop and sufficient flow limitation in relation with ramp metering. Also in this section we assume that there is a capacity drop on the freeway at the on-ramp. If there is no capacity drop, the benefits of local ramp metering are small [162].

Now we discuss these points in more detail:

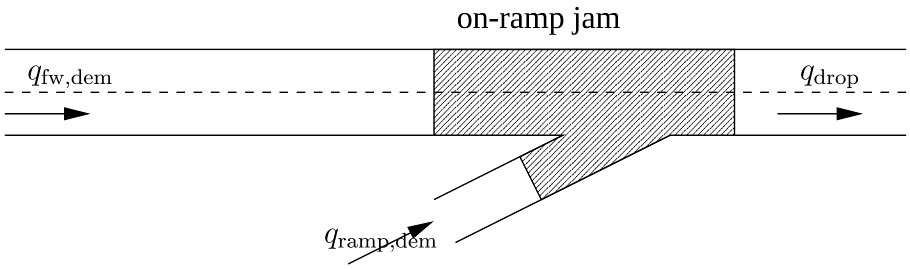


Figure 5.5: A traffic jam caused by high on-ramp flows.

- **Capacity drop.** Capacity drop at on-ramps is observed in field studies [20, 47, 153] and its value $q_{\mathrm{cap}} - q_{\mathrm{drop}}$ ranges from 0–15 % of the freeway capacity q_{cap}, where q_{drop} is the traffic flow after the breakdown has occurred (see also Figure 5.5).

  The capacity drop occurs when the density on the freeway is high, and the flow can be restored to a higher value if the density is reduced. A jam at an on-ramp can be triggered by too high demands, but also by a peak in the ramp demand or an upstream propagating shock wave on the freeway. After an on-ramp jam has been triggered it can remain existent for a long time, even if the peak in the on-ramp flow was short or when the shock wave has passed the on-ramp area. The high-density area at the on-ramp is often self-maintaining since the capacity of the ramp outflow has dropped from a value around capacity q_{cap} to q_{drop} ($< q_{\mathrm{cap}}$), and may not be sufficient anymore to accommodate the main-stream and ramp demands (see Figure 5.5). If $q_{\mathrm{fw,dem}} + q_{\mathrm{ramp,dem}} > q_{\mathrm{drop}}$ then the ramp jam will be self-maintaining. To reduce the density on the freeway at the on-ramp the ramp flow has to be limited to a value that is lower than before the creation of the jam. This is called hysteresis [83].

- **Sufficient flow limitation and demand scenario.**[6] Ramp metering is only useful if there is a jam and if ramp metering can sufficiently limit the flow to remove the high-density area. This seems obvious, but as we will see, in certain cases the ramp flow cannot be limited sufficiently. To explain this, we assume an on-ramp configuration as in Figure 5.5, and we consider the traffic behavior as a function of the freeway demand $q_{\mathrm{fw,dem}}$ and the ramp demand $q_{\mathrm{ramp,dem}}$ in Figure 5.6. The freeway demand is between 0 and the freeway capacity q_{cap} (in the figure assumed to be 4000 veh/h, line f) and the on-ramp demand is between 0 and the ramp capacity $q_{\mathrm{ramp,cap}}$ (assumed to be 2000 veh/h, line b). The ramp outflow capacity is assumed to be equal to q_{cap} and if $q_{\mathrm{fw,dem}} + q_{\mathrm{ramp,dem}} > q_{\mathrm{cap}}$ (above line d) the

[6] In this discussion we do not consider the occurrence of re-routing effects of ramp metering. If re-routing effects do occur, the discussion is similar, except for the case that re-routing would result in a ramp demand that is lower than $q_{\mathrm{r,min}}$

line	equation
a	$q_{\mathrm{r}} = q_{\mathrm{r,min}}$
b	$q_{\mathrm{r}} = q_{\mathrm{r,max}}$
c	$q_{\mathrm{r}} + q_{\mathrm{f}} = q_{\mathrm{drop}}$
d	$q_{\mathrm{r}} + q_{\mathrm{f}} = q_{\mathrm{cap}}$
e	$q_{\mathrm{f}} = q_{\mathrm{drop}} - q_{\mathrm{r,min}}$
f	$q_{\mathrm{f}} = q_{\mathrm{cap}}$

Table 5.1: The equations of the lines in Figure 5.6.

ramp will become congested, the outflow will drop to q_{drop} and a high-density region will be created on the freeway at the on-ramp. To remove this high-density region the total demand $q_{\mathrm{fw,dem}} + q_{\mathrm{ramp,dem}}$ has to be lower than q_{drop}, i.e., the point $(q_{\mathrm{ramp,dem}}, q_{\mathrm{fw,dem}})$ should be in the area below line c[7] (where the capacity drop assumed to be 10 %: $q_{\mathrm{drop}} = 0.9q_{\mathrm{cap}}$). The minimum metering rate is $q_{\mathrm{r,min}}$ (600 veh/h, which is a typical value for The Netherlands for a two-lane on-ramp, line a). Line e represents the freeway flow for which lines a and c intersect. For this flow ramp metering cannot limit the ramp flow sufficiently anymore to remove the high-density region. The areas of possible combination of freeway and ramp demands are labeled R–Z and are separated by the lines a–f.

Now we consider the traffic behavior in the areas R–Z:

R: The main-stream demand $q_{\mathrm{fw,dem}}$ is close to capacity, the ramp flow $q_{\mathrm{ramp,dem}}$ is small but the total demand $q_{\mathrm{fw,dem}} + q_{\mathrm{ramp,dem}}$ exceeds capacity, so there is a jam at the on-ramp. Since the ramp demand is below the minimum metering rate, ramp metering is not effective.

S: The total demand is between the dropped capacity and the capacity, so there may be a jam at the on-ramp. If there is a jam, ramp metering is not effective since the ramp demand is below the minimum metering rate.

T,U,V: The total demand is below the dropped capacity, so no jam is formed at the on-ramp. Ramp metering is not necessary.

W: The total demand is between the dropped capacity and the capacity, so there may be a jam at the on-ramp. If there is a jam, ramp metering can reduce the ramp flow to a value that results in a total inflow of the jammed area that is lower than the dropped capacity (to the left of line c). The jam will disappear and the outflow can be restored to a value that equals $q_{\mathrm{fw,dem}} + q_{\mathrm{ramp,dem}}$. If there is no jam, ramp metering is not necessary.

[7] We assume these boundaries to be straight lines, which may not be realistic. However, the interpretation of the diagram will not change if these curves have different (but more realistic) shapes.

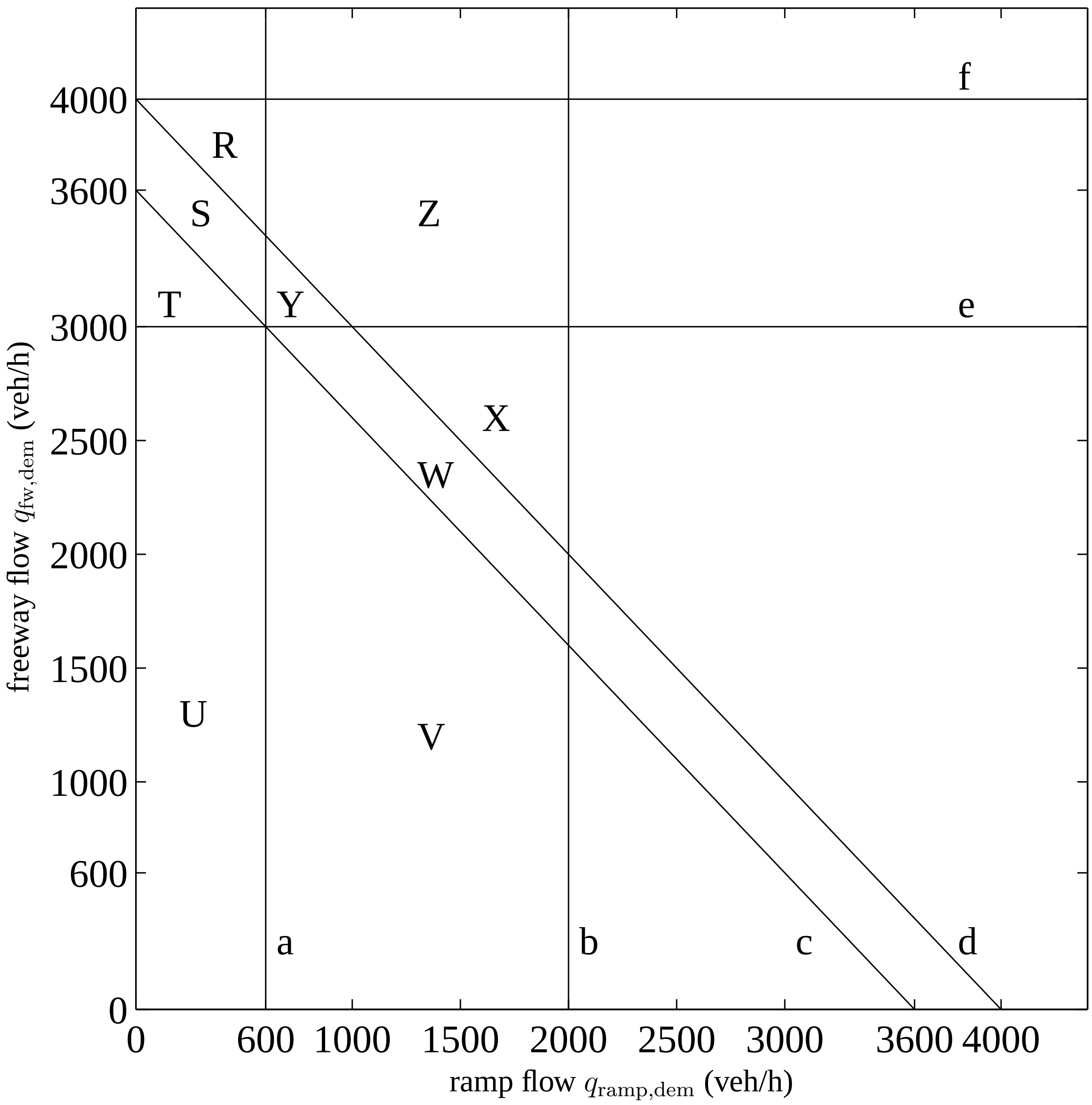


Figure 5.6: Ramp metering is only useful in areas W and X. The definition of the lines a–f is given in Table 5.1

X: The total demand exceeds capacity, so there is a jam at the on-ramp. Ramp metering can reduce the ramp flow to a value that results in a total inflow of the jammed area that is lower than the dropped capacity. The jam will disappear and the outflow can be restored to a value that is close to capacity, but smaller than the total demand.

Y: The total demand is between the dropped capacity and the capacity, so there may be a jam at the on-ramp. If there is a jam, ramp metering cannot reduce the ramp flow to a value that results in a total inflow of the jammed area that is lower than the dropped capacity. If there is no jam, ramp metering is not necessary.

Z: The total demand exceeds capacity, so there is a jam at the on-ramp. Ramp metering cannot reduce the ramp flow sufficiently.

The total area where ramp jams can occur is (R,S,W,X,Y,Z), However, from the explanation above, we may conclude that ramp metering is definitely effective for area X and possibly effective for area W. In other words, for effective ramp metering the on-ramp demand must be relatively high — which is typically not a restrictive requirement —, and the main-stream demand has to satisfy

$$q_{\mathrm{fw,dem}} < q_{\mathrm{drop}} - q_{\mathrm{r,min}}$$

which is significantly lower than capacity.

A typical scenario that often occurs at an on-ramp is when a shock wave is triggered from the ramp jam (see Figure 5.7). This shock wave has an outflow of approximately $0.7q_{\mathrm{cap}}$, which will be the main-stream demand $q_{\mathrm{fw,dem}}$ of the on-ramp. This means that the on-ramp state is in area W or X (not in V since there is a ramp jam), and ramp metering is useful. Although in this case ramp metering is useful, there is a remaining upstream traveling shock wave on the freeway that is unsafe, and increases travel time for all drivers on the freeway. The best solution would be to reduce the inflows on the freeway earlier, and possibly combine it with ramp metering.

Another typical scenario is when the ramp queue length is constrained, and when the constraint is violated the ramp metering is switched off. This further reduces the effectiveness of ramp metering and constrains the demands for which ramp metering is useful even further. In area X the demand structurally exceeds capacity, so ramp metering will be active until the ramp queue constraint is violated and ramp metering is switched off. If the demand is on the average in area W, but a jam is triggered by a peak in the ramp flow or a shock wave, it may be possible to resolve the ramp jam without violating the queue constraint, but there is no guarantee for this.

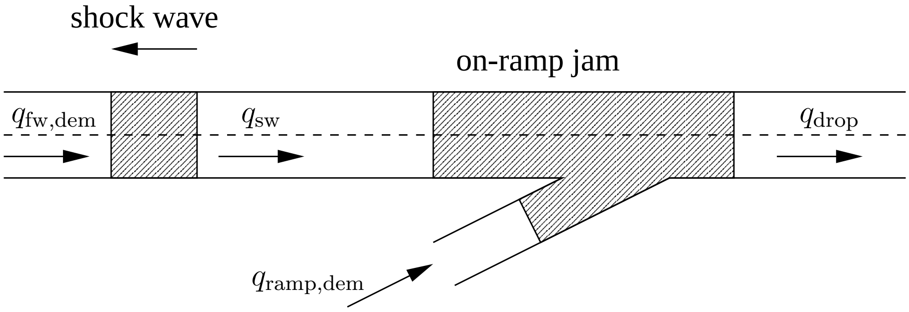


Figure 5.7: The shock wave upstream from the on-ramp jam feeds the on-ramp with a constant demand of approximately $0.7q_{\mathrm{cap}}$.

5.4 Conclusions

In this chapter we have presented conditions that are necessary for effective traffic control, under the assumptions that the total time spent (TTS) is to be minimized and that the cause of the performance degradation is the capacity drop or the blocking of traffic not traveling over the real bottleneck location.

The conditions necessary for effective control include

- the presence of the capacity drop or blocking in the real traffic situation,
- if model-based predictive control is applied, the ability of the traffic model to reproduce these phenomena, with a sufficiently high accuracy,
- the possibility to sufficiently reduce the inflow of the congested area,
- the network boundaries should be chosen such that the vehicles that are delayed by traffic control are inside the network,
- the network boundaries should be chosen such that the roads downstream can accommodate the improved traffic flows,
- the presence of traffic demands for which control is useful.

From these general conditions specific conditions are derived for speed limits and ramp metering. For speed limits, the traffic flow demand should be between the capacity flow and the flow after the capacity drop. This results in a metastable state, where the unstable shock wave can be converted into a wider but stable disturbance with a higher outflow. For ramp metering the analysis of freeway and ramp demands shows that the region for which ramp metering can improve the total time spent is relatively small compared to the region where congestion occurs. The main reason for this is that usually the ramp metering rate is bounded from below, and as a result the inflow of the congested area cannot be restricted sufficiently. Ramp metering will be effective if there is a ramp jam and the condition

$$q_{\mathrm{fw,dem}} < q_{\mathrm{drop}} - q_{\mathrm{r,min}}$$

is satisfied, where $q_{\mathrm{fw,dem}}$ is the freeway demand, q_{drop} the outflow of the ramp jam (after the capacity drop), and $q_{\mathrm{r,min}}$ the minimum ramp flow.

Chapter 6

Dynamic speed limit control

In this chapter we demonstrate the MPC control framework for traffic problems that can benefit from the use of speed limits. We discuss the integrated control of several combinations of control measures, such as speed limits and ramp metering, main-stream metering and ramp metering, and multiple speed limits.

In Section 6.1 we deal with the integrated control of ramp metering and speed limits, where the speed limits can prevent a traffic breakdown when ramp metering only is insufficient. Since the main effect of the speed limits in this section is to limit the flow when necessary, in Section 6.2 this set-up is compared with a set-up where the speed limits are replaced by main-stream metering. Another application of speed limits is found in Section 6.3 where speed limits are used to reduce or eliminate shock waves on freeways.

In each section we present a benchmark problem to illustrate the developed approach. For each benchmark problem we first give a general description of the problem. Next, we formulate the objective function, and the constraints of the optimization problem. Next, we present the set-up of the benchmark problem including the traffic scenario. Finally, we present the simulation results. The conclusions are stated in Section 6.4.

6.1 Integrated ramp metering and variable speed limits

In this section we consider traffic networks that are controlled by ramp metering and variable speed limits. We demonstrate that speed limits can complement ramp metering, when ramp metering alone is not efficient. The use of dynamic speed limits significantly reduces congestion and results in a lower total time spent.

In this section we focus on restrictive ramp metering and speed limits both aiming at improving the traffic flow by preventing a traffic breakdown as mentioned in Sections 2.1 and 2.2.

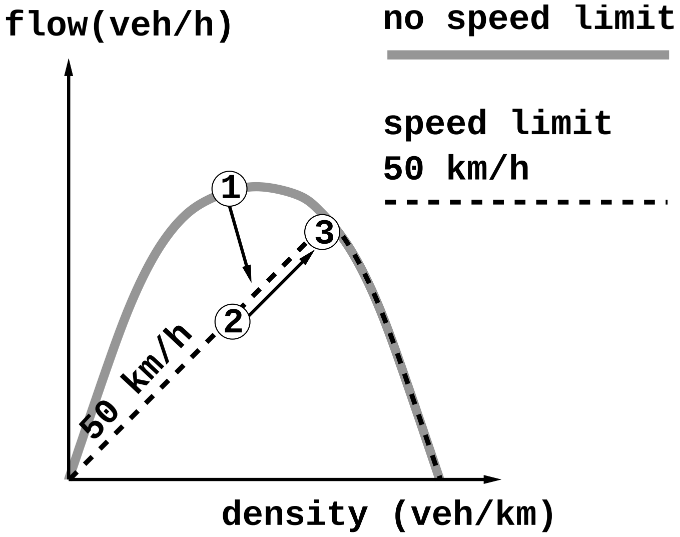


Figure 6.1: When traffic on the main road is in state 1, then it is nearly unstable and even a small flow from the on-ramp can cause a breakdown. Speed limits change the state from 1 to somewhere between 2 and 3, and change the shape of the fundamental diagram from the solid gray line to the dashed black line. The decrease of flow creates some space for the on-ramp traffic.

6.1.1 Problem description

It is clear that ramp metering is only useful when traffic is not too light (otherwise ramp metering is not needed) and not too dense (otherwise breakdown will happen anyway)[1]. This region is on the stable (left) side of the fundamental diagram, and close to the top (see state 1 in Figure 6.1), where a breakdown can happen. In practice there is often a constraint present regarding the operation of the ramp metering device[2], which has as a consequence that the minimum flow from the on-ramp is higher than zero[3]. Even these minimal flows can cause a breakdown when the traffic on the mainstream freeway is dense, which results in a reduced outflow and increased travel times.

Here we consider a situation where speed limits are imposed on the mainstream traffic while the on-ramp is metered (see Figure 6.2). The main idea is that when ramp metering is unable to prevent congestion on its own, adding variable speed limits could prevent a breakdown by limiting the inflow into the area where the traffic breakdown starts. The

[1] In this section we assume that the downstream traffic conditions are congestion free. If this assumption does not hold, ramp metering could be useful in a larger variety of scenarios.

[2] Examples of such constraints are: a minimum metering rate or a maximum queue length on the on-ramp. Both constraints prevent too long queues on the on-ramp, which could block other traffic streams on the secondary road network.

[3] Completely closing an on-ramp is in certain cases also an option. In that case that traffic is forced to use a route on the secondary roads.

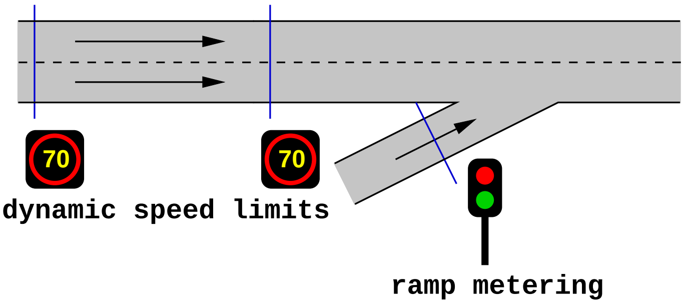


Figure 6.2: We consider a combination of speed limits and ramp metering as control measures.

speed limits change the shape of the fundamental diagram (see for an illustration Figure 6.1, and for the motivation see Section 3.3.1) and reduce the outflow of the controlled segment. Suppose the traffic state on the freeway is 1. When a speed limit is applied, the speed drops and the density increases, so the traffic state will move to a state somewhere between 2 and 3. However, because of the high traffic demand (state 1) the state will approach state 3, the capacity of the new fundamental diagram (dashed). Since this flow is lower than the capacity of the freeway without speed limit, there will be some space left to accommodate the traffic from the on-ramp and a breakdown is prevented. Consequently, the density in the on-ramp area on the freeway remains low and the outflow remains high.

A drawback of the above approach is that the mainstream flow will also decrease. But if the control is optimized properly, this flow drop will be always less than or equal to the flow drop of a breakdown, since otherwise breakdown would be the optimal situation. Another point of criticism could be that the approach keeps the controlled network congestion free, but at the cost of creating congestion at the entrances of the network. This is only partially true, because the controller will indeed delay the traffic sometimes to prevent a breakdown in the network, but afterward the flow will be higher than when the breakdown would have occurred. So, the inflow of the controlled part of the network will be decreased by the speed limits only for a short period of time. Unfortunately this still can cause congestion in upstream sections. A remedy could be to extend the size of the network with as many (uncontrolled) upstream sections as necessary to cover the congested area. In this way the congestion caused by the speed limits will not spill back to the mainstream origin queue and the congestion dynamics can be taken into account by the controller. Second, the network that is considered (i.e., evaluated and controlled) can be chosen larger, because the traffic is apparently so dense that the effects of the control measures reach beyond the bounds of the actual network.

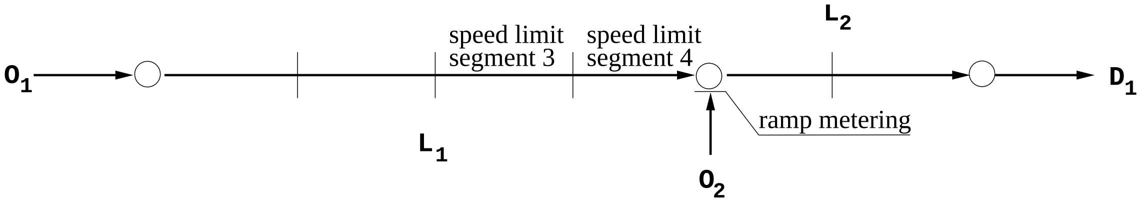


Figure 6.3: The benchmark network includes two sections with speed limits, and a metered on-ramp.

Objective function

The model predictive control algorithm finds the control signals $r_o(\ell)$ and $v_{\text{control},m,i}(\ell)$ for $\ell \in \{k_\text{c}, \ldots, k_\text{c} + N_\text{c} - 1\}$ that minimizes the selected objective function.

Note that similarly to Chapter 4 we distinguish between the controller time step length T_c and the simulation time step length T, and between the controller time step counter k_c and the model time step counter k. We assume that the controller time step length is an integer multiple of the simulation time step length: $T_\text{c} = MT$, with M a positive integer.

We select the following objective function:

$$J(k_\text{c}) = \sum_{k=Mk_\text{c}}^{M(k_\text{c}+N_\text{p})-1} \left\{ \sum_{(m,i)\in I_\text{all}} \rho_{m,i}(k) L_m \lambda_m + \sum_{o\in O_\text{all}} w_o(k) \right\} + \\ \sum_{\ell=k_\text{c}}^{k_\text{c}+N_\text{c}-1} \left\{ \xi_\text{ramp} \sum_{o\in O_\text{ramp}} \big(r_o(\ell) - r_o(\ell-1)\big)^2 + \right. \\ \left. \xi_\text{speed} \sum_{(m,i)\in I_\text{speed}} \left(\frac{v_{\text{control},m,i}(\ell) - v_{\text{control},m,i}(\ell-1)}{v_{\text{free},m}} \right)^2 \right\} ,$$

where I_all is the set of indexes of all pairs of segments and links and O_all is the set indexes of all origins, O_ramp is the set of indexes o of those on-ramps where ramp metering is present, and I_speed is the set of pairs of indexes (m,i) of the links and segments where speed control is present. This objective function contains two terms for the TTS (one term for the mainstream traffic and one term for the on-ramp queues), and two terms that penalize abrupt variations in the ramp metering and speed limit control signals respectively. The last two terms are weighted by the non-negative weight parameters ξ_ramp and ξ_speed.

6.1.2 Benchmark problem

Network and scenarios

The benchmark network for the experiments (see Figure 6.3) is chosen as simple as possible, since we want to focus on the relevant points of the approach presented in this section. The network consists of a mainstream freeway with two speed limits, and a metered on-ramp. The second speed limit is included to have more control over the state (i.e., speed, density) in the segment that is just before the on-ramp. The network considered consists of two origins (a mainstream and an on-ramp), two freeway links, and one destination. O_1 is the main origin and has two lanes with a capacity of 2000 veh/h each. The freeway link L_1 follows with two lanes, and is 4 km long consisting of four segments of 1 km each. Segments 3 and 4 are equipped with a variable message sign (VMS) where speed limits can be set. At the end of L_1 a single-lane on-ramp (O_2) with a capacity of 2000 veh/h is attached. Link L_2 follows with two lanes and two segments with length of 1 km each, and ends in destination D_1 with unrestricted outflow. We assume that the queue length at O_2 may not exceed 100 vehicles, in order to prevent spill-back to a surface street intersection.

We use the network parameters as found in [93]: $T = 10\,\mathrm{s}$, $\tau = 18\,\mathrm{s}$, $\kappa = 40$ veh/km/lane, $\nu = 60$ km^2/h, $\rho_{\max} = 180$ veh/km/lane, $\delta = 0.0122$, $a_1 = a_2 = 1.867$, $v_{\mathrm{free},m} = 102\,\mathrm{km/h}$ and $\rho_{\mathrm{crit},m} = 33.5$ veh/km/lane, for $m \in 1, 2$.

Furthermore, we assume that the desired speed is 10% higher than the displayed speed limit, that is $\alpha = 0.1$; and we set the signal variance penalty weights ξ_{ramp} and ξ_{speed} to 0.4.

The controller sampling time is chosen to be 1 min. During one sampling interval the control signals are assumed to be constant.

To examine the effect of the combination of variable speed limits and ramp metering a typical demand scenario is considered (see Figure 6.4): The mainstream demand has a constant, relatively high level and a drop after 2 h to a low value in 15 min. The demand on the on-ramp increases to near capacity, remains constant for 15 min, and decreases finally to a constant low value. The scenario was chosen such that the final (and steady) state of the network is the same for all scenarios (which means that all final densities are the same for all scenarios). This enables us to compare the TTS of the different scenarios. Under the given demand scenario the ‘no-control’ case, the ‘ramp metering only’ case and the ‘coordinated speed limits and ramp metering’ case are compared. Since the control optimizes TTS, the relevant quantity for comparing the following simulations is TTS.

The utilized numerical algorithm to solve the optimization problem is the MATLAB implementation of the SQP algorithm (fmincon). This optimization method has the advantage that the constraints can be formulated explicitly, without the use of a penalty term in the objective function. For the SQP algorithm we refer to Section 4.1.5.

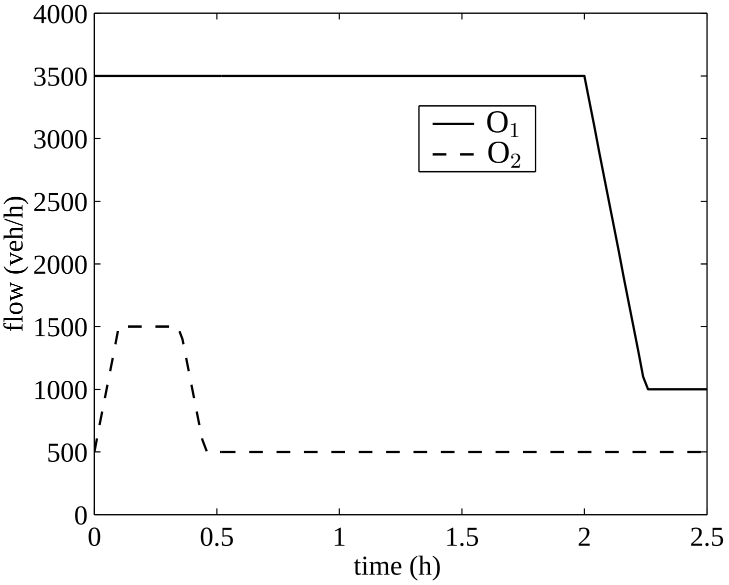


Figure 6.4: The demand scenario considered in the simulation experiments.

Results

The results of the 'no control' case are shown in Figure 6.5. As the demand increases on the on-ramp, the density on the main-stream on-ramp section (segment 1 of link 2) also increases and creates a congestion that propagates through link 1 and grows outside (upstream) the network. This causes a queue at origin 1 of approximately 150 vehicles. When the demand on the on-ramp drops to 500 veh/h the densities of segment 1 of link 2 up to segment 1 of link 1 gradually decrease but remain at a relatively high value (approximately 50 veh/km/lane). Finally, when the mainstream demand drops, the congestion dissolves and the network reaches steady state. The TTS is 1460.0 veh.h.

In Figure 6.6 we can see that the 'ramp metering only' case performs well (i.e., density in the on-ramp segment, link 2 segment 1, remains around 40 veh/km/lane, and the outflow of the network is high) until the maximum queue length is reached. After that point the performance breaks down in the 'ramp metering only' case, but in the 'coordinated speed limits and ramp metering' case the speed limits become active at this moment (see Figure 6.7). The speed limits reduce the inflow of the critical segment and cause a lower density (around 40 veh/km/lane) which enables a higher outflow. The TTS in the 'ramp metering only' case was 1382.6 veh.h, which is an improvement of 5.3 % compared to the 'no control' case. The TTS in the 'coordinated speed limits and ramp metering' case is 1251.0 veh.h, which is an improvement of 14.3 % with respect to the 'no control' case. So, we can conclude that the speed limits substantially improve the network performance.

For the MPC-controlled cases the optimal prediction horizon was found to be approx-

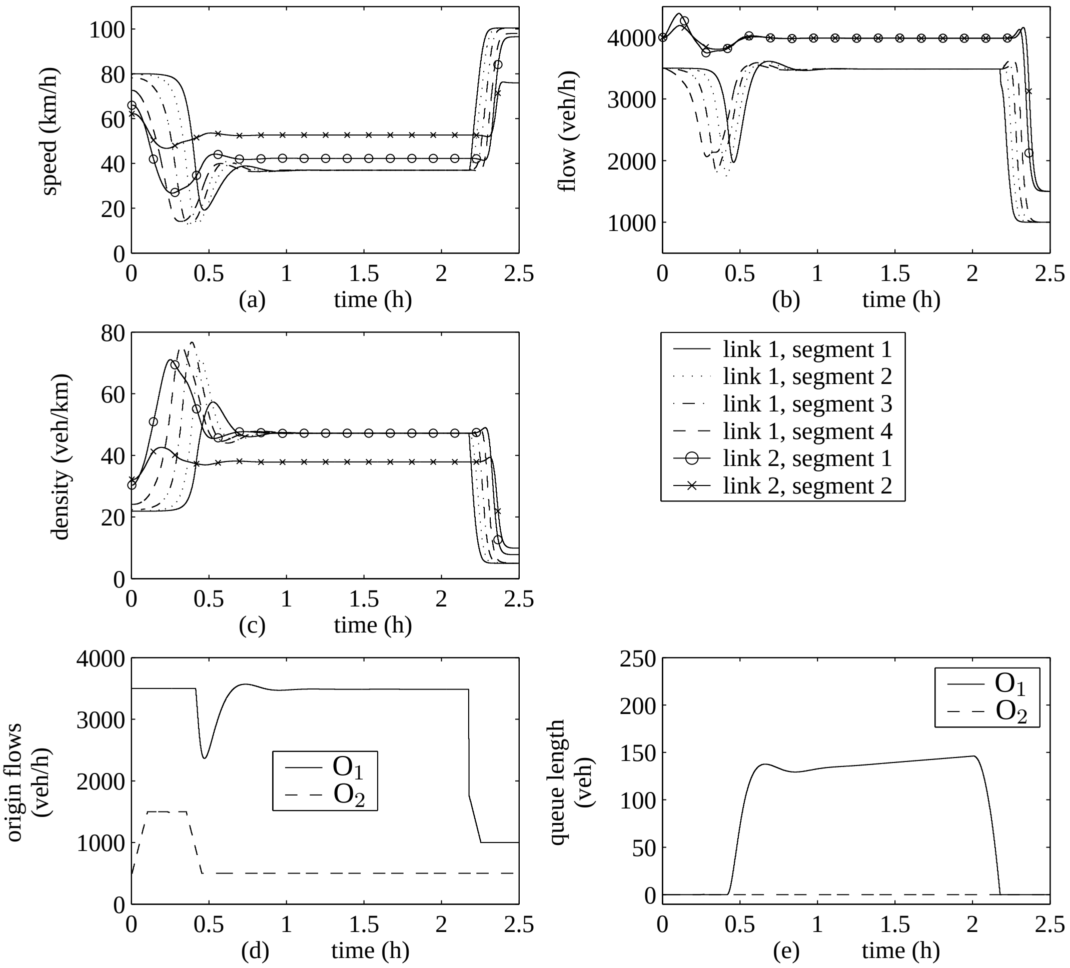


Figure 6.5: The simulation for the 'no control' case. The high demand on the on-ramp causes a congestion that remains existing until the main-stream demand drops to a low value.

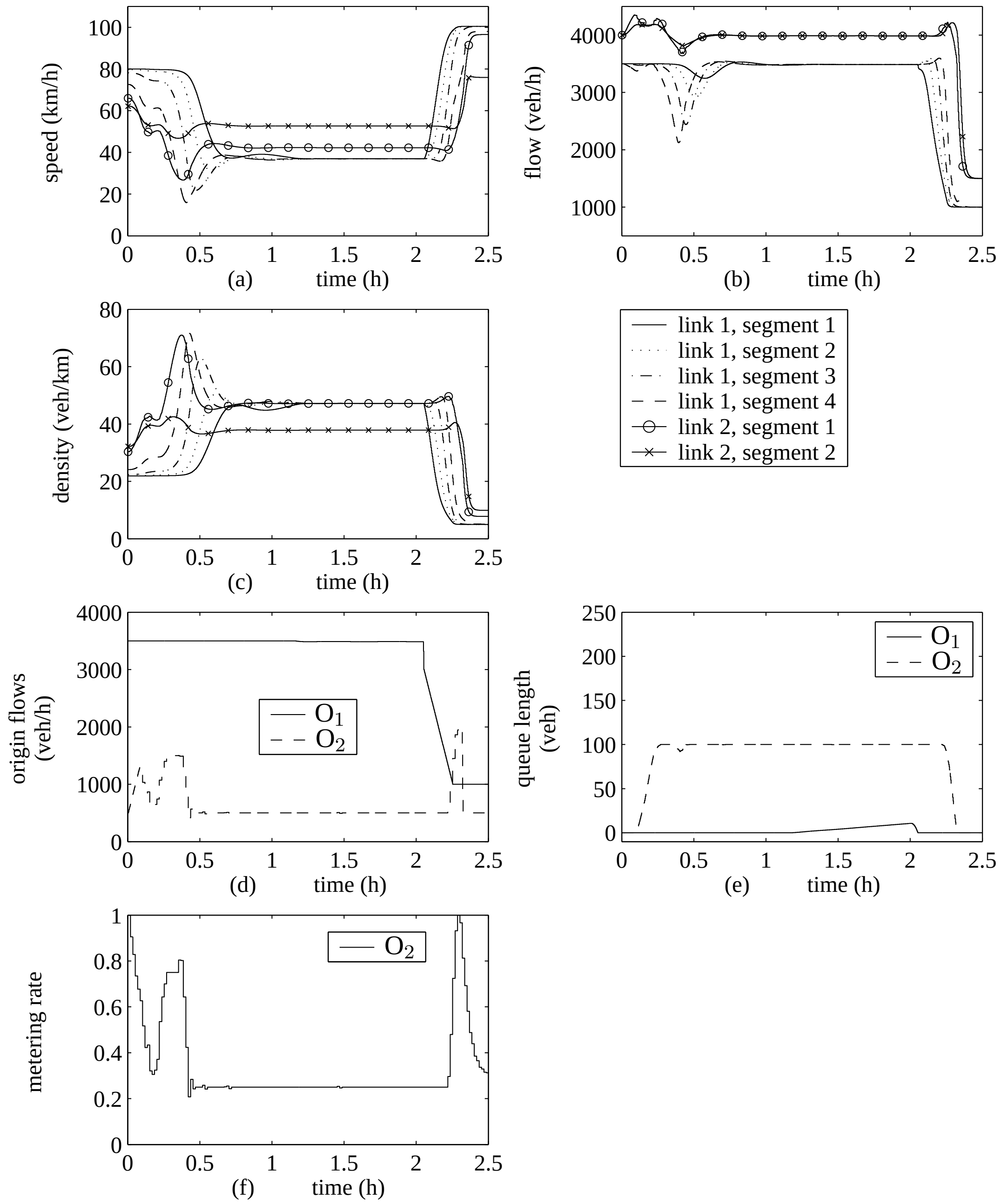


Figure 6.6: The simulation results for the ‘ramp metering only’ case. When the density approaches the critical density (33.5 veh/km), the ramp metering gradually switches on, and keeps the flow high (around 4200 veh/h). After the queue length on the on-ramp has reached the maximum queue length (100 veh), the ramp metering is forced to let the complete demand enter the freeway. This causes a congestion, and hence results in a reduced outflow. The queue at O_1 is slowly increasing until the main-stream demand drops.

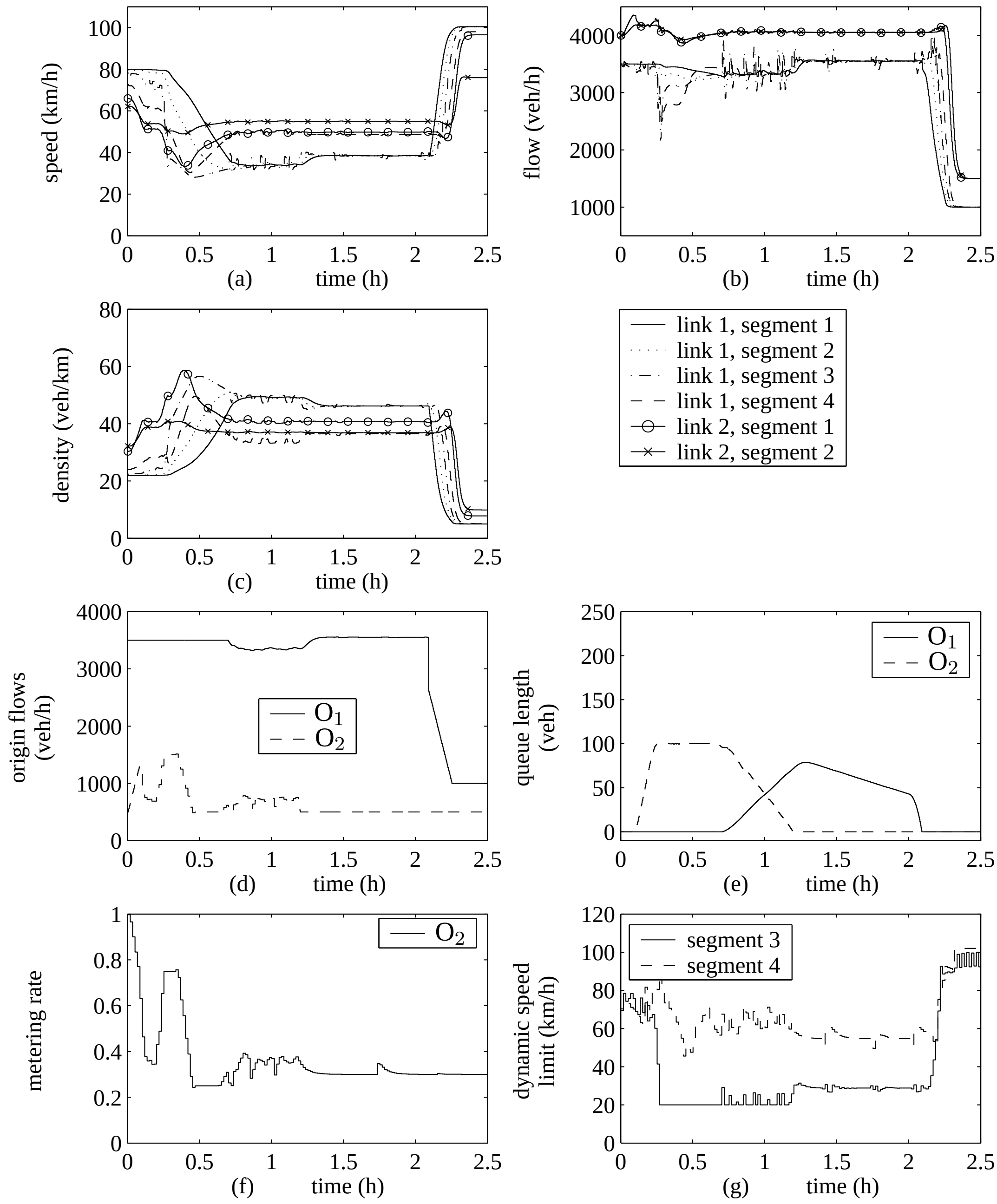


Figure 6.7: The results for the ‘coordinated speed limits and ramp metering’ case. In the beginning the ramp metering signal is similar to that of the ‘ramp metering only’ case, but when the density becomes too high (which may reduce the high outflow), the speed limits become active and reduce the inflow from the mainstream. This initially causes a longer queue on the mainstream, but the outflow is kept higher and both queues (O_1 and O_2) are decreasing even before the main-stream demand drops.

imately $N_{\mathrm{p}} = 7,$ which is in the order of the typical travel time through the network (4 km / 40 km/h = 0.1 h = 6 min). Shorter prediction horizons did not take the whole response of the system into account and result in insufficient control actions. Longer prediction horizons tend to take the future demand too much into account, which degrades the performance. This can be explained by the difference between the prediction horizon and the control horizon and the assumption that the control inputs are constant for $k_{\mathrm{c}} \in \{N_{\mathrm{c}}, \ldots, N_{\mathrm{p}} - 1\}$. If the difference between N_{p} and N_{c} is large, the control input for this period, which is represented by one optimisation parameter, will have a relatively large influence on the objective function. Consequently, the optimization will be focused more on this period (the end) than on the beginning of the control signal, while in the MPC cycle only the first sample of the optimized control signal is applied to the process. For this reason, a large difference between N_{p} and N_{c} may result in a lower performance. When the difference $N_{\mathrm{p}} - N_{\mathrm{c}}$ was kept constant (corresponding to a difference of 2 min) further increase of N_{p} caused only a small decrease of the TTS. A control horizon $N_{\mathrm{c}} = 3$ was sufficient for the 'ramp metering only' case, for the 'coordinated speed limits and ramp metering' case a control horizon of $N_{\mathrm{c}} = 5$ was necessary. Longer control horizons tended to result in more variance in the control signals.

The computation times on a 500 MHz Pentium III PC, were typically around 6 s and maximal 10 s for one MPC iteration in the 'coordinated speed limits and ramp metering' case (which is the most demanding case), with the controller implemented in Matlab and the traffic model in C. Since the controller sampling time is 1 min this is fast enough for real-time implementation.

6.2 Coordinated ramp metering and main-stream metering

Since one of the effects of a speed limit is that it holds back the demand, the comparison of ramp metering in combination with main stream metering instead of speed limits is also interesting. The main advantage of main-stream metering is that it can limit the flow stronger than dynamic speed limits. However, main-stream metering has also some disadvantages compared to speed limits:

1. Dynamic speed limit displays are already installed on most freeways[4], while main-stream metering requires the installation of new equipment.

2. Main-stream metering operates on only one location, while speed limits can operate on many consecutive segments on the freeway. When limiting the flow on one location only, the risk of creating a shock wave is higher, while coordinated speed limits can prevent the creation of shock waves (see [53]).

[4] In The Netherlands, but also in many other countries.

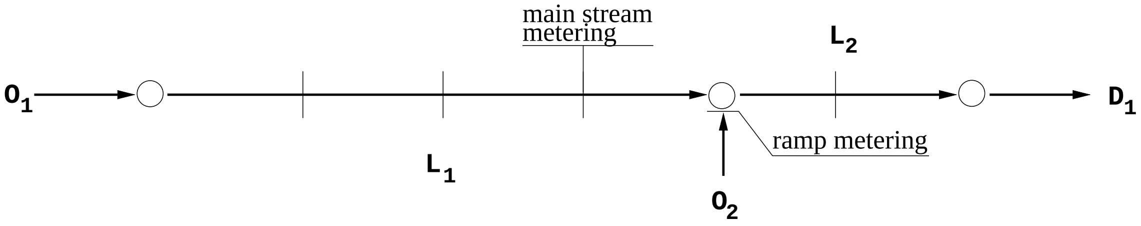


Figure 6.8: The benchmark network with main-stream metering and a metered on-ramp.

3. When main-stream metering is active the maximum flow is bounded by minimum red (and amber) times to approximately 75 % of the freeway capacity (as explained in Section 2.1.2), so there is no possibility to take control actions that result in flows in the range of 75 %–100 % of the freeway capacity.

An advantage of main-stream metering is that it can limit the flow much more than speed limits, which can be useful in very severe situations.

6.2.1 Benchmark problem

In this section we show the results of the 'coordinated main-stream metering and ramp metering' case for three different boundary conditions. The layout of the network is shown in Figure 6.8, which is practically the same as Figure 6.3, but now with main-stream metering instead of speed limits. The main-stream metering is chosen to operate on segment 3 of link 1, because that is the segment on which the speed limits have the largest effect (see Section 6.1.2).

In the first simulation the lower bound $\mathrm{LB}_{\mathrm{msm}}$ of the main-stream metering signal is chosen such that it results in a flow equivalent to the lower bound of the speed limits used in the simulations of the 'coordinated speed limit and ramp metering' case, $\mathrm{LB}_{\mathrm{msm}} = 0.62$ (see Figure 6.9). The interpretation of the main-stream metering signal behavior is very similar to the speed limit in the 'coordinated speed limits and ramp metering' case: main-stream metering becomes active when ramp metering is unable to keep the on-ramp segment of the freeway congestion free. The TTS is 1241.4 veh.h, which means an improvement of 15.0 % compared to the 'no control' case.

In the second case the lower bound of the main-stream metering was set to a low value ($\mathrm{LB}_{\mathrm{msm}} = 0.2$) (see Figure 6.10). This allows to fully utilize the advantage of main-stream metering over speed limits: main-stream metering can limit the flow much more. The TTS in this case is 1206.7 veh.h, which is an improvement of 17.4 % with respect to the 'no control' case. This performance is — as can be expected — better than in the first case.

When main-stream metering is active the maximum flow is bounded by minimum red (and amber) times to approximately 75 % of the freeway capacity, so there is no possibility to take control actions that result in flows in the range of 75 %–100 % of the freeway

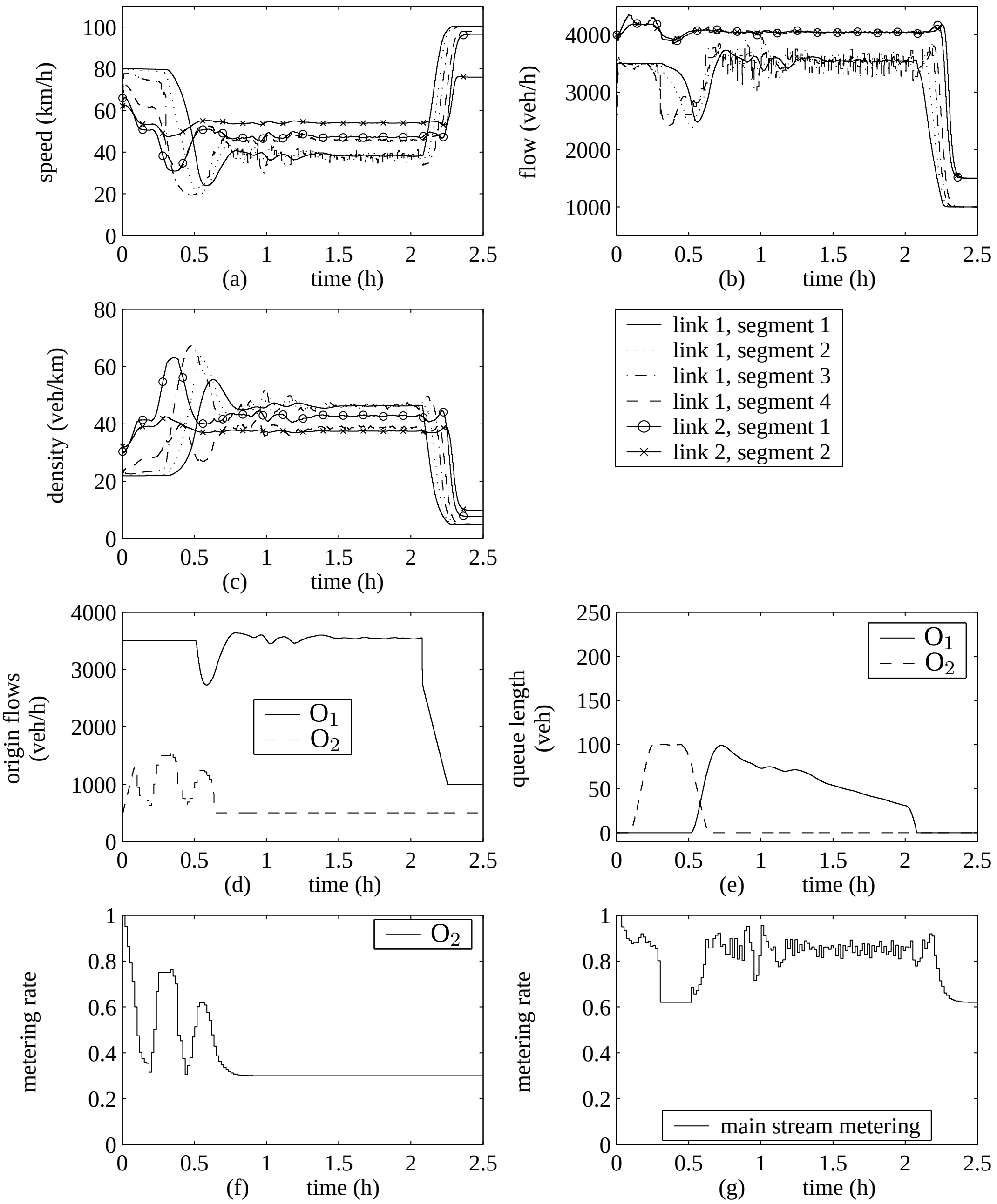


Figure 6.9: The results for the 'coordinated main-stream metering and ramp metering'. The behavior of the main stream metering is similar to behavior of the speed limits in the 'coordinated speed limits and ramp metering' case.

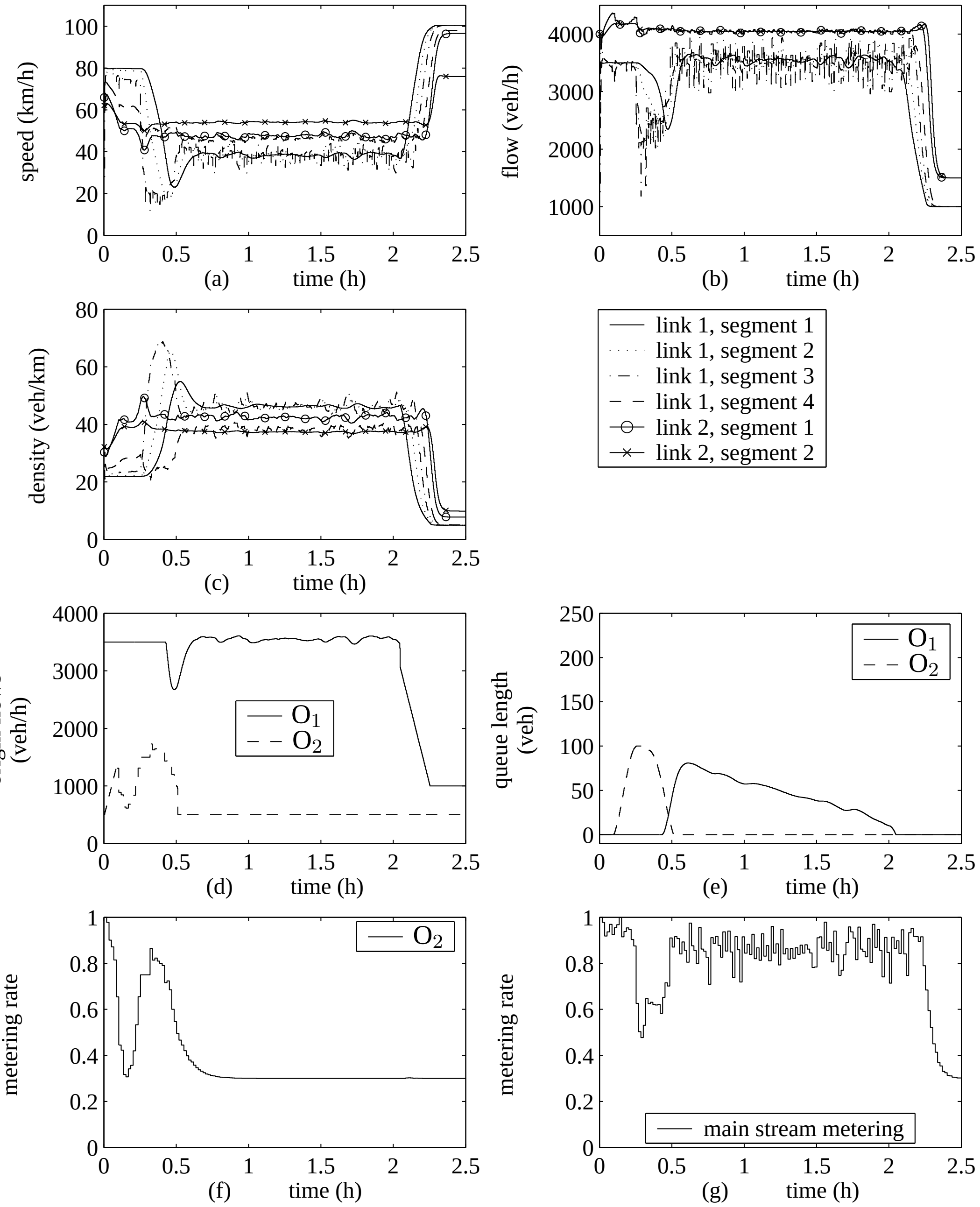


Figure 6.10: The results for the 'coordinated main-stream metering and ramp metering' with lower bound of $r_{\mathrm{msm}}(k_{\mathrm{c}})$ *set to* $\mathrm{LB}_{\mathrm{msm}} = 0.2$.

capacity. Therefore, we include a third simulation, where we take into account the on/off switching of the main-stream metering. So, the outflow of segment 3 of link 1 is unconstrained when main-stream metering is off, and is limited to $\mathrm{UB}_{\mathrm{msm}} q_{\mathrm{cap},m,i}$ (we take here 75 % of the nominal capacity of the freeway: $\mathrm{UB}_{\mathrm{msm}} = 0.75$) when main-stream metering is on. To approximate the on/off switching a similar approach is used as in [53]: the real-valued control signal $r_{\mathrm{msm}}(k_{\mathrm{c}})$ that results from one MPC iteration is rounded as

$$r_{\mathrm{msm,rounded}}(k_{\mathrm{c}}) = \begin{cases} 1 & \text{if} \quad (1 + \mathrm{UB}_{\mathrm{msm}})/2 \leq r_{\mathrm{msm}}(k_{\mathrm{c}}) \\ \mathrm{UB}_{\mathrm{msm}} & \text{if} \quad (1 + \mathrm{UB}_{\mathrm{msm}})/2 > r_{\mathrm{msm}}(k_{\mathrm{c}}) \geq \mathrm{UB}_{\mathrm{msm}} \\ r_{\mathrm{msm}}(k_{\mathrm{c}}) & \text{otherwise,} \end{cases}$$

where $\mathrm{UB}_{\mathrm{msm}} \in [0, 1]$ represents the highest metering rate, achieving the maximum flow during main-stream metering and $r_{\mathrm{msm,rounded}}(k_{\mathrm{c}})$ is applied to the traffic process. In Figure 6.11 the simulation results are presented. The TTS is 1224.4 veh.h which is an improvement of 16.1 %. This performance is very good. However, the controller oscillates (between switching on and off the main-stream metering) when the optimal metering rate would be somewhere in $[UB_{\mathrm{msm}}, 1]$ (compare this with the second case in Figure 6.10(g)). Such a frequent on/off switching is undesirable, because it may create shock waves and it may be unsafe. Therefore, in cases where only a limited flow reduction is necessary, speed limits are recommended.

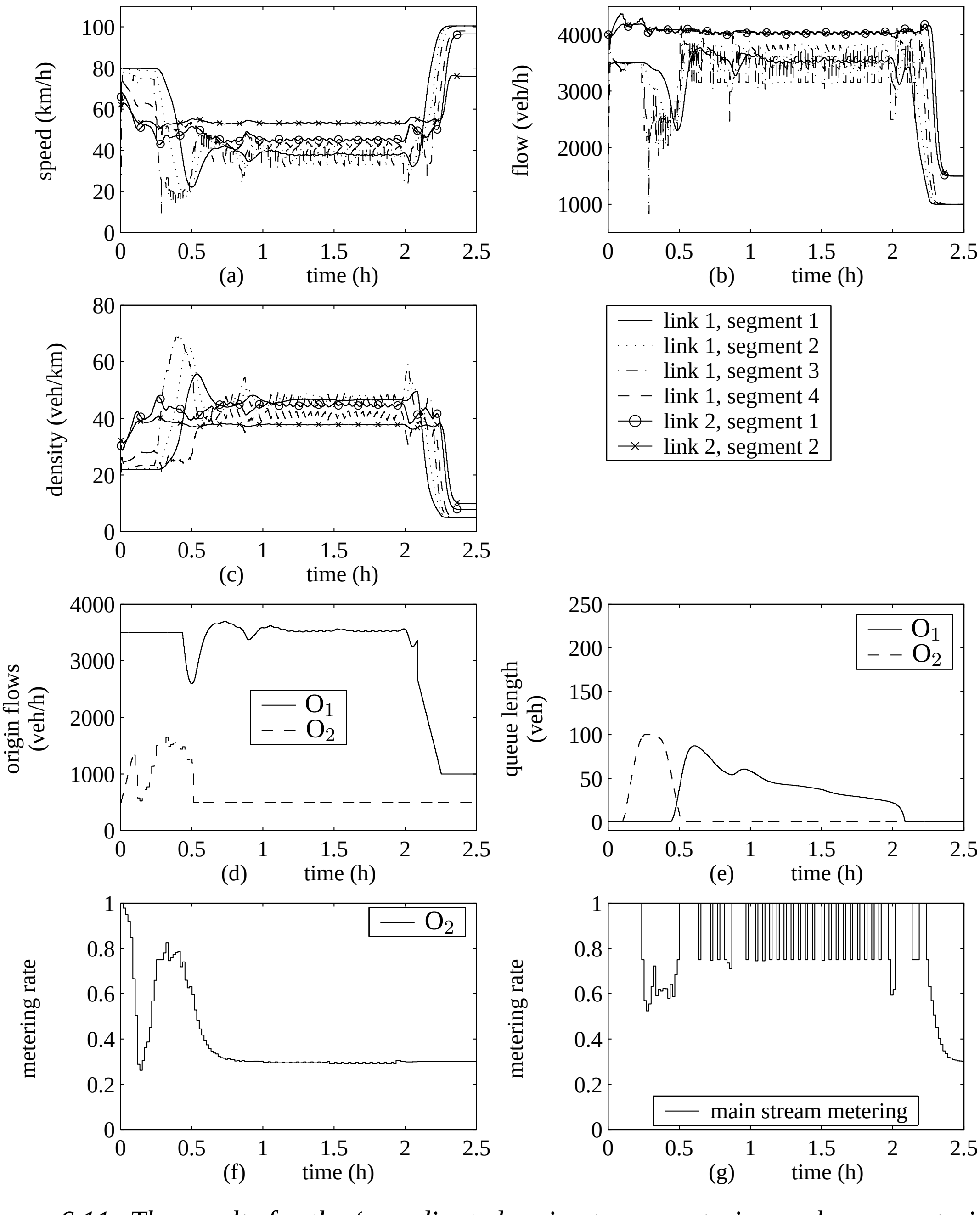


Figure 6.11: The results for the 'coordinated main-stream metering and ramp metering' with lower bound of $r_{\mathrm{msm}}(k_{\mathrm{c}})$ set to $\mathrm{LB}_{\mathrm{msm}} = 0.2$, and taking the maximal possible flow into account when main-stream metering is on. The controller shows oscillatory behavior when the optimal metering rate is between the upper bound and 1.

6.3 Shock wave reduction/elimination with coordinated variable speed limits

When freeway traffic is dense, shock waves may appear. These shock waves result in longer travel times and in sudden, large variations in the speeds of the vehicles, which could lead to unsafe situations. Dynamic speed limits can be used to eliminate or at least to reduce the effects of shock waves. However, coordination of the variable speed limits is necessary in order to prevent the occurrence of new shock waves and/or a negative impact on the traffic flows in other locations. In this section[5] we present an MPC approach to optimally coordinate variable speed limits for freeway traffic with the aim of suppressing shock waves. First of all, we optimize real-valued speed limits, such that the TTS is minimal. Next, we include a safety constraint that prevents drivers from encountering speed limit drops larger than, e.g., 10 km/h. Furthermore, to get a better correspondence between the computed and the applied control signals, we consider discrete speed limits.

Also in the case of speed limit control, prediction and coordination are necessary for an effective control strategy. Prediction is needed for two reasons: first, if the formation or the arrival of a shock wave in the controlled area can be predicted, then preventive measures can be taken. Second, the positive effect of speed limits on the traffic flow cannot be observed instantaneously,[6] so the prediction should at least include the point where the improvement can be observed.

Besides prediction and coordination, the speed limit control problem has other characteristics that impose certain requirements on the control strategy:

1. The speed limit signs used in practice display speed limits in increments of, e.g., 10 or 20 km/h. Therefore, the controller should produce discrete control signals.

2. For safety it is often required that the driver should not encounter a decrease in the displayed speed limit larger than a prespecified amount.

The control strategy presented in this section takes these two requirements into account.

6.3.1 Problem statement

It is well known (see, e.g., [85]) that some type of traffic jams move upstream with approximately 15 km/h. These jams can remain stationary for a long time, so every vehicle that enters the freeway upstream of the jammed area will have to pass the jammed area, which increases the travel time. Besides the increased travel time another disadvantage of the moving jams is that they are potentially unsafe.

[5] The material presented in this section is the result of the joint work with Pascual Breton.

[6] We will see that the speed limits have to slow down a part of the traffic first in order to dissolve the shock wave.

Lighthill and Whitham [104] introduced the term ‘shock wave’ for waves that are formed by several waves running together. At the shock wave, fairly large reductions in velocity occur very quickly. In this section we use the term ‘shock wave’ for any wave (the moving jammed areas), and we do not distinguish between waves and shock waves, because in practice any wave is undesired.

To suppress shock waves one can use speed limits in the following way. In some sections upstream of a shock wave, speed limits are imposed and consequently the inflow of the jammed area is reduced. When the inflow of the jammed area is smaller than its outflow, the jam will eventually dissolve. In other words, the speed limits create a low-density wave (with a density lower than in the uncontrolled situation) that propagates downstream. This low-density wave meets the shock wave and compensates its high density. As a result, the shock wave is reduced or eliminated.

A point of criticism could be that the approach reduces the shock wave, but at the cost of creating new shock waves upstream of the sections controlled by speed limits. However, if the speed limits are optimized properly, they will never create a shock wave that gives rise to delays that are higher than in the uncontrolled case. This can be explained in terms of the stable, metastable, and unstable traffic flow states observed by Kerner and Rheborn [85]. Stable means that any disturbance (no matter how large) will vanish without intervention. Metastable means that small disturbances will vanish, but large disturbances will create a shock wave. Unstable means that any disturbance (no matter how small) will trigger a shock wave. If speed limits are to dissolve shock waves, the traffic flow must be in the metastable state, because in the stable state there is not much to control, and in the unstable state any speed limit change will initiate a new shock wave. In the metastable state, the speed limits have the possibility (if the change of the speed limit values is sufficiently small) to limit the flow without creating large disturbances. See also Section 5.2 for more details on metastability. In the following sections we demonstrate how the proper speed limits can be found.

Objective function

We consider the following objective function:

$$\begin{aligned} J(k_\mathrm{c}) = T \sum_{k=Mk_\mathrm{c}}^{M(k_\mathrm{c}+N_\mathrm{p})-1} \left\{ \sum_{(m,i)\in I_\mathrm{all}} \rho_{m,i}(k) L_m \lambda_m + \sum_{o\in O_\mathrm{all}} w_o(k) \right\} + \\ \xi_\mathrm{speed} \sum_{\ell=k_\mathrm{c}}^{k_\mathrm{c}+N_\mathrm{c}-1} \sum_{(m,i)\in I_\mathrm{speed}} \left(\frac{v_{\mathrm{control},m,i}(\ell) - v_{\mathrm{control},m,i}(\ell-1)}{v_{\mathrm{free},m}} \right)^2 \quad , \end{aligned}$$

where I_all is the set of indexes of all pairs of segments and links and O_all is the set of all origins, and I_speed is the set of pairs of indexes (m,i) of the links and segments where speed control is applied. This objective function contains a term for the TTS, and a term that penalizes abrupt variations in the speed limit control signal. The variation term is weighted by the nonnegative weight parameter ξ_{speed}.

Constraints

In general, for the safe operation of a speed control system, it is required that the maximum decrease in speed limits that a driver can encounter ($v_{\text{max,diff}}$) is limited. A driver can encounter a new speed limit when he enters the next segment or when the speed limits are recalculated by the controller. Consequently, there are three situations where a driver can encounter a different speed limit value:

1. when the driver remains in the same segment and the speed limit is recalculated (and there are more speed limit signs on the same segment, such that the driver always can perceive the speed limit),
2. when a driver enters the next segment and the speed limit there in the same time step is different,
3. when the driver enters the next segment and the speed limit there is recalculated.

The maximum speed difference constraints for the three situations are formulated as follows:

$$\begin{aligned}
v_{\text{control},m,i}(\ell-1) - v_{\text{control},m,i}(\ell) \leq & v_{\text{max,diff}} && \text{for all } (m,i,\ell) \quad \text{such that} \\
&&& (m,i) \in I_{\text{speed}} \text{ and} \\
&&& \ell \in \{k_{\text{c}},\ldots,k_{\text{c}}+N_{\text{c}}-1\}, \\
v_{\text{control},m,i}(\ell) - v_{\text{control},m,i+1}(\ell) \leq & v_{\text{max,diff}} && \text{for all } (m,i,\ell) \quad \text{such that} \\
&&& (m,i) \in I_{\text{speed}} \text{ and} \\
&&& (m,i+1) \in I_{\text{speed}} \text{ and} \\
&&& \ell \in \{k_{\text{c}},\ldots,k_{\text{c}}+N_{\text{c}}-1\}, \\
v_{\text{control},m,i}(\ell-1) - v_{\text{control},m,i+1}(\ell) \leq & v_{\text{max,diff}} && \text{for all } (m,i,\ell) \quad \text{such that} \\
&&& (m,i) \in I_{\text{speed}} \text{ and} \\
&&& (m,i+1) \in I_{\text{speed}} \text{ and} \\
&&& \ell \in \{k_{\text{c}},\ldots,k_{\text{c}}+N_{\text{c}}-1\}.
\end{aligned}$$

In addition to the safety constraints, the speed limits are often subject to a minimum value $v_{\text{control,min}}$:

$$v_{\text{control},m,i}(\ell) \geq v_{\text{control,min}} \quad \text{for all } (m,i) \in I_{\text{speed}} \text{ and } \ell \in \{k_{\text{c}},\ldots,k_{\text{c}}+N_{\text{c}}-1\}.$$

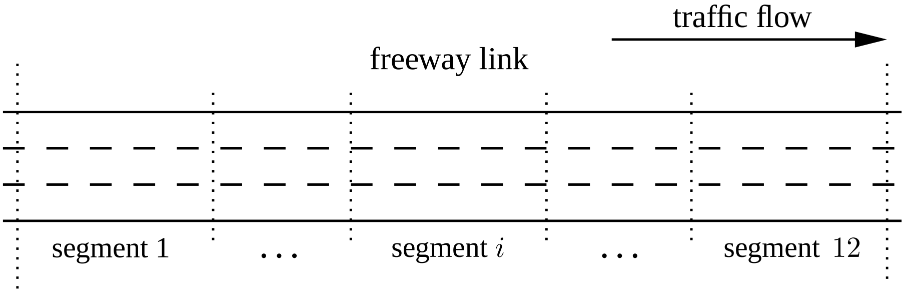


Figure 6.12: The freeway stretch for the benchmark problem consists of 12 segments of 1 km each.

In practice, the variable speed limit signs display speed limits in increments of, e.g., 10 or 20 km/h. Therefore, the controller should produce discrete control signals. This is expressed by the constraint

$$v_{\text{control},m,i}(\ell) \in \mathcal{V}_{m,i} \quad \text{for all } (m,i) \in I_{\text{speed}} \text{ and} \\ \ell \in \{k_{\text{c}}, \dots, k_{\text{c}} + N_{\text{c}} - 1\},$$

where $\mathcal{V}_{m,i}$ is the set of discrete speed limit values in segment i of link m. An example of such a set is $\mathcal{V} = \{50, 60, 70, 80, 90, 100, 110\}$.

6.3.2 Benchmark problem

In order to illustrate the control framework presented above, we will now apply it to a benchmark set-up for two scenarios. The benchmark network consists of a freeway link equipped with variable speed signs. For the first scenario we demonstrate the effectiveness of speed limits against shock waves, explain the tuning of N_{p} and N_{c}, and compare the performance of the controller with and without safety constraints. For the second scenario we show that the results are comparable to the first scenario and we explain the performance degradation that occurs when the real-valued speed limits are rounded downward to get the discrete speed limits.

Set-up with scenario 1

The benchmark set-up consists of one origin, one freeway link of 12 km, and one destination, as shown in Figure 6.12. The mainstream origin has two lanes with a capacity of 2000 veh/h each. The freeway link has two lanes, and consists of twelve segments of 1 km each. Segments 1 up to 5 and segment 12 are uncontrolled, while segments 6 up to 11 are equipped with a VMS where speed limits can be set. We choose to include the five uncontrolled upstream segments to be sure that the upstream boundary conditions do not play a dominant role. We use the same network parameters as in [93]: $T = 10\,\text{s}$, $\tau = 18\,\text{s}$,

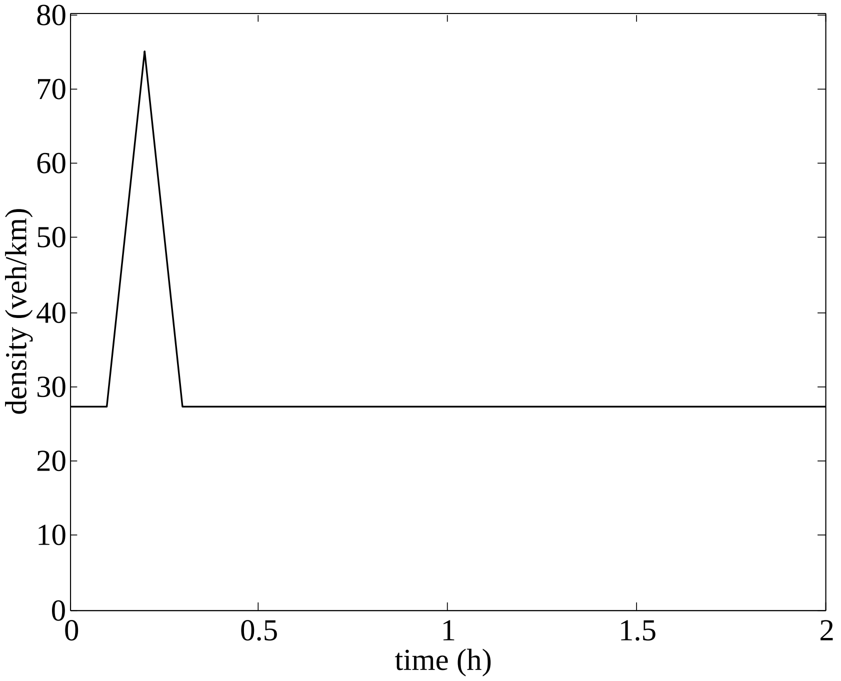


Figure 6.13: The downstream density profile for scenario 1.

$\kappa = 40$ veh/km/lane, $\rho_{\max} = 180$ veh/km/lane, $\rho_{\mathrm{crit},m} = 33.5$ veh/km/lane, $a_m = 1.867$ and $v_{\mathrm{free},m} = 102$ km/h.

Furthermore, we take $\eta_{\mathrm{high}} = 65\,\mathrm{km}^2$/h, $\eta_{\mathrm{low}} = 30\,\mathrm{km}^2$/h, $\alpha = 0.05$ and $\xi_{\mathrm{speed}} = 2$. For the variable speed limits we assume that they can change only every minute, and that they cannot be less than $v_{\mathrm{control,min}} = 50$ km/h. This is imposed as a hard constraint in the optimization problem. If there is a safety constraint, then $v_{\mathrm{max,diff}} = 10$ km/h. The input of the system is the traffic demand at the upstream end of the link and the (virtual) downstream density at the downstream end of the link. The traffic demand (inflow) has a constant value of 3900 veh/h, close to capacity (4000 veh/h). The downstream density equals the steady-state value of 28 veh/km, except for the pulse that represents the shock wave. The pulse was chosen large enough to cause a back-propagating wave in the segments, (see Figures 6.13 and the top of Figure 6.14). It is assumed that the upstream demand and downstream density is known, or predicted by an external algorithm. In practice, a combination of traffic measurements outside the controlled area and historical data could be used for prediction.

For the above scenario the tuning of N_{p} and N_{c} will be demonstrated, and the performance (TTS) of the real-valued and discrete controls with or without safety constraints are examined. In the discrete control case, the control values $v_{\mathrm{control},m,i}$ are in the set $\mathcal{V}_{m,i} = \mathcal{V} = \{50, 60, 70, 80, 90, 100, 110\}$ for all $(m, i) \in I_{\mathrm{speed}}$.

The solution of the real-valued speed control problem is calculated by the Matlab implementation of the SQP algorithm ‘fmincon’ (see also Section 4.1.5). The discrete control signal is a rounded version of the continuous optimization result. Three different types of discretization are examined: The first (‘round’) rounds the real control values to the nearest discrete value in the set $\mathcal{V}$, the second (‘ceil’) rounds them to the nearest discrete value in $\mathcal{V}$ that is higher than the real value, and the third (‘floor’) to the nearest discrete value in $\mathcal{V}$ that is lower than the real value.

This method of obtaining discrete control signals is heuristic but fast. It is also possible to use discrete optimization techniques such as tabu search [41, 43, 42], simulated annealing [36], or genetic algorithms [44, 31], but since (as we will see) for this set-up and input the discretization method results in a performance that is comparable to that of the real-valued case, it is not necessary to do so.

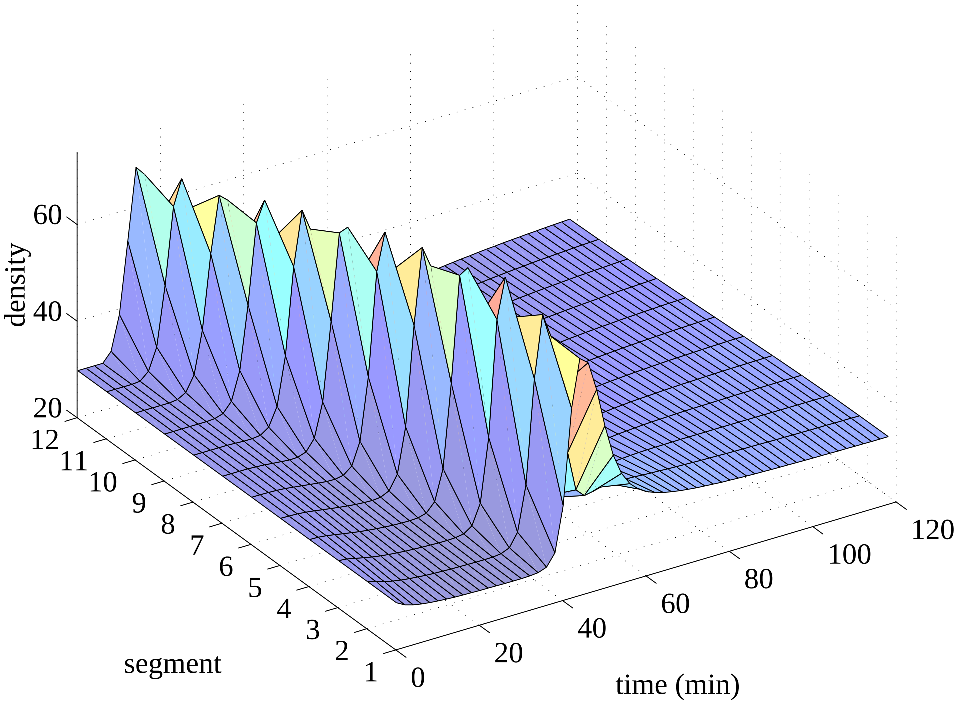
Note that it is not difficult to prove that the result of all of the three types of rounding will satisfy the safety constraints if the real-valued signal satisfies them and if $v_{\mathrm{max,diff}}$ is a multiple of the discretization step of the speed limits (here: a multiple of 10 km/h). Since it does not make much sense to set $v_{\mathrm{max,diff}}$ to another value than a multiple of the discretization step, this condition should not be a limitation.

The rolling horizon strategy is now implemented as follows. After the discretization, the first sample of the control signal is applied to the traffic system, and then the optimization–discretization steps are repeated. Note that this approach does not yield the same evolution and control signals as an approach in which first the real-valued signal is computed (using the rolling horizon approach) for the entire simulation period at once, rounded, and then applied for the whole simulation period. This is because in the first approach the different traffic behavior caused by the discretization is already taken into account in the each subsequent MPC iteration.

In the next section we will compare the performance of the discrete control to the performance achieved by the real-valued control without constraints, and the effect of introducing the safety constraints is examined.

Results

The results of the simulations of the ‘no control’ and the control case with real-valued speed limits without constraints are displayed in Figure 6.14. In the controlled case the shock wave disappears after approximately 90 min, while in the ‘no control’ case, the shock wave travels through the whole link. The speed limits are active in segments 6 up to 10; the speed limit in segment 11 has higher values than the critical speed and is not limiting the flow (see Figure 6.15). The active speed limits start to limit the flow at $t = 5$ min and create a low-density wave traveling downstream (the small dip in Figure 6.14 (bottom) and in the zoom-in version, Figure 6.16). This low-density wave meets the shock wave traveling upstream and reduces its density just enough to stop it. So, the tail of the shock wave has a fixed location while the head dissolves into free flow traffic, which means that the shock wave eventually dissolves completely.



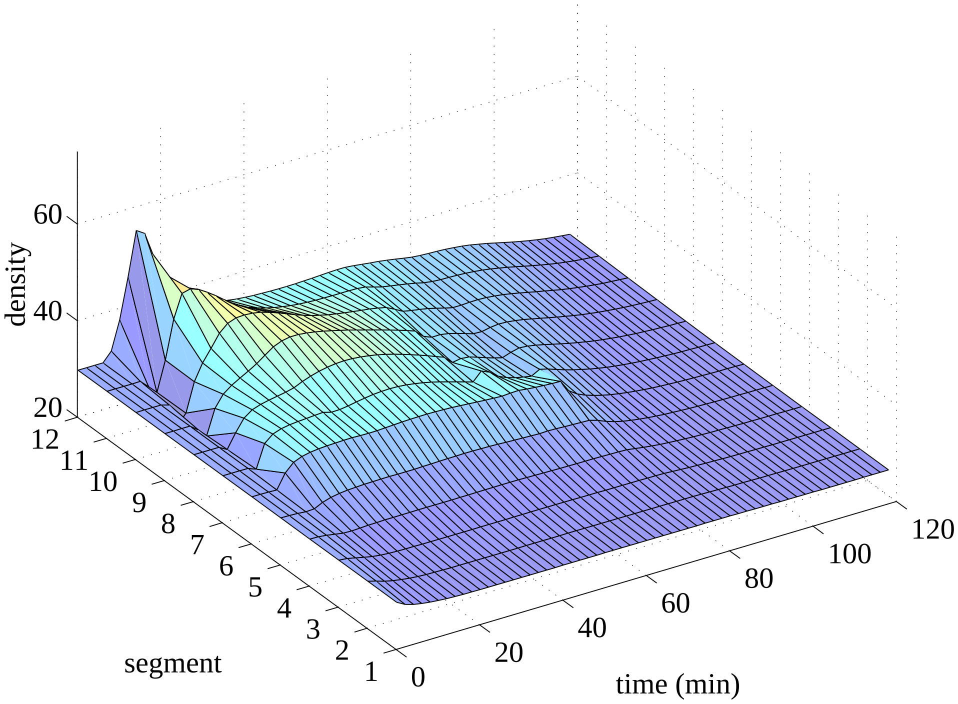


Figure 6.14: The shock wave propagates through the link in the 'no control' case (top). In the 'coordinated control' case, the shock wave disappears after approximately 90 min (bottom).

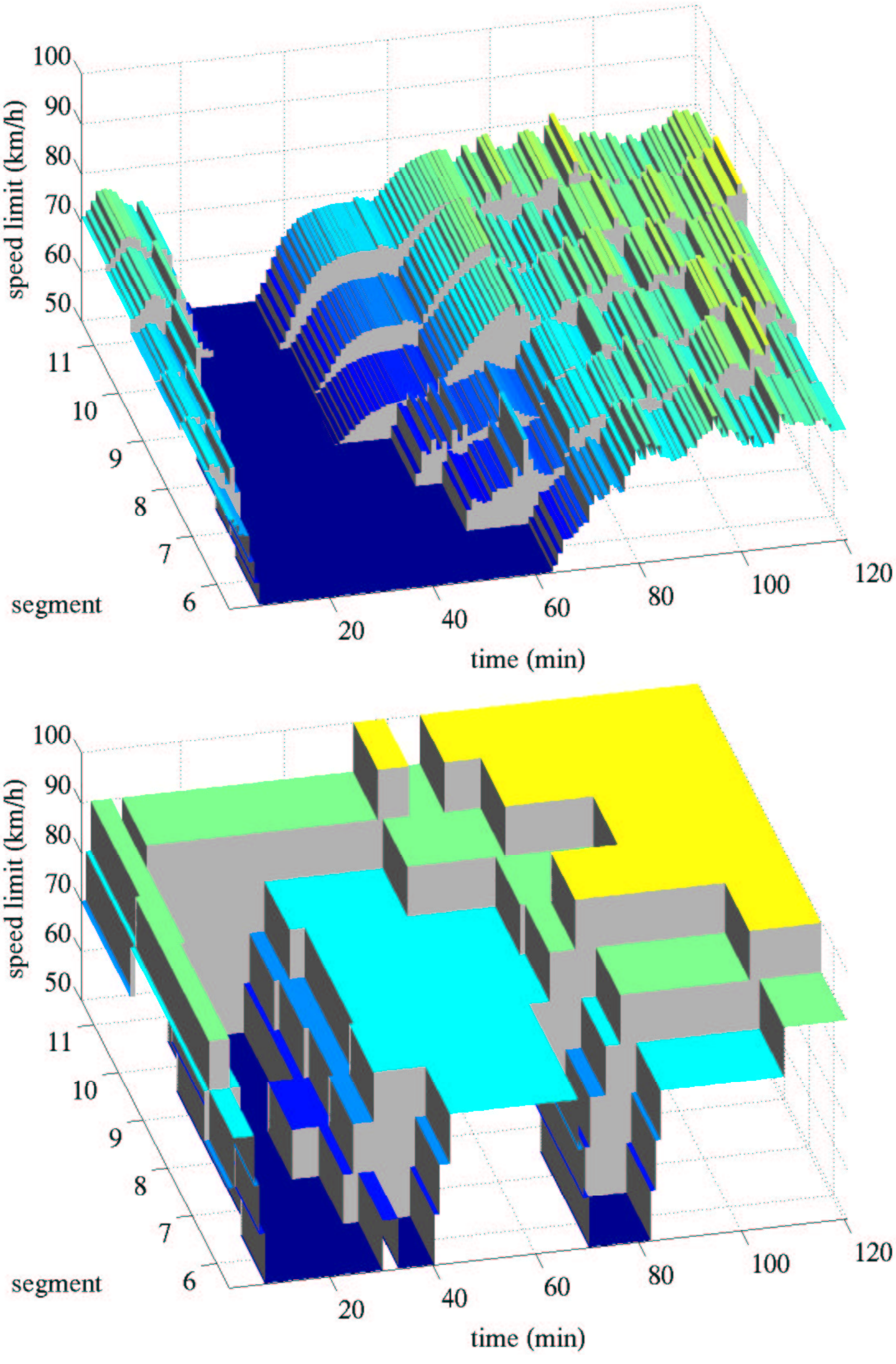


Figure 6.15: The speed limits for the real-valued case without safety constraints and $N_\mathrm{p} = 10$, $N_\mathrm{c} = 8$ *(top). The speed limits for the discrete ('ceil') case with safety constraints and* $N_\mathrm{p} = 10$, $N_\mathrm{c} = 8$.

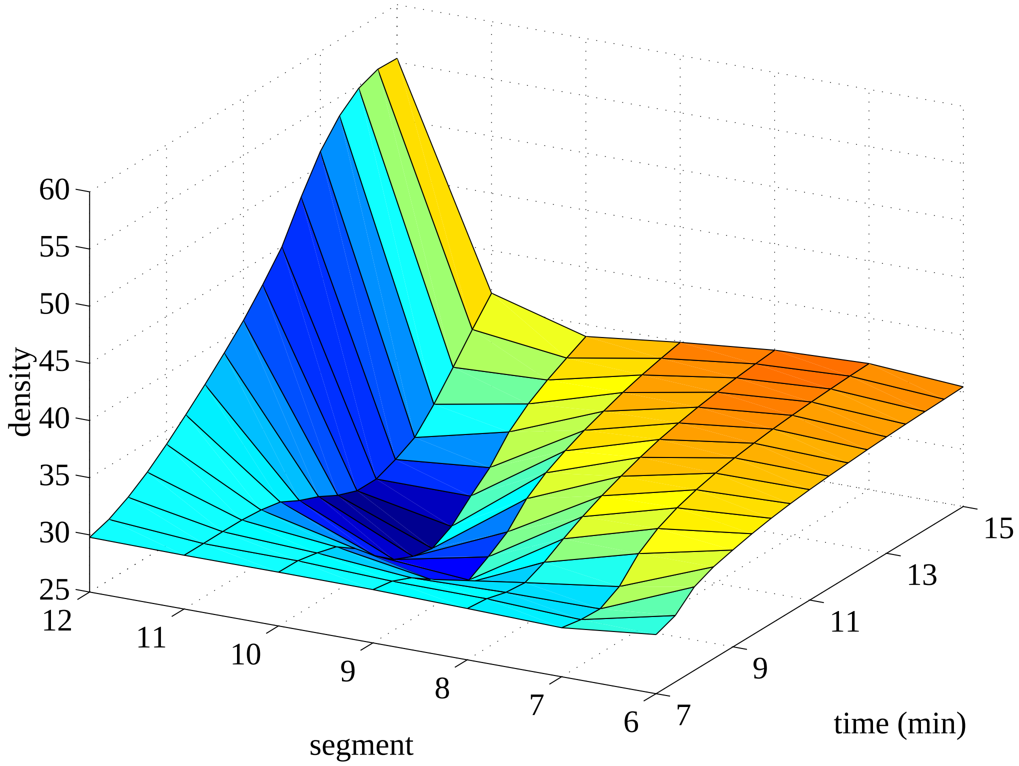


Figure 6.16: Zoom-in on the dip of Figure 6.14 (bottom) for the coordinated control case.

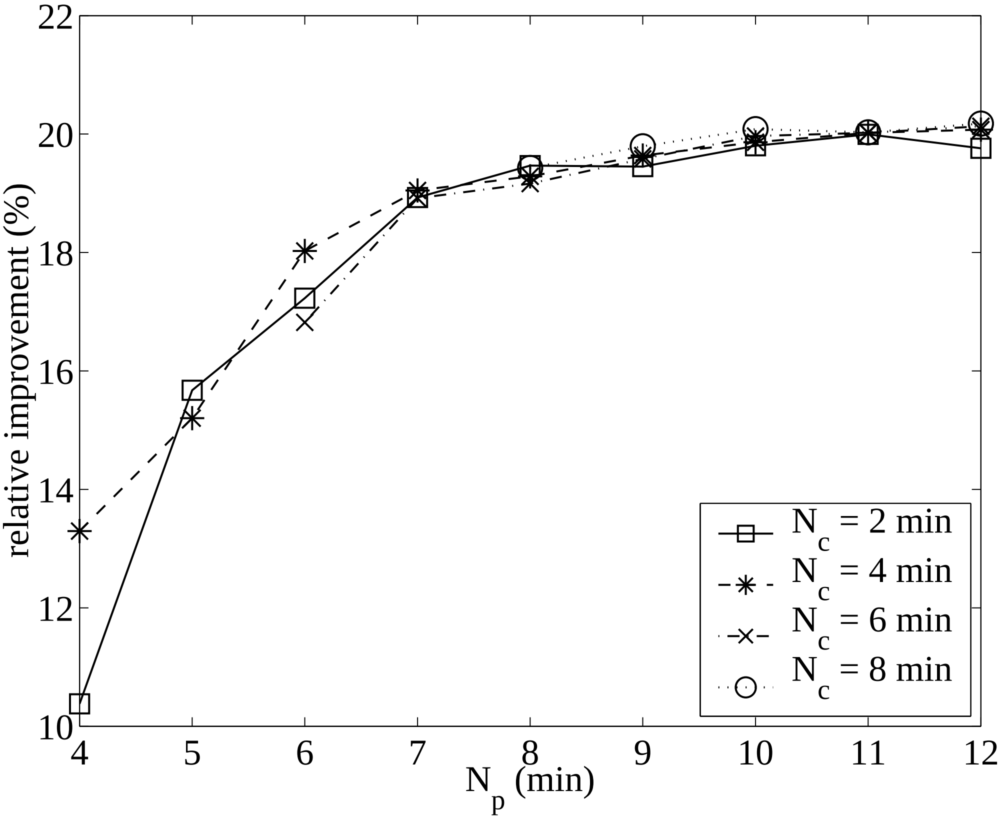


Figure 6.17: The relative improvement of the performance (TTS) for scenario 1 in the real-valued, unconstrained case compared to the 'no control' case as a function of N_p *for several values of* N_c*. The sensitivity to* N_p *is much higher than that to* N_c*.*

The speed limits persist until the shock wave (to be precise, the high-density region) is completely dissolved. The speed limits in Figure 6.15 start to increase after $t = 35$ min and return gradually to a high value that is not limiting the flow anymore.

The TTS was 1835.3 veh.h in the 'no control' case and 1466.7 veh.h in the controlled (real-valued, unconstrained) case, which is an improvement of 20.1 %.

The relative improvement of the performance as a function of N_p and N_c is shown in Figure 6.17. The performance depends stronger on N_p, but for $N_\mathrm{p} \geq 10$ (which is somewhat larger than the maximum travel time from segment 6 to the exit of the network as argued in Section 4.2.5) the graphs become nearly flat. We select $N_\mathrm{p} = 10$ and $N_\mathrm{c} = 8$ for the further analysis.

The result of the several types of discretization is shown in Table 6.1 for the simulations without safety constraints and in Table 6.2 for the simulations with safety constraints. The performance loss caused by the discretization is small in the 'round' and 'ceil' cases, but large for 'floor'. The cause for this performance degradation in the latter

Horizon		Relative improvement (%)			
N_p	N_c	real-valued	round	ceil	floor
9	4	19.6	17.5	17.9	-2.2
9	6	19.6	19.1	18.9	3.9
9	8	19.8	15.0	17.6	6.9
10	4	19.9	17.9	19.6	-1.1
10	6	20.0	19.6	19.3	2.9
10	8	20.1	15.2	18.3	5.9
11	4	20.0	18.0	19.8	-1.1
11	6	20.0	17.7	19.8	1.3
11	8	20.0	19.9	19.4	5.5
12	4	20.1	15.5	20.0	-2.2
12	6	20.1	19.7	20.0	1.3
12	8	20.2	19.8	20.0	5.7

Table 6.1: The relative improvement with respect to the ‘no control’ case of the performance (TTS) for scenario 1 for several combinations of N_p and N_c, and for the real-valued speed limits and the three discrete speed limits: ‘round’, ‘ceil’, and ‘floor’; without safety constraints.

case will be explained in Section 6.3.2.

In the other cases the inclusion of the constraints result in a small performance loss, which is in accordance with the general expectation that the introduction of extra constraints usually results in lower performance.

The performance improvement for $N_\mathrm{p} = 10$, $N_\mathrm{c} = 8$ in the constrained ‘ceil’ case is 17.3 %, which is very close to the improvement of the unconstrained ‘ceil’ case (18.3 %), and even comparable to the improvement of 20.1 % in the unconstrained real-valued case. Figure 6.15 (bottom) shows the values of the optimal speed limits discrete (‘ceil’) case with safety constraints and $N_\mathrm{p} = 10$, $N_\mathrm{c} = 8$.

Finally, the controller was implemented in Matlab and the traffic model in C. For this implementation the computation for 2 h of simulated time varied between 3 and 25 min on a 500 MHz Pentium III PC, which is at least four times faster than real time[7].

[7] For a full implementation in C an extra speed-up is expected. As a consequence MPC would also be suitable for larger networks, more control measures, or longer control and prediction horizons. Note, however, that the control of larger networks is a topic for future research.

Horizon		Relative improvement (%)			
N_p	N_c	real-valued	round	ceil	floor
9	4	19.4	16.4	18.0	0.2
9	6	19.5	19.3	19.0	12.3
9	8	19.4	18.4	11.4	11.9
10	4	19.5	15.5	18.5	1.4
10	6	19.6	19.4	18.0	9.0
10	8	19.7	19.1	17.2	11.0
11	4	19.6	15.4	18.2	0.4
11	6	19.7	19.8	19.6	7.3
11	8	19.9	19.7	19.3	5.5
12	4	19.7	14.7	19.3	1.8
12	6	19.9	19.9	19.7	12.5
12	8	19.9	19.3	19.6	13.4

Table 6.2: The relative improvement with respect to the ‘no control’ case of the performance (TTS) for scenario 1 for several combinations of N_p and N_c, and for the real-valued speed limits and the three discrete speed limits: ‘round’, ‘ceil’, and ‘floor’; with safety constraints.

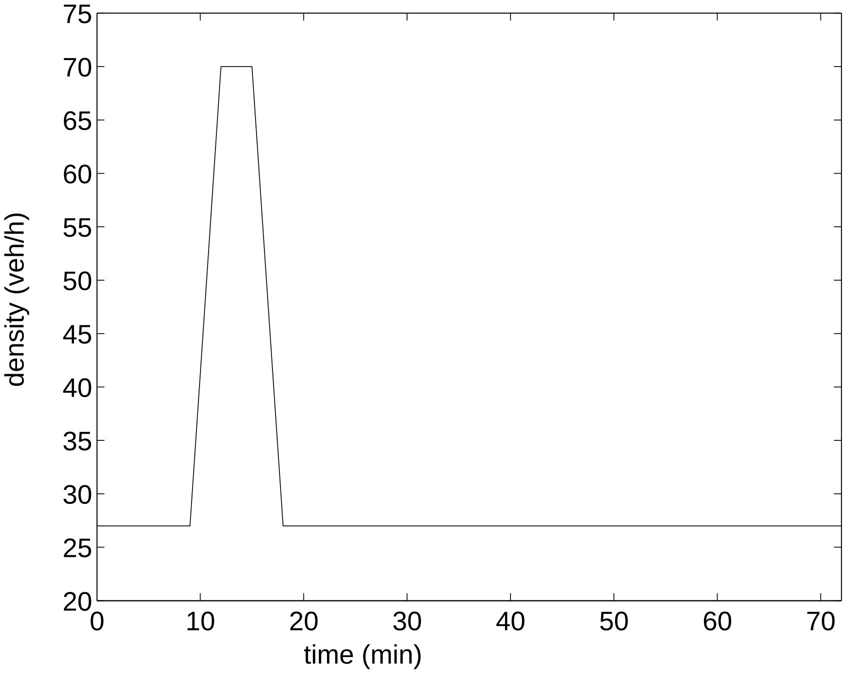


Figure 6.18: The downstream density profile for scenario 2.

Set-up with scenario 2

We use the same set-up as for scenario 1 but with another shock wave scenario, which is shown in Figure 6.18.

Results

The results of the simulations of the 'no control' and the control case with real-valued speed limits and without constraints are displayed in Figures 6.19. In the controlled case the shock wave disappears after approximately 2 h, while in the 'no control' case, the shock wave travels through the whole link. The speed limits are active in segments 6 up to 10; the speed limit in segment 11 has higher values than the critical speed and is not limiting the flow (see Figure 6.20). The active speed limits start to limit the flow at time $t = 4$ min and create a low-density wave traveling downstream (see the small dip in Figure 6.19). This low-density wave meets the shock wave traveling upstream and reduces its density just enough to stop it. So, the tail of the shock wave has a fixed location while the head dissolves into free flow traffic as in the uncontrolled situation, which means that the shock wave eventually dissolves completely.

The speed limits persist until the shock wave (to be precise, the high-density region) is completely dissolved. The speed limits in Figure 6.20 start to increase after $t = 17$ min and return gradually to a high value that is not limiting the flow anymore.

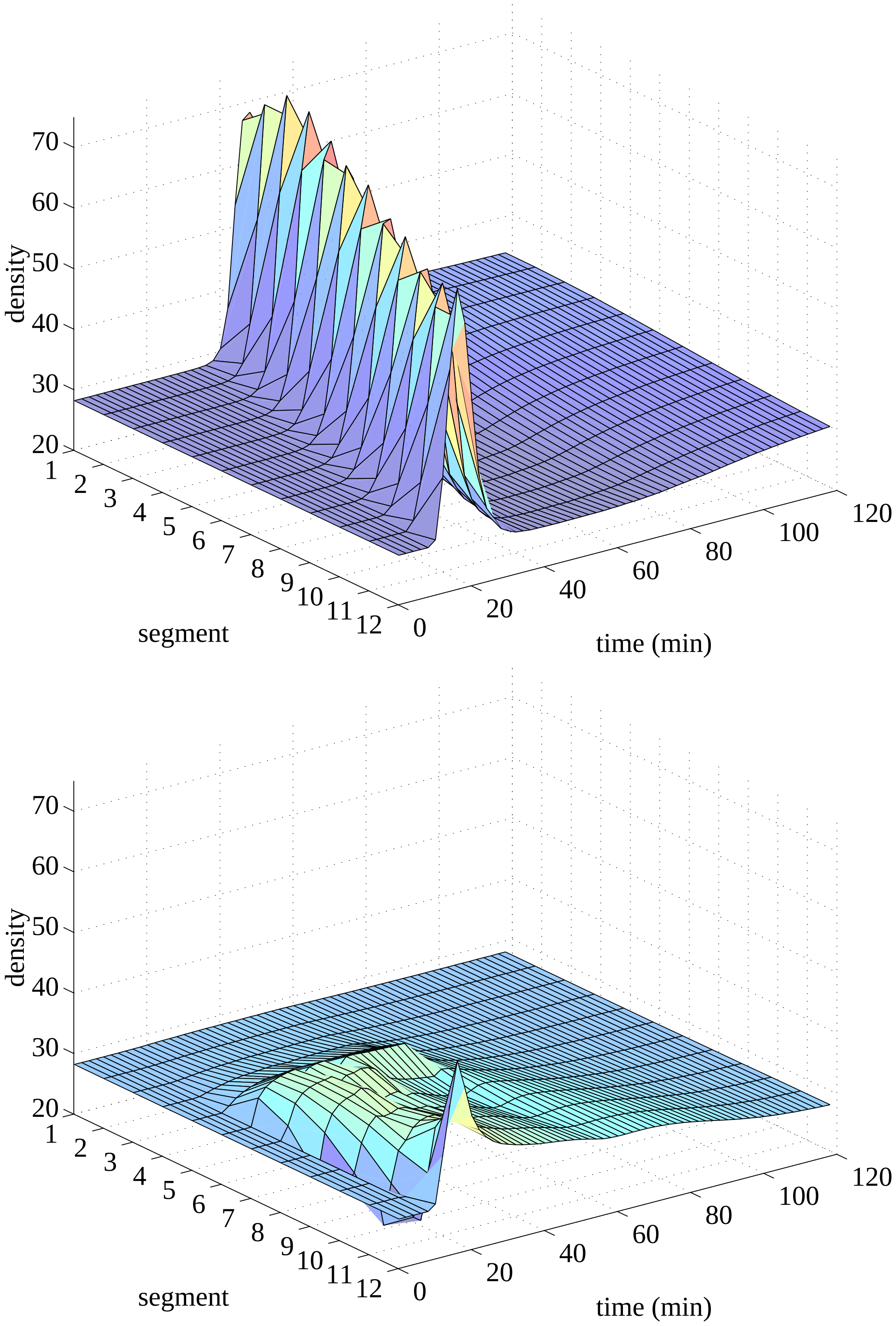


Figure 6.19: The shock wave propagates through the link in the no control case (top). In the coordinated control case, the shock wave disappears after approximately 2 h (bottom).

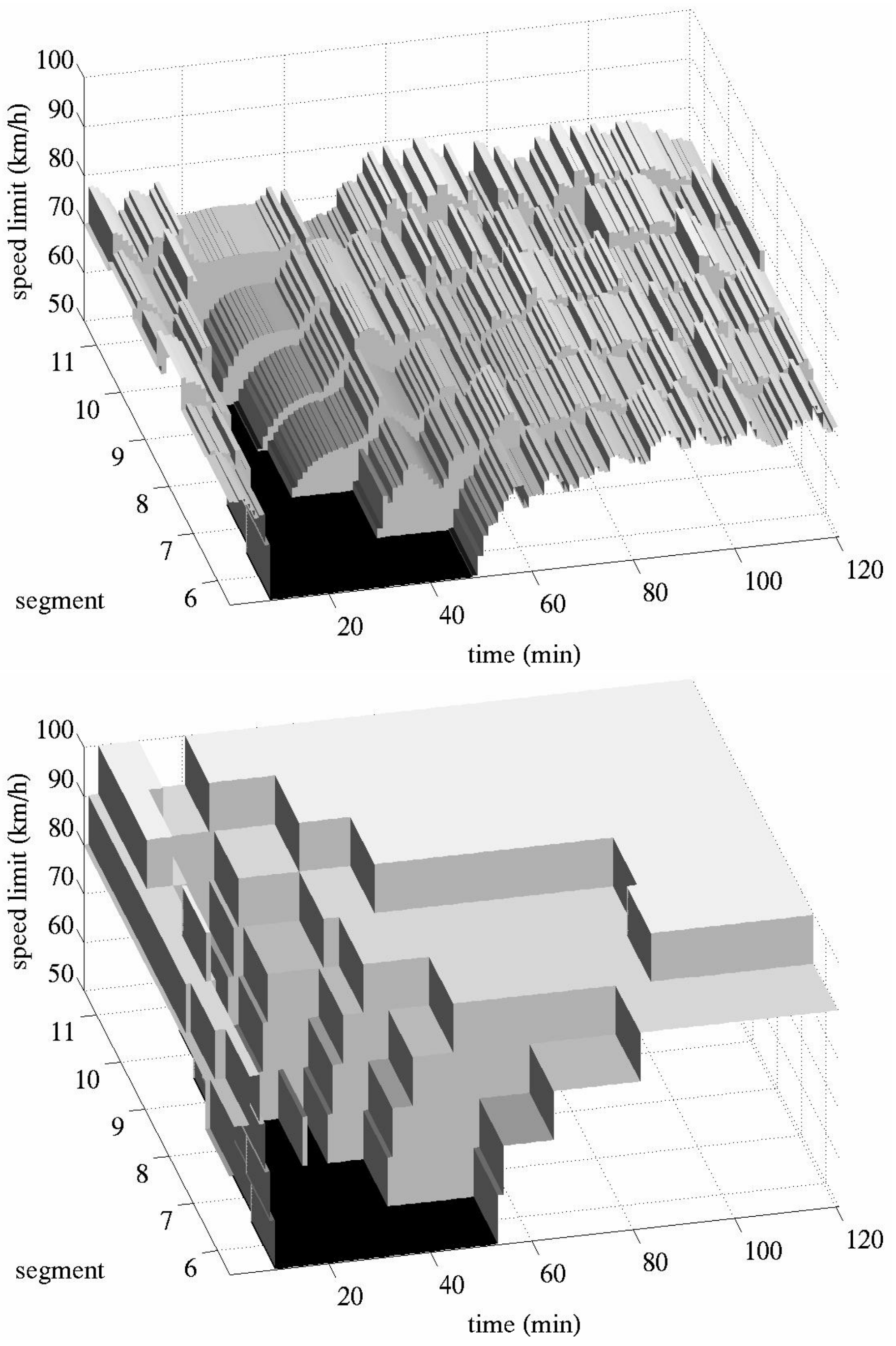


Figure 6.20: The speed limits for the real-valued case without safety constraints and $N_{\mathrm{p}} = 11$*,* $N_{\mathrm{c}} = 8$ *(top). The speed limits for the discrete (ceil) case with safety constraints and* $N_{\mathrm{p}} = 11$*,* $N_{\mathrm{c}} = 8$*. For the purpose of visibility, the travel direction is opposite to that in Figure 6.19.*

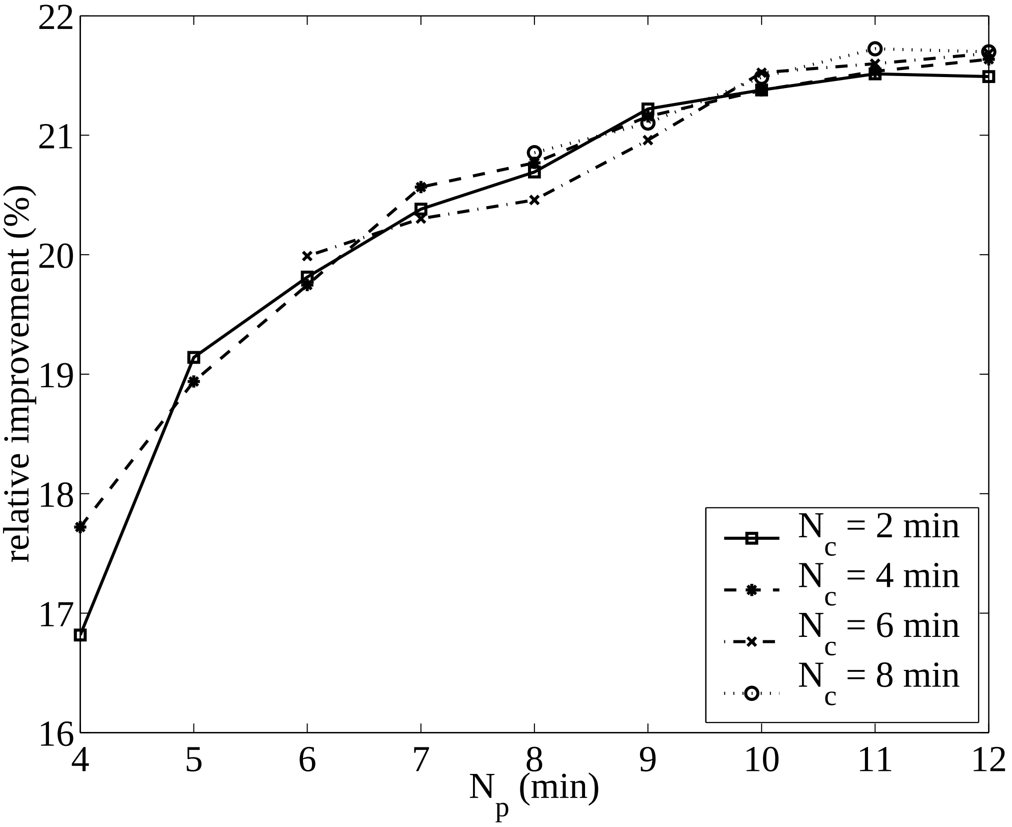


Figure 6.21: The relative improvement of the performance (TTS) for scenario 2 in the real-valued, unconstrained case compared to the 'no control' case as a function of N_p for several values of N_c. The sensitivity to N_p is much higher than that to N_c.

The TTS was 1862.0 veh.h in the 'no control' case and 1458.0 veh.h in the controlled (real-valued, unconstrained) case, which is an improvement of 21.7 %.

The relative improvement of the performance as a function of N_p and N_c is shown in Figure 6.21. The performance depends stronger on N_p, but for $N_\mathrm{p} \geq 10$ (which is somewhat larger than the maximum travel time from segment 6 to the exit as argued in Section 4.2.5) the graphs become nearly flat. For further analysis we chose $N_\mathrm{p} = 11$ and $N_\mathrm{c} = 8$.

The result of the several types of discretization is shown in Table 6.3. The performance loss caused by the discretized speed limits is small in the 'round' and 'ceil' cases, but large for 'floor'. The performance degradation in case of 'floor' can be explained by the slow dynamics of the traffic process, and the detailed explanation is given in Section 6.3.3.

The results of including the safety constraints are comparable to the results without safety constraints, see Table 6.3. Figure 6.20 shows the values of the optimal speed limits discrete (ceil) case with safety constraints and $N_\mathrm{p} = 11$, $N_\mathrm{c} = 8$.

		Relative improvement (%)							
Horizon		unconstrained				constrained			
N_p	N_c	cont.	round	ceil	floor	cont.	round	ceil	floor
9	4	21.1	20.2	21.5	1.9	20.9	19.4	18.5	3.9
9	6	20.9	21.1	21.4	3.8	21.1	20.2	21.1	15.1
9	8	21.1	15.9	21.4	12.1	21.1	20.7	20.4	15.1
10	4	21.4	20.1	21.5	0.5	21.2	20.0	21.2	0.2
10	6	21.6	20.5	21.7	14.2	21.4	20.9	21.4	12.7
10	8	21.5	21.1	21.7	5.2	21.4	20.8	21.4	16.3
11	4	21.5	19.8	21.5	-2.8	21.3	19.5	21.3	4.2
11	6	21.6	21.1	21.7	3.7	21.4	21.1	21.4	14.2
11	8	21.7	21.1	21.7	8.1	21.5	21.0	21.7	13.0
12	4	21.6	20.3	21.6	-1.1	21.5	19.0	21.5	-0.3
12	6	21.7	21.3	21.8	7.3	21.5	21.4	21.4	14.2
12	8	21.7	21.6	21.8	8.9	21.5	21.5	21.4	15.0

Table 6.3: The relative improvement of the performance (TTS) for scenario 2 for several combinations of N_p and N_c, and for the real-valued speed limits and the three discrete speed limits: round, ceil, and floor.

Finally, the computation times were similar to scenario 1.

6.3.3 Effects of rounding

In this section we give a more detailed explanation of the performance degradation for ‘floor’ rounding as described in Section 6.3.2. The degradation occurs for the ‘floor’ type of rounding of the real-valued control signal (resulting from the MPC optimization), but does not occur for ‘round’ and ‘ceil’ types of rounding. In this document we show and explain that if the discrete speed limit step size is increased, the degradation occurs in all cases, but for the ‘floor’ type of rounding the most. The performance degradation occurs for both the constrained and unconstrained cases (see Table 6.3). Therefore, in the remainder of this section we will consider unconstrained cases only.

Analysis of the performance degradation for ‘floor’ rounding

The performance degradation in case of ‘floor’ can be explained by the relatively slow dynamics of the traffic process and the step size of the discrete speed limits. To explain this we describe the behavior of the closed-loop system consisting of the traffic model and the MPC-controller, for the case of an arriving shock wave and discrete MPC speed limit control with the ‘floor’ type of rounding.

Initially, when the shock wave enters the link, the flow is restricted by low speed limits (here 50 km/h). When the congestion starts resolving, the (optimized, real-valued) speed limits will increase to enable the traffic to accelerate. These speed limits will be just above[8] the natural evolution[9] of the speed, (such that they are not limiting anymore) or if necessary (determined by the optimization) below the natural evolution of the speed. This value is rounded downward by floor, and in the next MPC iteration the actual speeds will be lower than the speed limit resulting from the continuous optimization. Since the dynamics of the traffic process is relatively slow, the speeds usually do not increase within the controller sampling time (1 min) with more than 10 km/h. This means that 'floor' will result in the same low value, which keeps the average speed and (out)flow low. This process is repeated for each MPC iteration.

In Figure 6.22 a snapshot of the MPC procedure for speed limits rounded with 'floor' is shown (with $N_\mathrm{p} = 11$, $N_\mathrm{c} = 8$). For visibility purposes we show the speed limits for one segment only. The left vertical line is the current time instant, the right vertical line represents the end of the prediction horizon. Between these two lines the speed limit signals are optimized. In case of discrete speed limits the signals between the two vertical lines are approximated by the discrete signals. Figure 6.23 is a zoom-in of Figure 6.22, where we can see that the optimal real-valued speed limit value in segment 11 at the current time (see the left vertical line; this speed limit equals approximately 52 km/h) is higher than one controller time step earlier (50 km/h). However, if 'floor' is used the current speed limit (52 km/h) is rounded to 50 km/h again, because the increase in speed limit is small. Since the continuous optimization rarely results in a speed limit jump (increase) that is larger than the discretization step, 'floor' will tend to round the signals to the same low value.

It is clear that if the step size of the discretized speed limits is smaller, then the probability of repeated downward rounding of the speed limits is smaller, and the performance will be better. To verify this explanation we will now investigate whether the performance of 'floor' improves when the step size of the discrete speed limits is reduced.

We compare several speed limit step sizes with $N_\mathrm{p} = 11$ and $N_\mathrm{c} = 8$. The results are given in Table 6.4. We can conclude from the table that in general the performance improves if the speed limit step size is decreased, and that the performance of 'floor' breaks down first when the speed limit step size is increased. These findings are in accordance with our expectations[10].

[8] Higher speed limits will not occur, because they do not change the traffic behavior, and consequently do not improve the performance.

[9] We mean by natural evolution of the speed the evolution that would occur if no speed limits were present.

[10] It is remarkable that the performance of 'floor' slightly improves when the speed limit step size is increased from 10 km/h to 15 km/h. This may be caused by the specific traffic scenario, or by the fundamental difference between the MPC optimization (for a horizon of length N_p) and optimization for the whole simulation length.

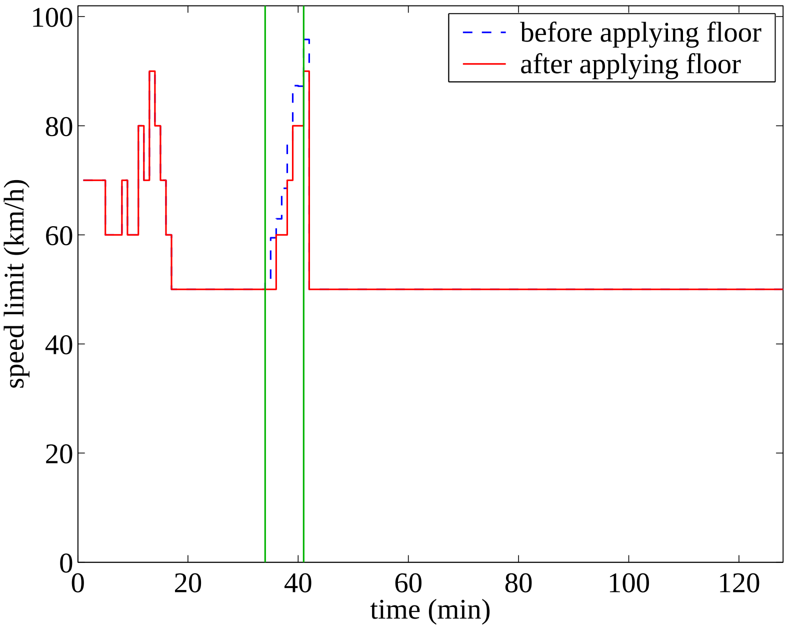


Figure 6.22: Snapshot of the speed limit (in segment 11) resulting from the MPC procedure rounded with floor. The signal is shown before and after rounding. The left and right vertical lines indicate respectively the current time instant and the end of the prediction horizon.

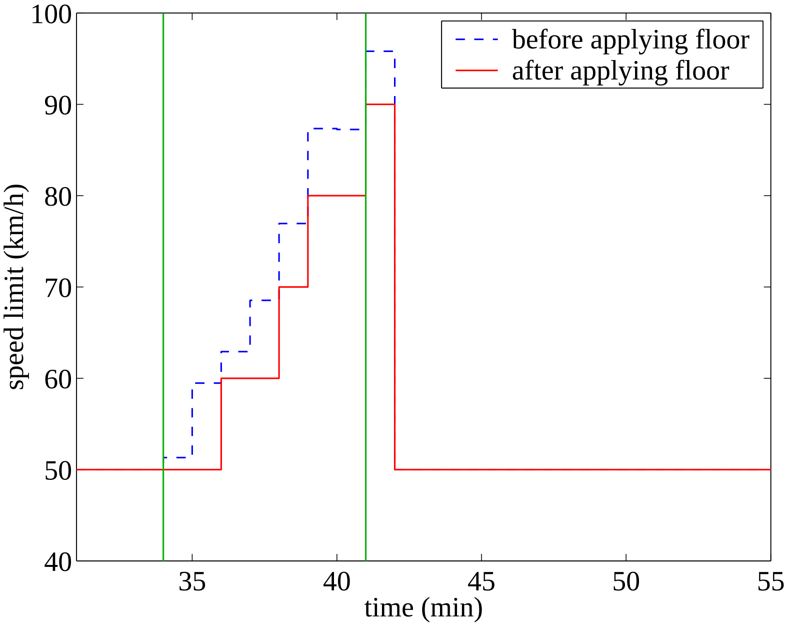


Figure 6.23: Zoom-in of Figure 6.22

ΔSL (km/h)	Relative improvement (%)		
	round	ceil	floor
0 (real-valued)	21.7	21.7	21.7
5	21.7	21.6	18.1
10	20.8	21.7	3.3
15	19.0	2.9	4.5
20	1.6	2.9	-12.8

Table 6.4: The relative improvement of the performance (TTS) for several speed limit step sizes (ΔSL) for the real-valued speed limits and the three discrete speed limits: 'round', 'ceil', and 'floor'.

6.4 Conclusions

We have applied the MPC framework to several traffic problems that can benefit from the use of speed limits. The main purpose of the control was for all problems to find the control signals that minimize the total time that vehicles spend in the network (i.e., TTS).

In Section 6.1 we have dealt with the integrated control of ramp metering and speed limits, where the speed limits can prevent a traffic breakdown when ramp metering only is insufficient. Since the main effect of the speed limits in this section is to limit the flow when necessary, in Section 6.2 this set-up was compared with a set-up where the speed limits are replaced by main-stream metering.

Speed limits proved to be useful when ramp metering was unable to keep the on-ramp segment of the freeway congestion free. This idea was illustrated by a simple example network, where the cases 'ramp metering only' and 'coordinated ramp metering and speed limits' were compared for a typical demand scenario. We found that the coordinated case results in a network that has a higher outflow and a significantly lower TTS. Compared to the 'no control' case the TTS improvement in the 'ramp metering only' case was 5.3 % and in the 'coordinated speed limits and ramp metering' case 14,3 %.

Since the main effect of the speed limits in such situations is that they hold back the traffic, we have also compared the 'coordinated ramp metering and speed limits' scenario with the 'coordinated ramp metering and main-stream metering' scenario (where the speed limits are replaced by a main-stream metering device). The comparison was made for several bounds on the main-stream metering signal. If the bounds were chosen such that the maximal flow limitation is equal to the maximal flow limitation of the speed limits, the improvement of the TTS was 15.0 % (compared to the 'no control' case), which is very close to the improvement achieved by the case with the speed limits. If the bounds were chosen such that the flow limitation can be stronger (such that the control signal does not hit the bound), the improvement of the TTS was 17.4 %. If the on/off switching of the main-stream metering device was taken into account, the improvement of the TTS was 16.1 %.

The interpretation of these result is that the choice between speed limits and main-stream metering should be made based on the demands on the on-ramp and the freeway. If the speed limits can limit the flow sufficiently, i.e., the flow corresponding to the lowest speed limits is such that the ramp flow and the main flow can be accommodated, then speed limits should be used. If not, main-stream metering should be used, which can limit the flow much more. The preference for speed limits is motivated by the advantages of speed limits compared to main-stream metering. First, the maximum flow for main-stream metering is limited to approximately 75 % of the nominal capacity of the freeway when main-stream metering is on. This can cause oscillatory behavior when only a lighter flow limitation is necessary. Second, main-stream metering limits the flow only at one location which may cause shock waves. Opposed to this, speed limits can limit the flow more gradually, and since there are often installed more speed limit signs on freeway stretch they can prevent shock waves. A possible approach that utilizes the advantages of both speed limits and main-stream metering would be to use both measures and switch between the two, depending on the severity of the traffic congestion.

In Section 6.3 we have applied speed limits to reduce or eliminate shock waves on freeways. The MPC framework was applied to a benchmark network consisting of a link of 12 km, where 6 segments of 1 km are controlled by speed limits. The controller was evaluated for two different downstream density scenarios for the shock wave entering from the downstream end of the link. Both scenarios gave similar results. It was shown that coordinated control with real-valued speed limits (base case) is effective against shock waves. The performance loss caused by discrete speed limits and the inclusion of safety constraints was also examined. The performance of the discrete safety-constrained speed limits was comparable to that of the base case if the discrete speed limits were generated by ‘round’ or ‘ceil’. In all of these cases the coordination of speed limits eliminated the shock wave. The controlled case resulted in a network where the outflow was sooner restored to capacity, and in a decrease of the TTS of respectively 17 % and 21 %.

Chapter 7

Integrated optimal route guidance and ramp metering

In this chapter[1] we propose a traffic control approach that integrates ramp metering and dynamic route guidance using the MPC framework. The main objective of the control is to minimize the TTS in the network by providing accurate travel times while taking into account the effect of other traffic control measures, such as ramp metering. By aiming at minimizing the TTS as well as the difference between travel times shown on the DRIPs and the travel times actually realized by the drivers, the interests of both the individual drivers as well as the road administration are pursued. Simulation results for a case study show that the proposed integrated MPC traffic control results in a lower TTS while at the same time the drivers get accurate travel time information.

This chapter is organized as follows. In Section 7.1 we give an introduction the problem of dynamic route guidance via DRIPs in combination with other control measures, such as ramp metering. In this chapter we use the destination-oriented mode of METANET introduced in Section 3.2 to model the traffic on the freeways and secondary roads. In Section 7.2 we introduce a model for the reaction of the drivers to route guidance messages, and in Section 7.3 we explain how the individual travel times are estimated, which are necessary to calculate the difference between the predicted and realized travel times. Next, we present the objective function used in the model predictive control approach in Section 7.4, and finally we illustrate our approach for a case study in Section 7.5.

[1] The material presented in this chapter is the result of the joint work with Abdessadek Karimi.

7.1 Introduction

Dynamic route guidance is used to inform drivers about current or expected travel times and queue lengths so that they can reconsider their choice for a certain route.

On DRIPs usually one of the three types of information is displayed: *travel times*, *delays*, or *congestion lengths* on the alternative routes leading to a common destination. If we assume that the drivers want to minimize their travel time, the information about delays and congestion lengths is not very useful, since a shorter delay does not necessarily mean a shorter travel time (the nominal travel times on the alternative routes may also be different) and the relation between congestion length and travel time depends on the speed, which may also be different on the alternative routes.

The only remaining option is displaying travel times, but there is still a choice between displaying instantaneous travel times and predicted travel times. The disadvantage of instantaneous travel times is that the difference between the instantaneous travel time and the travel time experienced by the drivers may get large under changing traffic conditions. E.g., if there is a traffic jam with increasing length, the jam may be much longer when a driver arrives at the tail of the jam than when he saw the message on the DRIP. As the main goal of dynamic route guidance is to help drivers in such situations, displaying predicted travel times is a better option. Also in order to keep compliance high, the travel time prediction error of the DRIPs should be small.

However, using predicted travel times may result in splitting rates that are not optimal from a system point of view. To achieve the desired splitting rates, messages that are incorrect may be necessary. In other words, there is a conflict between informing drivers and controlling the traffic towards a better performance (cf. [90]. In this chapter we resolve the conflict by applying a control strategy that provides accurate travel time predictions while at the same time the network performance is optimized using DRIP messages and ramp metering. This means that the displayed travel times are not considered as the system *output,* but as the system *input* (control signal). So, in this chapter *optimized* travel times are introduced, which are simultaneously optimized with the other control signals.

Combination of on-ramp metering and dynamic route guidance with the use of an optimal control strategy has been studied in [35, 91]. There, optimal split rates are calculated at points where drivers can choose between alternatives using METANET-DTA, and the ramp metering rates are calculated with the ALINEA feedback algorithm or also taken into account within the optimization routine. However, after calculating optimal split rates as done in [35, 91], it is rather hard to find those control measures that realize the optimal splitting rates.

7.2 Driver route choice modeling

The METANET model presented in Section 3.2 describes the evolution of the traffic flows in a traffic network. One of the variables in this model is the routing choice parameter β, which is the result of the drivers' behavior, and which in our case will be influenced by the travel times shown on the dynamic route information panels (DRIPs). Hence, we also require a model that describes how drivers react to travel time information and how they adapt their route choice.

A well-known behavior model is the logit model [23, 157], which is used to model all kinds of consumer behavior based on the cost of several alternatives. The lower the cost of an alternative, the more consumers will choose that alternative. Also in traffic modeling these kinds of models are used. In that case consumers are the drivers, and the cost is the comfort, safety, or travel time of the alternative routes to reach the desired destination. The logit model calculates the probability that a driver chooses one of more alternatives based on the difference in travel time between the alternatives.

Assume that we have two possible choices m_1 and m_2 at node n to get to destination j. For the calculation of the split rates out of the travel time difference between two alternatives the logit model results in

$$\beta_{n,m,j}(k) = \frac{\exp(\sigma\,\vartheta_{n.m,j}(k))}{\exp\left(\sigma\,\vartheta_{n,m_1,j}(k)\right) + \exp\left(\sigma\,\vartheta_{n,m_2,j}(k)\right)}$$

for $m = m_1$ or $m = m_2$, where $\vartheta_{n,m,j}(k)$ is the travel time shown on the DRIP at node n to travel to destination j via link m. The parameter σ describes how drivers react on a travel time difference between two alternatives. The higher σ, the less travel time difference is needed to convince drivers to choose the fastest alternative route. In Figure 7.1 an example is given for the logit model for several values of σ.

7.3 Calculation of individual travel times

The calculation of the individual travel times is necessary to determine the difference between the realized travel times and the travel times shown on the DRIP. This calculation is inspired by [25], and is done by tracking vehicles at every simulation step. When a vehicle passes a bifurcation node with a DRIP, the information is stored. When the vehicle leaves the network its realized travel time is computed, and the difference between the realized travel time and the predicted travel time is included in the performance function (see equation (7.1) below).

Let us now discuss how the travel times are determined. Every N simulation steps some virtual vehicles are inserted into the network and their progress through the network is tracked at every simulation step. More specifically, for each virtual vehicle ζ the

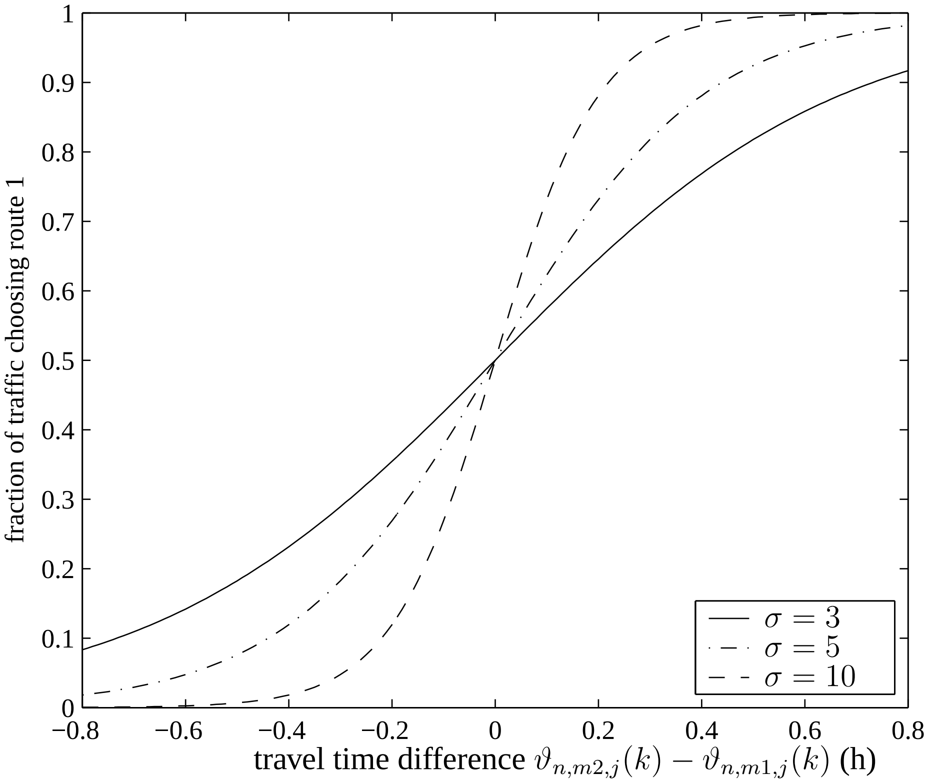


Figure 7.1: An example of the resulting splitting rates as a function of the travel time differences according to the logit model. It is assumed that $\vartheta_{n,m2,j}(k) = 1$ h.

following information is tracked during the simulation:

1. The route the vehicle is going to travel.

2. The link and the segment in which the vehicle currently is located, and its position s in this segment.

3. The predicted travel times shown in the DRIPs that the vehicle passed.

4. The realized travel time τ of the vehicle from the DRIPs it has already passed to the current position.

5. Whether or not the vehicle has left the network, and, if applicable, the time the vehicle left the network.

In order to track the position of the vehicles and to record the travel times, the METANET model has to be expanded as follows. Based on the METANET model equations given in Section 3.2 we can determine the time-dependent speed profile for all routes of a given network. Then the current position $s_{\zeta,m,i}(k)$ of vehicle ζ in segment i of link m is updated as follows:

$$s_{\zeta,m,i}(k+1) = s_{\zeta,m,i}(k) + v_{m,i}(k)T,$$

where $v_{m,i}(k)$ is the mean speed on segment i of link m at simulation step k. If the updated position $s_{\zeta,m,i}(k+1)$ is larger than the length L_m of segment i of link m, we put the vehicle ζ in the next segment of its route (say, segment i' of link m'), and we adapt the (new) position $s_{\zeta,m',i'}(k+1)$ accordingly.

The travel time $\tau_{\zeta,\omega}(k)$ of vehicle ζ from DRIP ω to its current position is updated as follows:

$$\tau_{\zeta,\omega}(k+1) = \tau_{\zeta,\omega}(k) + T \quad .$$

7.4 Control Strategy

To solve the integrated control of dynamic route guidance and ramp metering we apply the MPC framework presented in Chapter 4. Here, we present only the objective function used in the MPC controller, since the other elements are the same as in previous chapters.

Note that similarly to Chapter 4 we distinguish between the controller time step length T_c and the simulation time step length T, and between the controller time step counter k_c and the model time step counter k. We assume that the controller time step length is an integer multiple of the simulation time step length: $T_\mathrm{c} = MT$, with M a positive integer.

7.4.1 States, control signals, and objective function

The state vector of the traffic network consists of the partial densities for every segment of each link, the mean speed of every segment of every link, and the partial queues at every origin. The control vector consists of the ramp metering rates and the displayed travel times at bifurcation nodes. The process disturbance or external input vector consists of the demands and the composition rates at the origins.

The objective function is the weighted sum of the TTS, a prediction error term, and a small control signal variation penalty:

$$\begin{aligned} J(k_\mathrm{c}) = &\xi_1 T \sum_{k=Mk_\mathrm{c}}^{M(k_\mathrm{c}+N_\mathrm{p})-1} \left[\sum_{(m,i)\in I_\mathrm{all}} \rho_{m,i}(k) L_m \lambda_m + \gamma \sum_{o\in O_\mathrm{all}} w_o(k) \right] + \\ &\xi_2 \sum_{\zeta\in\mathcal{V}(k_\mathrm{c})} \sum_{\omega\in\mathcal{D}(\zeta)} (\vartheta_\text{pred}(\zeta,\omega)) - \vartheta_\text{real}(\zeta,\omega))^2 + \\ &\xi_3 \sum_{\ell=k_\mathrm{c}}^{k_\mathrm{c}+N_\mathrm{c}-1} \|u(\ell) - u(\ell-1)\|^2 \ , \end{aligned} \tag{7.1}$$

where $u(\ell)$ is the vector containing all control signals, I_all is the set of pairs of indexes (m,i) of all links and segments in the network, O_all is the set of all origins, $\mathcal{V}(k_\mathrm{c})$ is the set of indexes of all vehicles that left the network in the period $[k_\mathrm{c}, k_\mathrm{c} + N_\mathrm{p} - 1)$, $\mathcal{D}(\zeta)$ is the set of indexes of DRIPs that vehicle ζ has encountered, variable $\vartheta_\text{pred}(\zeta,\omega)$ is the travel time shown on the DRIP ω for vehicle ζ, and $\vartheta_\text{real}(\zeta,\omega)$ is the actually realized travel time for vehicle ζ from DRIP ω to its destination.

The first term in the objective function (7.1) is the TTS (both on the freeways and in the on-ramp queues, where the relative contribution of the latter is determined by the weighting factor γ). The second term penalizes (and tends to equalize) the travel time prediction errors. Note that these prediction errors are calculated for the virtual reference vehicles as explained in Section 7.3. The third term penalizes the control variance. The ξ_i's are weighting factors for the different terms of the objective function. The values for the ξ_i's depend on the traffic policy imposed by the road administrator. In this chapter highest priority is given to the TTS, followed by the prediction error, and the lowest priority is given to the control signal variation term.

7.5 Case study

7.5.1 Set-up

The network for the case study is chosen to be simple but containing all essential elements. The network is shown in Figure 7.2, and consists of two origins O_1, O_2 and two

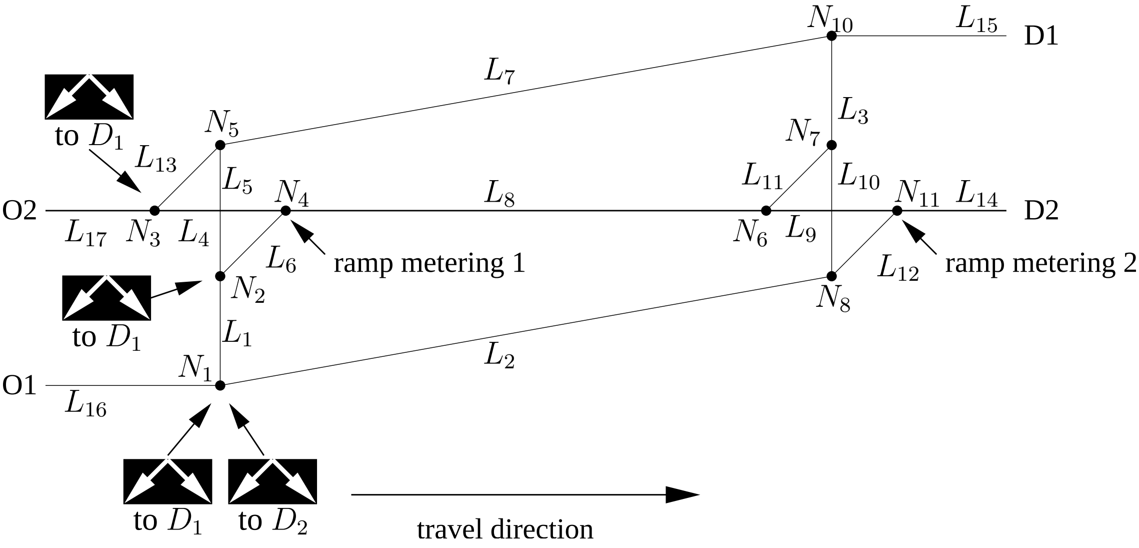


Figure 7.2: The traffic network of the case study consists of two origins O_1, O_2, and two destinations D_1, D_2. The network contains a freeway (consisting of links L_{17}, L_4, L_8, L_9, and L_{14}), and secondary roads (consisting of the other links). There are four DRIPs and two on-ramp metering installations (indicated by the symbol RM).

destinations D_1, D_2. Origin O_2 and destination D_2 are on the freeway (which consists of links L_{17}, L_4, L_8, L_9 and L_{14}), whereas O_1 and D_1 are on the secondary road network (which consists of the other links). Each link consists of one or more segments of 1 km except the on-ramp links (L_6,L_{12}), which have a length of 700 m.

Only one direction is considered, and that is from O_1, O_2 to D_1, D_2. For several origin-destination pairs, drivers can choose whether they travel via the freeway or via the secondary roads. There are three alternative routes from O_1 to D_1, two alternatives from O_1 to D_2 and from O_2 to D_1, and one way to travel from O_2 to D_2. DRIPs are installed at the bifurcation nodes N_1, N_2, and node N_3 as follows:

- At node N_1 two DRIPs are installed, one for destination D_1 and one for destination D_2. The DRIP for destination D_1 shows three travel times because there are three alternative routes from node N_1 to destination D_1: L_1–L_6–L_8–L_{11}–L_3–L_{15}, L_1–L_5–L_7–L_{15}, and L_2–L_{10}–L_3–L_{15}. The DRIP at node N_1 for destination D_2 shows two travel times: for routes L_6–L_8–L_9–L_{14} and L_2–L_{12}–L_{14}.

- At node N_2 there is only one way to travel to destination D_2, and there are two alternatives to travel to destination D_1: L_6–L_8–L_{11}–L_3–L_{15} and L_5–L_7–L_{15}.

- At node N_3 there is only one alternative to travel to destination D_2, and there are two alternatives to travel to destination D_1: L_{13}–L_7–L_{15} and L_4–L_8–L_{11}–L_3–L_{15}.

The on/off-ramps are situated at points where the secondary road crosses the freeway. At each on-ramp a ramp metering system is installed. Traffic from a secondary road that wants to travel via freeway has to cross one of the two on-ramps. Traffic from the freeway that wants to travel to destination D_1 has to cross one of the off-ramps.

7.5.2 Scenario

We consider the following scenario:

- At the start of the simulation we have a capacity reduction at destination D_2, which results in a shock wave originating at D_2. The tail of the shock wave propagates upstream until the downstream end of freeway link L_8 is congested. Calculations show that in this case the alternative routes from origin O_1 to destination D_1 get faster, resulting in more traffic choosing these alternative routes.

- The simulation starts from a steady state situation in which we have the following flows or demands: 600 veh/h for the origin-destination (OD) pair (O_1, D_1), 1400 veh/h for (O_1, D_2), 900 veh/h for (O_2, D_1), and 2100 veh/h for (O_2, D_2). However, 5 min after the simulation has started, the total demand at O_2 increases, resulting in a flow of 1200 veh/h for OD pair (O_2, D_1), and 2800 veh/h for (O_2, D_2).

7.5.3 Model and controller parameters

The METANET parameters used for the simulation of the case study network are based on the METANET validation as described in [89]. The additional parameters introduced with the model extensions in Section 3.3 are also defined. More specifically, in our case study two values for ν are defined as is also done in [15, 62]. We select $\eta_{\text{low}} = 35\,\text{km}^2/\text{h}$, and $\eta_{\text{high}} = 60\,\text{km}^2/\text{h}$. The capacity of the freeway links is chosen as 2200 veh/h, and the capacity of the secondary road links is chosen as 1500 veh/h. The free flow speed $v_{\text{free},m}$ is 120 km/h for freeway links, and 80 km/h for secondary road links. Furthermore, we have $\tau = 20\,\text{s}$, $\rho_{\text{max}} = 180\,\text{veh/km/lane}$, $\kappa = 40\,\text{veh/km}$, and $v_{\text{min}} = 7\,\text{km/h}$ as the minimum speed (cf. [89]).

For the controller we have taken $T_{\text{c}} = 5\,\text{min}$. The prediction horizon $N_{\text{p}} = 12$ corresponds to a prediction of 1 h ahead. For the control horizon we take $N_{\text{c}} = 9$, which corresponds to a period of 45 min, which is shorter than the prediction horizon, but long enough to get a good performance. The weighting parameters were chosen $\xi_1 = \xi_3 = 1$. For the weight for prediction error ξ_2, we simulate both $\xi_2 = 0$ and $\xi_2 = 1$ in order to illustrate the effect of the prediction error term.

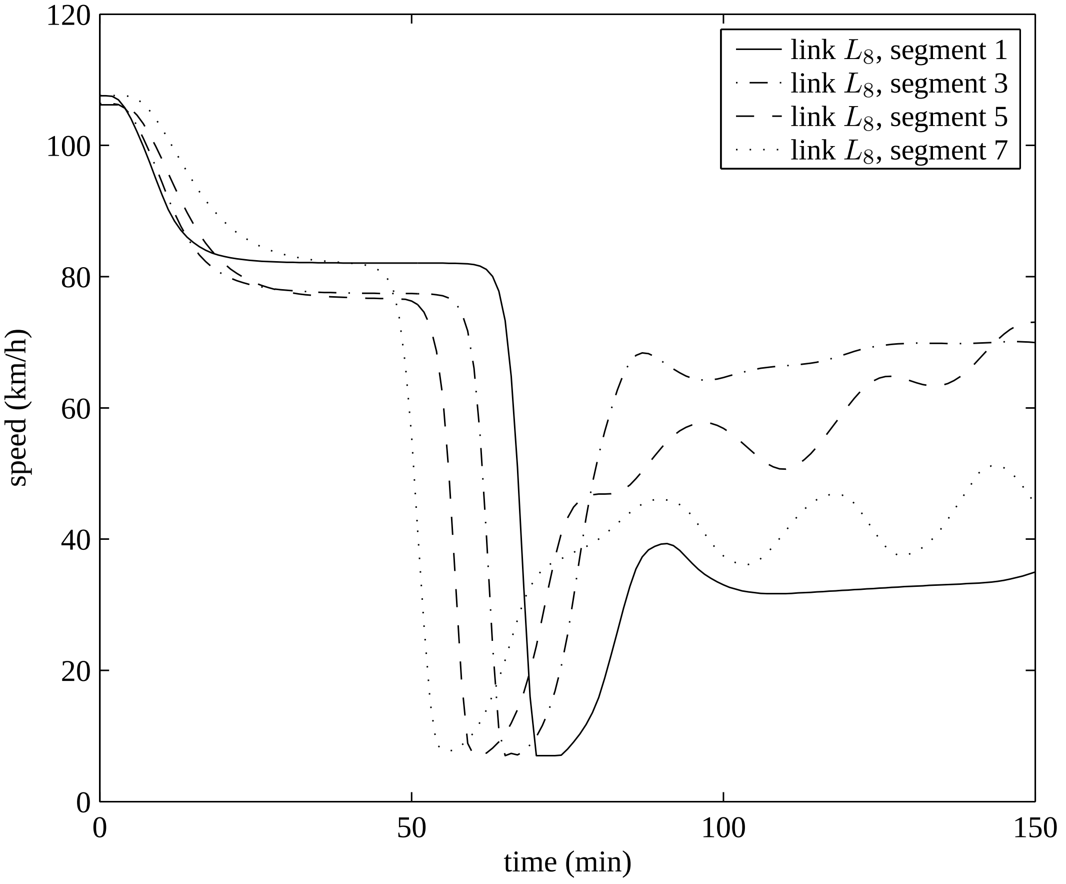


Figure 7.3: Evolution of the mean speed on some segments of link L_8 in the 'no control' case. When no control measures are activated, the shock wave propagates through link L_8, and speed on this decreases significantly.

7.6 Simulation results

We have simulated the network of the case study for the scenario given above both with and without MPC control. Below we discuss some of the most relevant results of these simulations.

Figure 7.3 shows the evolution of the speed on the freeway link L_8 when no control measures are active. This link is the main part of the freeway, and it is also used by traffic that is destined to secondary road destinations. Due to the shock wave entering via destination D_2 at the beginning of the simulation period, the speeds on the freeway are reduced significantly. Since drivers are not informed about the alternative routes, which could reduce their travel times, they still choose to travel via link L_8. The lack of information drivers receive when there is no dynamic route guidance active results in the inefficient use of some secondary road links, such as link L_{10}. Although link L_{10}

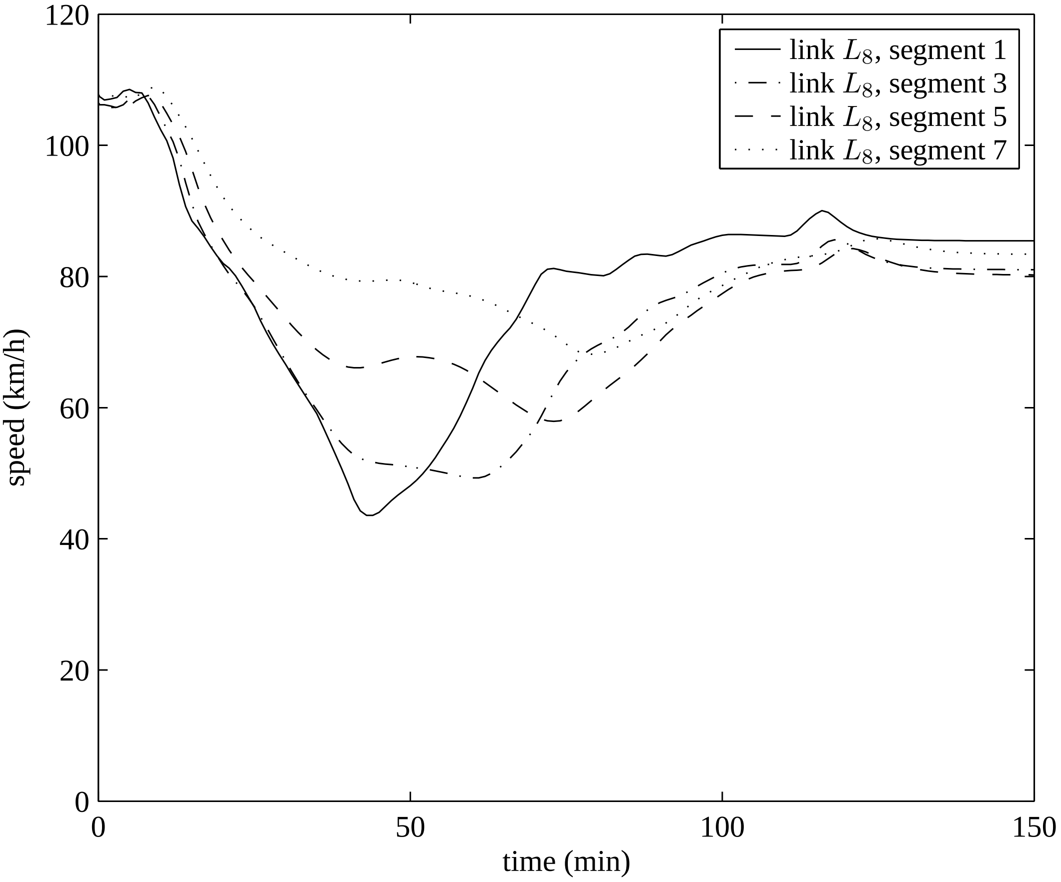


Figure 7.4: Evolution of the mean speed on the segments of link L_8 when route guidance and ramp metering is applied. Ramp metering and rerouting results in more traffic choosing alternative routes via the secondary roads, which leads to less traffic on link L_8 and increased speeds with respect to the 'no control' case.

can be optimally used for the rerouting of traffic flow, the link is almost unused in the uncontrolled case.

When dynamic route guidance is active and MPC is applied, we get an improvement in the mean speed over the freeway as is shown in Figure 7.4. The freeway is relieved from congestion because of the rerouting due to dynamic route guidance, which results in more traffic choosing for alternative routes via the secondary roads. This leads to less traffic on link L_8 and increased speeds with respect to the 'no control' case. Furthermore, the ramp metering reduces the inflow of traffic destined to the freeway destination and thereby improves the throughput on the freeway. As a consequence, the shock wave is damped significantly.

The evolution of the flows in link L_{14} in the uncontrolled and the controlled case is represented in Figures 7.5 and 7.5. In the uncontrolled case, the flow as well as the speeds

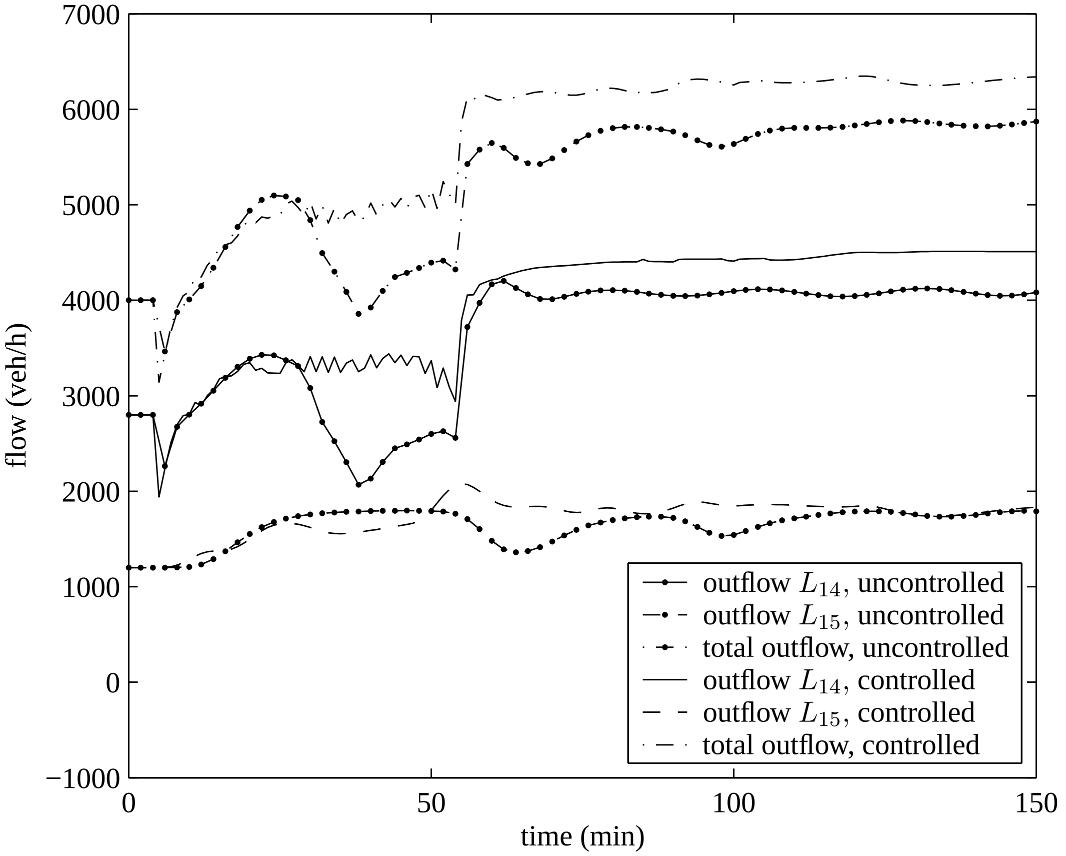


Figure 7.5: Evolution of the outflow of the exit links L_{14} and L_{15} and the total outflow in the uncontrolled and controlled cases. The outflow of the controlled case is higher.

on this link are reduced significantly due to the shock wave. The outflow of the network is reduced due to the shock wave, which results in congestion and lower speeds in the rest of the network. When dynamic route guidance and ramp metering are activated the flow on link L_{14} improves significantly. The outflow from the network also increases since the effect of the shock wave is reduced by the rerouting and ramp metering.

In Figure 7.5 we observe oscillations in the first 350 s. The oscillations do not originate from the ramp metering, because the ramp metering signal varies more slowly. The real cause is unknown and is a topic for future research.

Figure 7.6 shows the metering rates at ramp links L_6 and L_{12} for the controlled case.

The traffic from origin O_2 destined to D_1 is routed via link L_{13}. Traffic from origin O_2 destined to D_2 has no alternative than to travel via the highway. Traffic that originates in O_1 and is destined to D_1 is routed via link L_2, while traffic from O_1 destined to D_2 is routed via link L_1. The traffic that is routed via link L_1 to destination D_2 has to travel via the on-ramp link L_6. Although there is no critical situation on link L_8 at the on-ramp on

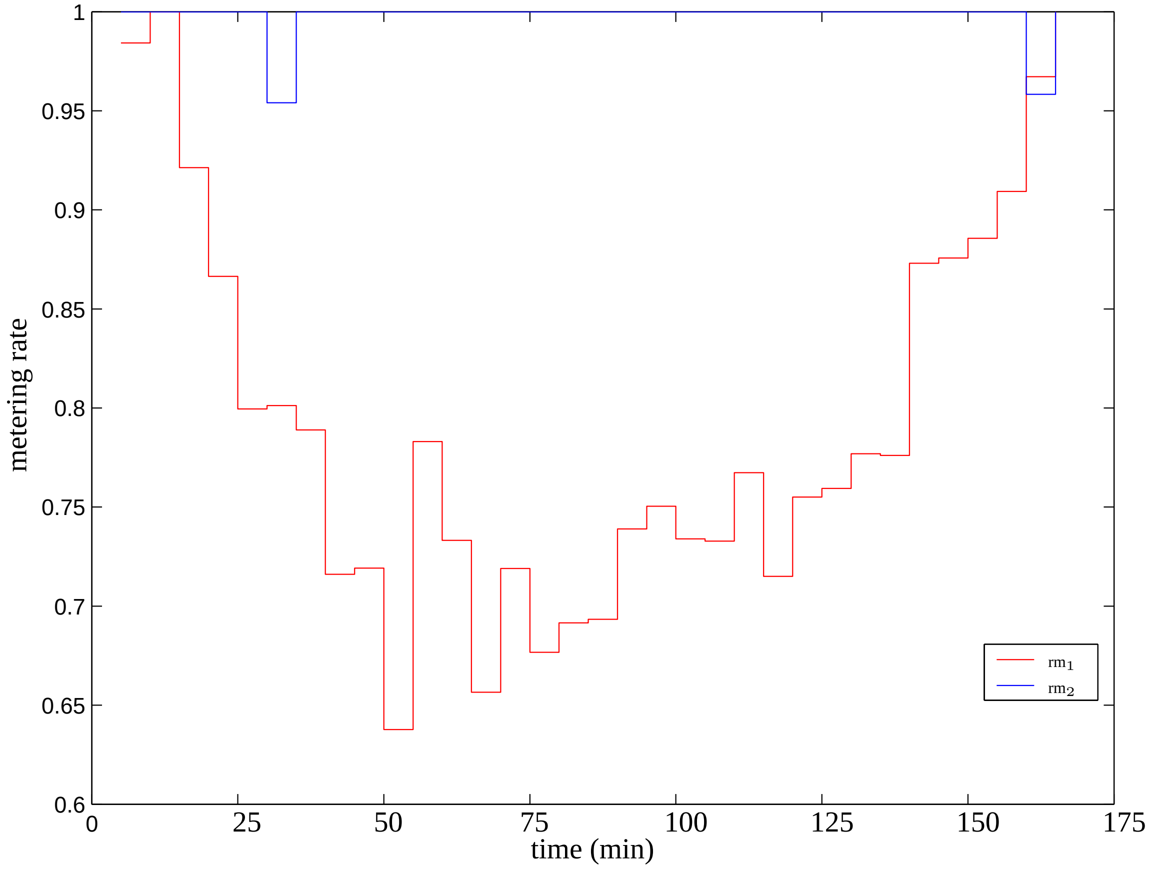


Figure 7.6: Metering rates of the metering systems at ramp link L_6 (rm$_1$) and ramp link L_{12} (rm$_2$).

link L_6 the metering anticipates on the fact that if all traffic is admitted to the freeway this can cause the shock wave not to be reduced optimally. The metering admits at least 60 % of the traffic on on-ramp link L_6 to enter the freeway. The ramp metering causes queues on the on-ramp link L_6 to spill back to link L_1, which results in a queue of 2 km on link L_1.

The TTS in case of ‘no control’ is 6365.4 veh.h compared to 4530.6 veh.h in the case of MPC control, which corresponds to an improvement of 28.8 %.

Figure 7.7 shows the difference in prediction error between not taking the prediction error into account in the objective function ($\xi_2 = 0$) and taking the prediction error into account ($\xi_2 = 1$). The travel times shown on the DRIPs and the metering rates are optimized in both cases. Each dot in Figure 7.7 represents one vehicle that has left the network. The optimal reference shown in the figure corresponds the optimized travel times shown on the DRIPs being equal to the travel times realized by the drivers. The angles α^+ and α^- are representative for the maximum errors in case the optimized travel times were too low and too high respectively. It is a subject for future research why the optimized travel times are, in most cases, higher than the realized travel times.

7.7 Conclusions

We have considered the problem of traffic control based on MPC with ramp metering and dynamic route guidance as the traffic control measures.

The first issue addressed in this chapter is the integration of dynamic route guidance and ramp metering. The approach we have chosen is to use the DRIP as both a control tool *and* an information provider to the drivers, and ramp metering as a control tool to redistribute the delays over the on-ramp and the freeway. The drivers’ reaction to the travel times shown on the dynamic route information panels is modeled by the logit model. The travel times shown on the DRIPs are optimized travel times, which are chosen such that the reactions of the drivers and the control actions of ramp metering are taken into account. This results in one optimization that optimizes both the ramp metering and the travel times shown on the DRIPs at the same time such that on the one hand the TTS in the network is reduced by optimally rerouting traffic over the available alternative routes in the network, but on the other hand the difference between the travel times shown on the DRIPs and the travel times actually realized by the drivers is also kept as small as possible.

The second issue addressed in this chapter is whether the proposed approach leads to an improvement of the traffic system in congested situations such as described in the case study. The simulations show that rerouting of traffic and on-ramp metering using MPC leads to an improvement in performance of 28.8 % for the case study.

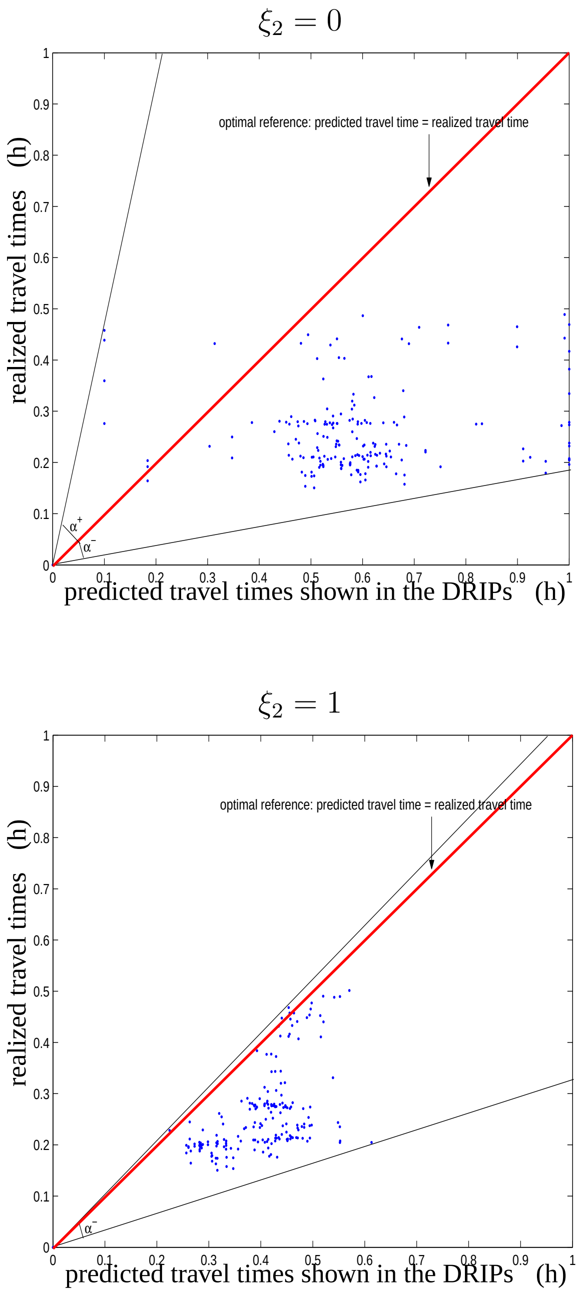


Figure 7.7: Plot of the travel time prediction error in case the weighting factor ξ_2 for the prediction error term in the objective function is set to 0 (top) and to 1 (bottom). The angles α^+ and α^- are representative for the maximum relative errors of the shown travel times.

Chapter 8

Mixed urban–freeway networks

8.1 Introduction

In this chapter[1] we consider traffic control for networks containing both urban roads and freeways.

Freeway traffic control measures such as ramp metering often allow a better flow, and higher speeds and throughput on the freeway at the cost of queues at the on-ramp, which may spill back and block urban roads. On the other hand, many cities try to get the vehicles out of the urban road network as soon as possible, thereby displacing the congestion to the neighboring ring roads and freeways. Moreover, freeway control measures that improve the traffic flow towards urban roads are only effective if these roads can accommodate the increased flows. If not, the problems are only shifted towards the urban area. This shift of congestion between urban roads and freeways, and vice versa, is often made worse by the fact that in many countries urban, regional, and freeway roads are managed and controlled by different traffic management bodies, each with their own traffic policies and objectives. However, the situation sketched above is certainly not optimal. By considering an integrated and coordinated approach the performance (taking into account the trade-off between the often conflicting objectives and interests of different traffic management bodies) of the overall network can be significantly improved. Therefore, our goal in this chapter is to develop an integrated traffic control approach for coordinated control of mixed urban and freeway traffic networks that makes an appropriate trade-off between the performance of the urban and freeway traffic operations, and that prevents a shift a problems from urban roads to freeways, and vice versa.

For urban traffic networks systems such as UTOPIA/SPOT, and SCOOT [143] use an integrated approach that coordinates the operation of several traffic signal set-ups in a city to obtain a smoother flow and/or a better circulation. For freeway traffic networks several

[1] The material presented in this chapter is the result of the joint work with Monique van den Berg.

authors [46, 91, 126, 125] have considered a coordinated approach in which many different control measures (such as ramp metering, route guidance, variable speed limits, etc.) are coordinated on a larger scale, which results in a better overall performance. However, up to now little attention has been paid to integrated control of networks consisting of both urban and freeway roads.

We once again, use a model predictive control (MPC) approach (see also Chapter 4). As MPC requires a model to predict the future evolution of the traffic flow, a first requirement is a model that describes the evolution of the traffic in a mixed urban/freeway traffic network. In this chapter we will develop a macroscopic traffic model for networks containing both urban roads and freeways. We opt for a macroscopic model that yields a sufficiently accurate description of the evolution of the traffic flows for given traffic demands, traffic conditions, and output restrictions on the one hand, and that can be simulated sufficiently fast — so that is can be used in on-line traffic control — on the other hand. In particular, we use the extended version of the destination independent METANET [109, 156] traffic flow model to model the freeway traffic. For the urban network we use a modified and extended model that is based on the Kashani model [82]. This model has the advantage that the travel time from the entrance of the link to the end of the queue waiting at the traffic light to leave the link is taken into account. This also allows for the modeling of arriving platoons at the intersections.

Furthermore, we also model the interface between the urban and the freeway model. This results in an integrated model for mixed freeway and urban traffic networks, which is especially suited for use in a model-based predictive traffic control approach. We propose such an approach, and we illustrate it using a synthetic case study.

This chapter is organized as follows. Recall that in Sections 3.2–3.3 we have introduced the METANET model that is used to describe the traffic on the freeway. While this model is also used in this chapter, we do not describe it here again. In Section 8.2 the model used for the urban areas is given. Section 8.3 contains the formulas used for the traffic on the on-ramps and off-ramps, i.e., for the interface between the urban and the freeway network, which results in an integrated model for mixed urban and freeway networks. Since the two models have different simulation time step lengths, the variables of the overall model have to be computed in a specific order, which we discuss in Section 8.4. Next, in Section 8.5 we present the control signal, the objective function and the constraints used in the MPC approach for coordinated control of mixed urban and freeway networks. Preliminary simulation results for a synthetic case study are then given in Section 8.6.

We repeat here the definition of the basic METANET variables that will also be used in this chapter:

m	link index
i	segment index
T_f	time step of the freeway simulation (in hours; a typical value is about 10/3600 h = 10 s)
k_f	freeway time step counter
N_m	number of segments in freeway link m
λ_m	number of lanes in freeway link m
L_m	length of the segments in link m (km)
τ, κ, a_m, η	constant parameters reflecting street geometry, vehicle characteristics, drivers' behavior, etc.
$\rho_{\mathrm{crit},m}$	critical density of the segments of link m (veh/km/lane)
ρ_max	maximum density (veh/km/lane)
$v_{\mathrm{free},m}$	speed that the vehicles tend to drive at under free flow conditions on link m (km/h)
C_o	capacity of on-ramp o (veh/h)
$\rho_{m,i}(k_\mathrm{f})$	density of segment i of freeway link m at time $t = k_\mathrm{f}T_\mathrm{f}$ (veh/km/lane)
$v_{m,i}(k_\mathrm{f})$	speed in segment i of freeway link m at time $t = k_\mathrm{f}T_\mathrm{f}$ (km/h)
$w_o(k_\mathrm{f})$	length of the queue on on-ramp o at time $t = k_\mathrm{f}T_\mathrm{f}$ (veh)
$q_o(k_\mathrm{f})$	flow that enters freeway from origin o (veh/h)
$q_{m,i}(k_\mathrm{f})$	flow leaving segment i of freeway link m in $[k_\mathrm{f}T_\mathrm{f}, (k_\mathrm{f}+1)T_\mathrm{f})$ (veh/h)
$d_o(k_\mathrm{f})$	demand flow arriving at the origin of freeway link f in the time interval $[k_\mathrm{f}T_\mathrm{f}, (k_\mathrm{f}+1)T_\mathrm{f})$ (veh/h).
Q_n	flow that enters freeway node n

Remark 8.1.1 As 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 for the freeway network model (k_f), the urban model (k_u), and the controller (k_c). For the sake of simplicity, we assume that T_u is an integer divisor of T_f, and that T_f is an integer divisor of T_c:

$$T_\mathrm{f} = L\,T_\mathrm{u}, \quad T_\mathrm{c} = M\,T_\mathrm{f} = M L\,T_\mathrm{u} \enspace ,$$

with L and M positive integers. □

8.2 Urban model

Several authors have already developed models to describe traffic flows in urban traffic networks [35, 82, 105]. Recall that we will use the model for on-line traffic control, and that we have to select a model that offers an appropriate trade-off between accuracy and computational complexity.

Our model to describe the traffic in the urban parts of the network is based on the Kashani model [82], but it has the following extensions:

- We use horizontal queues, which allows us to take into account the blocking effect that arises when a link is full of vehicles and other vehicles from an upstream intersection cannot enter anymore.
- We use turning-direction-dependent queues, which correctly model the queue dynamics if one turning direction is blocked and another is free.
- We use short time steps compared to the Kashani model, which uses the cycle time of the traffic signal set-up as the simulation time step. Such a large simulation time step poses problems when we want to model the blocking effect accurately. Furthermore, we also want to allow different cycle times for different traffic signal installations, we will use a fixed simulation step T_{u} (typically 1 to 5 s) for the urban network that is independent of the cycle times of the traffic signal installations.

Remark 8.2.1 Here, we point out some notational issues:

- The nodes in the combined model includes the METANET nodes and the urban intersections and are indexed sequentially, therefore any node is has a unique index. Similarly, all links (freeway, urban, on-ramps, off-ramps) are indexed sequentially, and thus uniquely.
- In the following definitions of the traffic variables and the remainder of this chapter we mean by origins of the urban intersection s all upstream intersections (*nodes*) and off-ramps that are directly connected by a link with s. By destinations of the urban intersection s we mean all downstream *links* (including on-ramps) that are directly connected by a link with s.
- For variables related to turning directions have an additional index that indicates the link which the drivers want to turn to.

□

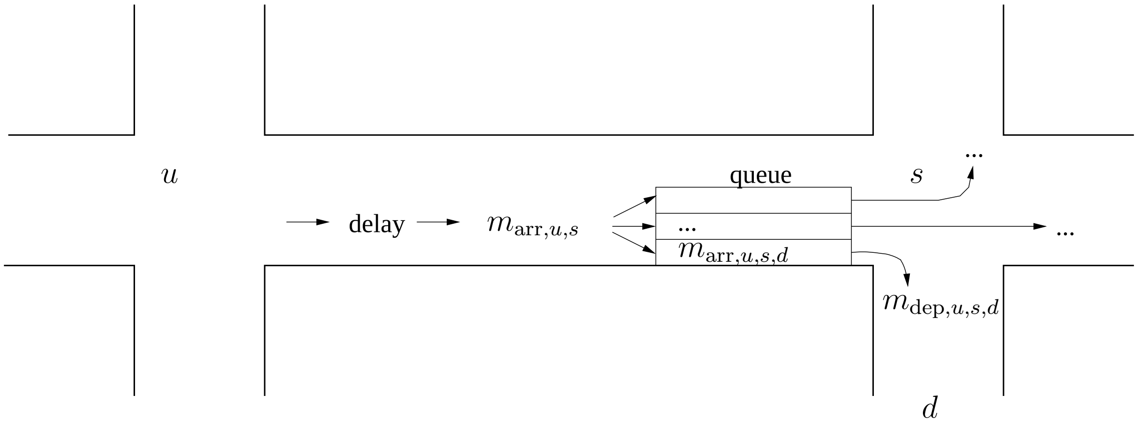


Figure 8.1: Variables used in the urban model.

The new model is described using the following parameters (see also Figure 8.1):

s, n, u	intersection indexes (node)
T_u	time step used for the urban simulation (h)
k_u	urban time step counter
U_s	set of origins of intersection s
d	link index (when it is a destination)
O_s	set of leaving links of node (intersection) s
$x_{u,s,d}(k_\mathrm{u})$	queue length at time $t = k_\mathrm{u}T_\mathrm{u}$ (veh) at intersection s, for traffic that goes from origin u to link d
$l_{s,n}$	link connecting intersections s and n
$\beta_{u,s,d}(k_\mathrm{u})$	fraction of the traffic arriving from origin u at intersection s that wants to go to link d in the time interval $[k_\mathrm{u}T_\mathrm{u}, (k_\mathrm{u}+1)T_\mathrm{u})$
$L_{s,n}$	length of link $l_{s,n}$ (veh)
$L_{\mathrm{km},s,n}$	length of link $l_{s,n}$ (km)
L_vehicle	average length of the vehicles (km)
$S_{s,n}(k_\mathrm{u})$	available free space of link $l_{s,n}$ at time $t = k_\mathrm{u}T_\mathrm{u}$ (veh) (i.e., the buffer capacity $L_{s,n}$ minus the number of vehicles that are already present at time $t = k_\mathrm{u}T_\mathrm{u}$)
$m_{\mathrm{arr},u,s}(k_\mathrm{u})$	number of vehicles arriving at the tail of the queue in link $l_{u,s}$ during the time interval $[k_\mathrm{u}T_\mathrm{u}, (k_\mathrm{u}+1)T_\mathrm{u})$
$m_{\mathrm{arr},u,s,d}(k_\mathrm{u})$	number of vehicles arriving at the tail of the queue in link $l_{u,s}$ with destination link d during the time interval $[k_\mathrm{u}T_\mathrm{u}, (k_\mathrm{u}+1)T_\mathrm{u})$
$m_{\mathrm{dep},u,s,d}(k_\mathrm{u})$	number of vehicles departing from link $l_{u,s}$ towards link d in $[k_\mathrm{u}T_\mathrm{u}, (k_\mathrm{u}+1)T_\mathrm{u})$
$m_{\mathrm{dep},s,d}(k_\mathrm{u})$	number of vehicles departing from intersection s towards link $l_{s,d}$ in $[k_\mathrm{u}T_\mathrm{u}, (k_\mathrm{u}+1)T_\mathrm{u})$

$g_{u,s,d}(k_\mathrm{u})$ indicates whether the traffic sign at intersection s for the traffic going from u to d is green[2](1) or red (0) during $[k_\mathrm{u}T_\mathrm{u},(k_\mathrm{u}+1)T_\mathrm{u})$

$C_{u,s,d}(k_\mathrm{u})$ capacity of intersection s for traffic arriving from u and turning to d at time $t=k_\mathrm{u}T_\mathrm{u}$ (veh/h)

$v_{s,n}$ free-flow speed[3]for the urban traffic between the entrance of the link $l_{s,n}$ and the tail of the queue at intersection n (km/h)

$\delta_{s,n}(k_\mathrm{u})$ time required to reach the tail of the queue waiting in link $l_{s,n}$ at time $t=k_\mathrm{u}T_\mathrm{u}$ (units of urban time steps)

$w_{o,m}(k_\mathrm{u})$ queue length on on-ramp o (veh) coming from intersection s waiting to depart towards freeway link m at time $t=k_\mathrm{u}T_\mathrm{u}$.

$x_{u,s,d}(k_\mathrm{u})$ queue length link $l_{u,s}$ (veh) waiting to depart towards link d at time $t=k_\mathrm{u}T_\mathrm{u}$.

$\lambda n,s$ the number of lanes in urban link $l_{n,s}$

The new model is formulated as follows. The traffic leaving the link $l_{u,s}$ toward link d is given by

$$m_{\mathrm{dep},u,s,d}(k_\mathrm{u})=\begin{cases}0 & \text{if } g_{u,s,d}(k_\mathrm{u})=0,\\ \min\left(x_{u,s,d}(k_\mathrm{u})+m_{\mathrm{arr},u,s,d}(k_\mathrm{u}),S_{s,d}(k_\mathrm{u}),T_\mathrm{u}C_{u,s,d}\right) & \\ & \text{if } g_{u,s,d}(k_\mathrm{u})=1.\end{cases}$$

The free space $S_{s,d}$ in link $l_{s,d}$ is equal to the number of vehicles that can enter the link. It imposes an implicit constraint on the number of vehicles that can depart towards each link, $m_{\mathrm{dep},u,s,d}$, and it can never be larger than the link length. It is computed as

$$S_{s,d}(k_\mathrm{u}+1)=S_{s,d}(k_\mathrm{u})-m_{\mathrm{dep},s,d}(k_\mathrm{u})+\sum_{u\in U_s}m_{\mathrm{dep},u,s,d}(k_\mathrm{u}) \quad . \tag{8.1}$$

Another constraint is that the total flow from several directions must be smaller than or equal to the storage space in the destination link d. These different flows do not have to have the same value because not all the queues from which they are coming have the same length and not all incoming flows have the same priority to enter the link. This results in a ratio between the different flows. To illustrate how the effective values of $m_{\mathrm{dep},u,s,d}(k_\mathrm{u})$ can be computed we assume there are two origins, and so two queues from which vehicles want to drive into the same link[4].(Let $m_{\mathrm{dep,int},1}(k_\mathrm{u})$ and $m_{\mathrm{dep,int},2}(k_\mathrm{u})$ denote the number

[2] We use here the notion of effective green which takes into account the acceleration behavior at the beginning of the green phase, and the length of the amber phase.

[3] We assume, for the sake of simplicity, that this speed is equal for all links.

[4] The extension to a link with more queues with vehicles waiting to enter is straightforward.

of vehicles intending to enter the link $l_{s,n}$ 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 $m_{\text{dep,int},1}(k_\text{u}) + m_{\text{dep,int},2}(k_\text{u}) \leq S_{s,n}(k_\text{u})$, then

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

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

$$\begin{cases} m_{\text{dep},1}(k_\text{u}) = m_{\text{dep,int},1}(k_\text{u}) \quad \text{and} & \\ m_{\text{dep},2}(k_\text{u}) = S_{s,n}(k_\text{u}) - m_{\text{dep,int},1}(k_\text{u}) & \text{if } m_{\text{dep,int},1}(k_\text{u}) \leq \frac{1}{2}S_{s,n}(k_\text{u}), \\ & \\ m_{\text{dep},1}(k_\text{u}) = m_{\text{dep},2}(k_\text{u}) = \frac{1}{2}S_{s,n}(k_\text{u}) & \text{if } m_{\text{dep,int},1}(k_\text{u}) \geq \frac{1}{2}S_{s,d}(k_\text{u}). \end{cases}$$

The traffic departing to link $l_{s,n}$ can be computed as

$$m_{\text{dep},s,n}(k_\text{u}) = \sum_{u \in U_s} m_{\text{dep},u,s,n}(k_\text{u}) \;\; .$$

These vehicles drive from the beginning of the link $l_{s,n}$ towards the tail of the queue waiting on the link. This gives a time delay $\delta_{s,n}(k_\text{u})$:

$$\delta_{s,n}(k_\text{u}) = \text{ceil}\left(\frac{S_{s,n}(k_\text{u})\, L_\text{vehicle}}{v_{s,n}} \right) \;\; ,$$

where $\text{ceil}(x)$ with x a real number denotes the smallest integer larger than or equal to x. The traffic arriving at the tail of the queue should be added to the traffic that arrived in the queue in earlier time steps. This results in:

$$m_{\text{arr},s,n}(k_\text{u} + \delta_{s,n}(k_\text{u})|k_\text{u}) = m_{\text{arr},s,n}(k_\text{u} + \delta_{s,n}(k_\text{u})|k_\text{u} - 1) + m_{\text{dep},s,n}(k_\text{u}) \;\; , \tag{8.2}$$

where $m_{\text{arr},s,n}(k|l)$ is the number of vehicles expected to arrive at the end of the queue at time k based on knowledge at time $l(\leq k)$, and $m_{\text{arr},s,n}(k) = m_{\text{arr},s,n}(k|k)$. The traffic $m_{\text{arr},s,n}(k_\text{u})$ reaches the tail of the queue in link $l_{s,n}$, and divides itself over the sub-queues according to the turning rates $\beta_{u,s,d}(k_\text{u})$, where d is the link which this fraction of the drivers wants to turn to. The number of vehicles arriving at the end of each sub-queue is then given by:

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

Finally, the sub-queue lengths are updated as follows:

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

8.3 On-ramps and off-ramps

Both the urban model and the freeway model have now been presented. The next step is to make the connection between the two models. This connection consists of on-ramps and off-ramps.

8.3.1 On-ramps

Consider an on-ramp o that connects intersection s of the urban network to node n of the freeway network. The traffic that enters the on-ramp from the urban network is given by $m_{\mathrm{dep},s,n}(k_{\mathrm{u}})$. This traffic has a delay given by $\delta_{s,n}(k_{\mathrm{u}})$, and $m_{\mathrm{arr},s,n}(k_{\mathrm{u}})$ is determined similarly as in (8.2). After reaching the tail of the queue on the on-ramp, the traffic divides itself over the different directions m (freeway links):

$$m_{\mathrm{arr},s,n,m}(k_{\mathrm{u}}) = \beta_{s,n,m}(k_{\mathrm{u}}) m_{\mathrm{arr},s,n}(k_{\mathrm{u}}) \ .$$

These vehicles arrive at the tail of the on-ramp queue. The queue length $w_{s,r,m}(k_{\mathrm{u}})$ is computed as

$$w_{o,m}(k_{\mathrm{u}}+1) = w_{o,m}(k_{\mathrm{u}}) + m_{\mathrm{arr},s,n,m}(k_{\mathrm{u}}) - m_{\mathrm{dep},s,n,m}(k_{\mathrm{u}}) \ .$$

The number of departures at the front of the on-ramp queue depends on the available space on the freeway, which space depends on the density on the freeway. This results in a maximum flow that can leave the on-ramp:

$$q_{\mathrm{max},o,m}(k_{\mathrm{f}}) = \begin{cases} C_o \left(1 - \dfrac{\rho_{m,1}(k_{\mathrm{f}}) - \rho_{\mathrm{crit},m}}{\rho_{\mathrm{max}} - \rho_{\mathrm{crit},m}}\right) & \text{if } \rho_{m,1}(k_{\mathrm{f}}) > \rho_{\mathrm{crit},m} \, , \\ C_o & \text{otherwise}. \end{cases}$$

The flow $q_{\mathrm{dep},r,m}(k_{\mathrm{f}})$ that enters the freeway is then given by

$$q_{\mathrm{dep},r,m}(k_{\mathrm{f}}) = \min\left(\frac{1}{T_{\mathrm{f}}}\left[w_{s,r,m}(k_{\mathrm{f}}L) + \sum_{k_{\mathrm{u}}=k_{\mathrm{f}}L}^{(k_{\mathrm{f}}+1)L-1} m_{\mathrm{arr},s,r,m}(k_{\mathrm{u}})\right], q_{\mathrm{max},o,m}(k_{\mathrm{f}})\right) \ .$$

This flow should be translated into the number of vehicles that leaves the on-ramp. This is done by distributing the flow equally over the urban time step[5]:

$$m_{\text{dep},s,r,m}(k_\text{u}) = \frac{q_{\text{dep},r,m}(k_\text{f})T_\text{f}}{L} \qquad \text{for } k_\text{u} = Lk_\text{f}, ..., L(k_\text{f}+1)-1 \ .$$

The free space $S_{s,r}(k_\text{u})$ is also computed using equation (8.1).

8.3.2 Off-ramps

Consider the off-ramp o that connects freeway node n to urban intersection s. When it is assumed that no car can enter and leave the link within one time freeway step, the departing traffic does not depend on the arriving traffic. Therefore, the departing traffic from intersection s can be computed first, and afterward the traffic entering the link $l_{n,s}$ can be computed.

The flow leaving the freeway cannot be larger than allowed by the free space on the off-ramp. This free space depends on the length of the off-ramp, on the queue currently waiting on it, and on the traffic that is going to leave the link $l_{r,s}$ during the period $[k_\text{f}T_\text{f}, (k_\text{f}+1)T_\text{f})$. The flow that wants to enter the off-ramp is a fraction of the flow on the freeway:

$$q_{\text{dep},\text{demand},n,s}(k_\text{f}) = \beta_{n,o}(k_\text{f})\, Q_n(k_\text{f}) \ .$$

This flow is not always able to enter the off-ramp, due to the maximum capacity C_o of the off-ramp, and the free space on the off-ramp. This free space in fact varies over the time interval $[k_\text{f}T_\text{f}, (k_\text{f}+1)T_\text{f})$, as vehicles are leaving at the front of the queue during the time interval, and so the free space grows. This results in the following expression for the actual flow that arrives at the off-ramp from the freeway:

$$\begin{aligned} &q_{\text{dep},n,s}(k_\text{f}) = \\ &\quad \min\left(q_{\text{dep},\text{demand},n,s}(k_\text{f}), C_o, \frac{1}{T_\text{f}}\left[S_{n,s}(Lk_\text{f}) + \sum_{\ell=Lk_\text{f}}^{L(k_\text{f}+1)-1} \sum_{d\in D_s} m_{\text{dep},n,s,d}(\ell)\right]\right) \ . \end{aligned}$$

The flow entering the off-ramp is translated into the number of vehicles per urban time step. Similarly to on-ramps we assume equal distribution of the flow over the urban time steps:

$$m_{\text{dep},r,s}(k_\text{u}) = \frac{q_{\text{dep},m,r}(k_\text{f})T_\text{f}}{L} \qquad \text{for } k_\text{u} = k_\text{f}L+1, \ldots, (k_\text{f}+1)L \, .$$

This traffic undergoes a delay $\delta_{r,s}(k_\text{u})$ and then enters the urban network.

[5] Since the freeway model does not contain any information about the distribution of the flow *withing* one freeway time step, we assume equal distribution, for the sake of simplicity.

A constraint for the flow in METANET can be implemented by adjusting the speed of the traffic. The flow is computed using equation (3.1). A way to influence the flow is changing the speed in the last segments of incoming links of node n. These speeds can be adapted as follows:

$$v_{m,N_m}(k_\mathrm{f})_\mathrm{new} = \begin{cases} v_{m,N_m}(k_\mathrm{f})_\mathrm{old} & \text{if } q_{\mathrm{dep,demand},n,s}(k_\mathrm{f}) \le q_{\mathrm{dep},n,s}(k_\mathrm{f}), \\ v_{m,N_m}(k_\mathrm{f})_\mathrm{old}\dfrac{q_{\mathrm{dep},n,s}(k_\mathrm{f})}{q_{\mathrm{dep,demand},n,s}(k_\mathrm{f})} & \text{otherwise,} \end{cases}$$

where $v_{m,N_m}(k_\mathrm{f})_\mathrm{old}$ is the value originally computed using equation (3.1). The density of the off-ramp is computed with:

$$\rho_{\mathrm{off},o}(k_\mathrm{f}) = \frac{L_{n,s,-}S_{n,s}(L(k_\mathrm{u}+1)-1)}{L_{\mathrm{km},n,s}\,\lambda_{n,s}} \ , \tag{8.3}$$

where $\lambda_{n,s}$ is the number of lanes in link $l_{n,s}$.

8.4 Overall model

If we combine the model equations presented in Sections 3.2–3.3 for the freeway network, the urban network of Section 8.2, and their interface respectively of Section 8.3, we get a model for the mixed urban and freeway network.

Note that to be able to compute all the variables, some attention should be payed to the order in which they are determined. Now we have to explain how we can compute the variables for freeway step $k_\mathrm{f}+1$ using the variables of step k_f. We will now briefly discuss the order in which the equations should be processed. For the time period $[k_\mathrm{f}T_\mathrm{f}, (k_\mathrm{f}+1)T_\mathrm{f})$ (which corresponds to freeway time index k_f and urban time indexes $k_\mathrm{f}L, \ldots, (k_\mathrm{f}+1)L-1$) all the variables are assumed to be known. These variables are: density, speed, flows, and origin queue lengths for the freeways, queue lengths, free space and arriving vehicles for the urban network, and queue lengths and free space for the ramps. To compute the values of the variables for the next time step, we apply the following computation order:

1. Simulate the urban traffic (with the on-ramp outflows and off-ramp inflows excluded) for urban time steps $(k_\mathrm{f}+1)L, \ldots, (k_\mathrm{f}+2)L-1$. This also gives the arrivals at the on-ramps and the traffic leaving the off-ramps, which makes it possible to compute the free space on the off-ramps.

2. Compute the on-ramp traffic. The amount of traffic that will enter the freeway from the on-ramps, $q_{\mathrm{dep},r,m}(k_\mathrm{f})$, is distributed evenly over the whole freeway time step (or

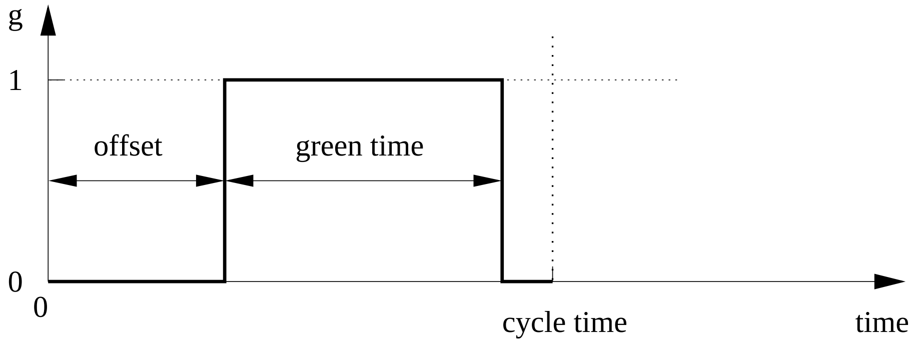


Figure 8.2: The conversion from the offset and the green time to the binary signal g in case the sum of the offset and the green time is smaller than the cycle time.

in urban time steps: over the period given by the urban time steps $(k_\mathrm{f})L, \ldots, (k_\mathrm{f} + 1)L - 1$), and used to compute the evolution of the queue length on the on-ramp.

3. Compute the off-ramp traffic for freeway time step k_f. The traffic is able to enter the off-ramp is computed based on the traffic that wants to enter the off-ramp and the amount of free space that is available at the end of period given by the urban time steps $(k_\mathrm{f})L, \ldots, (k_\mathrm{f} + 1)L - 1$.

4. Now the freeway traffic can be simulated. For $k_\mathrm{f} + 1$ the speeds, flows and densities are determined. The flow $q_{\mathrm{dep,demand},n,s}(k_\mathrm{f} + 1)$ that wants to enter the off-ramp is computed.

This order of computing makes it possible to simulate the whole network without redundant computations and predictions.

8.5 Control strategy

8.5.1 Model predictive control

We will apply a model-based predictive control strategy for coordinated traffic control of mixed urban and freeway networks. The MPC framework was described in Chapter 4. Here we consider only the control signal, objective function, and constraints.

8.5.2 Control signal, objective function, and constraints

The control signal contains the offsets of the phases of each intersection and the durations of the green times. The cycle time is assumed to be fixed, but the extension to a variable cycle time is straightforward. The continuous offset and green times need to be converted into the binary signal $g_{u,s,d}$ before the prediction model can be run. Both the offset and

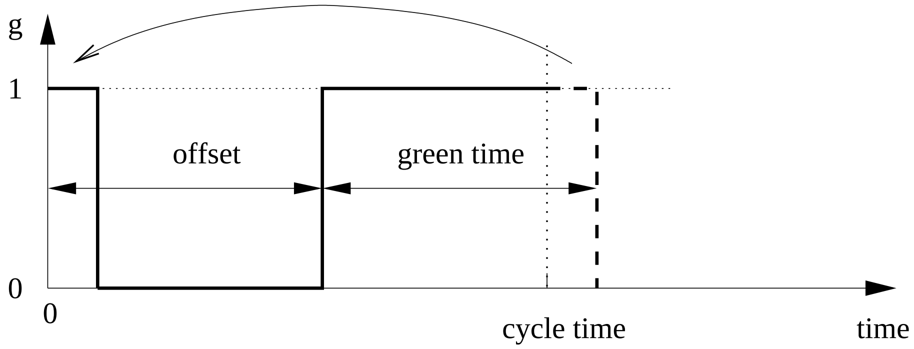


Figure 8.3: The conversion from the offset and the green time to the binary signal g in case the sum of the offset and the green time is larger than the cycle time.

the green time are constrained to be between zero and the cycle time. For the conversion of the offset and the green time we distinguish two situations: when the sum of the offset and the green time is smaller than the cycle time, and when the sum is larger. In the first case the conversion is accomplished as shown in Figure 8.2: the signal g starts with zero (red) up to the offset, where it turns to one (green) for the duration of the green time, and finally it turns to zero (red) again up to the end of the cycle time. The binary values of g can be found by sampling the resulting signal for each k_{u}. In case that the sum of the offset and the green time is larger than the cycle time a similar procedure is followed, and the part of the green phase reaching into the next cycle is transferred to the beginning of the cycle, as shown in Figure 8.3.

We choose the total time spent (TTS) as cost function because it can easily be computed for the urban part as well as for the freeway part. To compute the TTS in the urban part of the network, the number of vehicles in each link, $n_{\mathrm{veh},s,n}$, is required:

$$n_{\mathrm{veh},s,n}(k_{\mathrm{u}}) = L_{s,n} - S_{s,n}(k_{\mathrm{u}}) \ ,$$

where n can be any intersection connected to intersection s. The number of vehicles must be computed for all the urban links, on-ramps and off-ramps.

Assume we are at time $t = k_{\mathrm{c}}^{*}T_{\mathrm{c}}$. The TTS will be computed over a period $t = [k_{\mathrm{c}}^{*}T_{\mathrm{c}}, (k_{\mathrm{c}}^{*} + N_{\mathrm{p}})T_{\mathrm{c}})$. Define k_{u}^{*} and k_{f}^{*} such that $k_{\mathrm{c}}^{*}T_{\mathrm{c}} = k_{\mathrm{u}}^{*}T_{\mathrm{u}} = k_{\mathrm{f}}^{*}T_{\mathrm{f}}$. The end of the period then corresponds to $t = (k_{\mathrm{u}}^{*,\mathrm{end}} + 1)T_{\mathrm{u}} = (k_{\mathrm{f}}^{*,\mathrm{end}} + 1)T_{\mathrm{f}}$, so $k_{\mathrm{f}}^{*,\mathrm{end}} = M(k_{\mathrm{c}}^{*} + N_{\mathrm{p}}) - 1$ and $k_{\mathrm{u}}^{*,\mathrm{end}} = (k_{\mathrm{c}}^{*} + N_{\mathrm{p}})ML - 1$. The TTS in the urban network in the period

$[k_\mathrm{c}^* T_\mathrm{c}, (k_\mathrm{c}^* + N_\mathrm{p})T_\mathrm{c})$ is then given by:

$$\begin{aligned}\mathrm{TTS}_\mathrm{urban}(k_\mathrm{c}^*) = T_\mathrm{u} \sum_{k_\mathrm{u}=k_\mathrm{u}^*}^{k_\mathrm{u}^{*,\mathrm{end}}} \Bigg(&\sum_{(s,n)\in I_\mathrm{urban}} n_{\mathrm{veh},s,n}(k_\mathrm{u}) \ + \sum_{(s,n)\in I_\mathrm{on}} n_{\mathrm{veh},s,n}(k_\mathrm{u}) + \\ &\sum_{(s,n)\in I_\mathrm{urban,orig}} n_{\mathrm{veh},s,n}(k_\mathrm{u}) + \sum_{(s,n)\in I_\mathrm{off}} n_{\mathrm{veh},s,n}(k_\mathrm{u}) \Bigg) \ ,\end{aligned}$$

where I_urban is the set of pairs of indexes (s, n) for all urban links $l_{s,n}$ in the network, similarly I_on is the set of pairs of indexes for on-ramp links, I_off is the set of pairs of indexes for off-ramp links, and $I_\mathrm{urban,orig}$ is the set of pairs of indexes for all urban origins in the network.

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

$$\mathrm{TTS}_\mathrm{freeway}(k_\mathrm{c}^*) = T_\mathrm{f} \sum_{k_\mathrm{f}=k_\mathrm{f}^*}^{k_\mathrm{f}^{*,\mathrm{end}}} \sum_{(m,i)\in I_\mathrm{all}} L_m \lambda_m \rho_{m,i}(k_\mathrm{f}) \ ,$$

where I_all is the set of pairs of indexes (m, i) of all freeway links and segments in the network.

The two above formulas together give the TTS for the entire network. Two positive weighting factors ξ_1, ξ_2 are introduced to give more or less importance to one of the two parts:

$$\mathrm{TTS}(k_\mathrm{c}^*) = \xi_1 \mathrm{TTS}_\mathrm{freeway}(k_\mathrm{c}^*) + \xi_2 \mathrm{TTS}_\mathrm{urban}(k_\mathrm{c}^*) \ .$$

Furthermore, we can impose constraints such as maximum queue lengths at intersections, and at on-ramps or off-ramps, minimum and maximum green times, etc.

8.6 Case study

In order to illustrate the model and the feasibility of the MPC control approach we have selected a test network (see Figure 8.4), that contains some essential elements of mixed urban and freeway networks. The test network consists of a two-way freeway with two on-ramps and two off-ramps. 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 intersections (B, D and E) where no turning is allowed (for simplicity).

Five different (more or less arbitrary) traffic scenarios have been simulated. For each of the scenarios we have considered fixed-time control and MPC. For the fixed-time con-

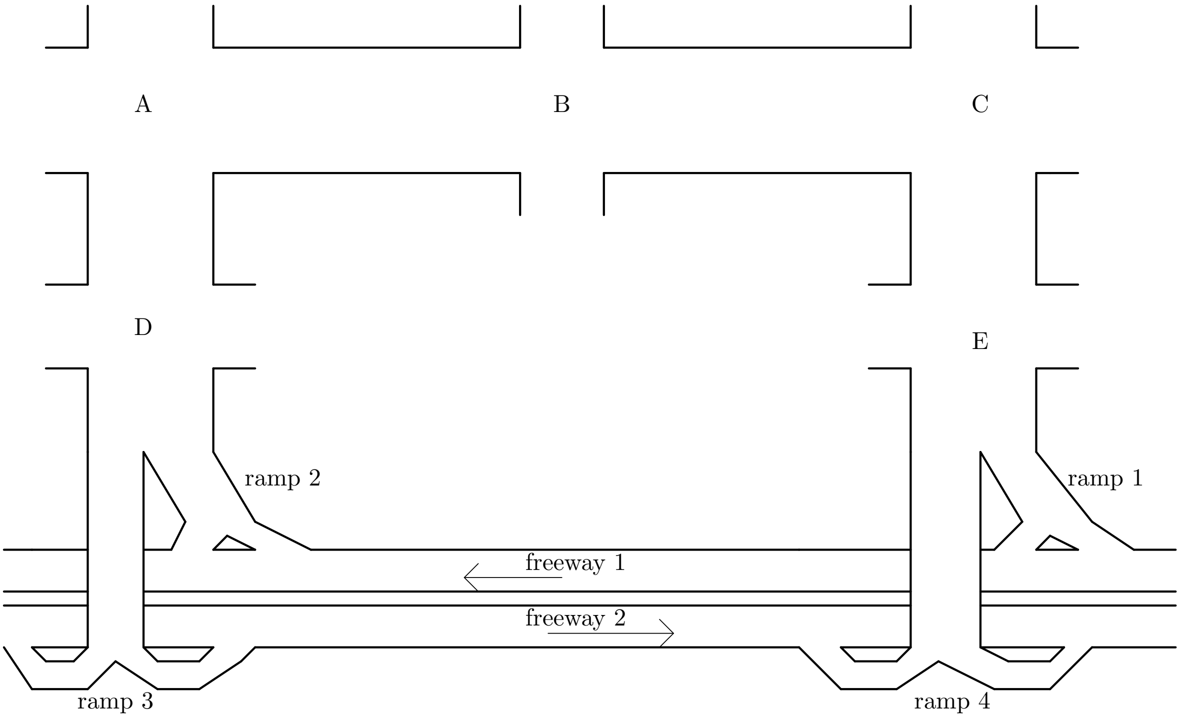


Figure 8.4: Mixed urban and freeway network used in the case study.

trol we have assumed that the intersection control cycle consists of two phases both receiving 50 % of the cycle time. The lights are simulataneously green for the queues in the north and west directions, and simultaneously for the east and west directions. For the MPC control we also have assumed a fixed cycle time, but optimized the phases and the relative green times.

The scenarios have been simulated for 2000 s (approximately 30 min), with $T_\mathrm{c} = 120$ s, $T_\mathrm{f} = 120$ s, $T_\mathrm{u} = 1$ s, $N_\mathrm{p} = 8$ (16 min), $N_\mathrm{c} = 4$ (8 min), $\xi_1 = \xi_2 = 1$ (time spent in urban and freeway are equally important), and a fixed cycle time of 120 s.

The five scenarios represent some typical traffic situations by varying the demands appearing at the origins, and varying the turning rates in the network. The traffic situations for the scenarios are described as follows:

1. **Marginally saturated intersections.** The demands are chosen such that almost all queues are cleared at the end of the green phase, and such that there is no congestion on the freeway.

2. **Morning rush, traffic traveling into the city.** A large part of the freeway traffic leaves the freeway towards the city. This creates a blocked off-ramp and some congestion on the freeway.

3. **Evening rush, traffic leaving the city.** Most traffic at the urban intersections turns towards the freeway, which causes blocking in the urban network, and a traffic jam on the freeway.

4. **Congestion at an urban intersection.** All lights in all directions at intersection D are set to red. Consequently, the urban area gets congested.

5. **Congestion on the freeway.** There is a bottleneck at the exits of both freeways. This creates a congestion on the freeways that eventually propagates to the urban area.

Comparing the fixed-time controller with the MPC controller for these cases we have observed that in general the MPC controller allowed more green time for the directions that will not lead over the congested areas. Consequently, those vehicles that do not pass through the congested part will exit the network earlier, which results in a lower TTS. While this general behavior of the controller is in accordance with what we may expect, further examination of the resulting MPC controller, control signals and traffic behavior is a topic for future research.

We show in Table 8.6 the improvements achieved by MPC for the five scenarios. The goal of these preliminary simulations was to get an impression about the feasibility of the approach, and further investigations are necessary. The results are encouraging, but it is difficult to judge the quality of the improvements. Therefore, the examination of more simple scenarios where the optimal behavior can be predicted beforehand on an analytical basis is also a topic for future research.

	Scenario	fixed time	MPC	improvement
1.	Marginally saturated intersections	642.3	593.1	7.6%
2.	Morning rush, traffic into the city	641.4	601.4	6.2%
3.	Evening rush, traffic leaving the city	730.7	670.8	8.2%
4.	Congestion at urban intersection	1104.1	1061.2	3.9%
5.	Congestion at freeway	963.4	901.3	6.4%

Table 8.1: The TTS and the improvements achieved using MPC instead of fixed time control

8.7 Conclusions

We have proposed a new, extended model for urban traffic that is based on Kashani's model, but that has the following additional features: horizontal queues, a shorter time step, and turning direction dependent queues, which results in a more accurate description of the urban traffic. The urban model was combined with the METANET freeway traffic flow model, and with a model that described the interaction between the urban and the freeway model. This resulted in an overall model for mixed urban and freeway traffic networks. Next, we have used this model as a basis for a model-based predictive control approach for coordinated traffic control of mixed urban and freeway traffic networks. The model and the control approach have been illustrated via a preliminary case study, for which MPC control resulted in a reduction of 4-8 % of the total time spent with respect to fixed-time control. Comparison of the MPC approach with other urban traffic control techniques is a topic for future research.

Chapter 9

Conclusions and further research

9.1 General conclusions

In this thesis we have studied the control approach called *model predictive control* (MPC) that has the properties that are excellently suited for network-oriented traffic control. Here we summarize these properties, and explain their relevance for traffic control:

- **Multiple traffic control measures can be integrated** since MPC can handle multiple-input multiple-output systems. As there are typically several traffic control measures available and the traffic situation (state) is characterized by speeds, flows and densities (and possibly other quantities) the controller should be a multiple-input, multiple-output controller. The MPC controller integrates the traffic control measures such that they serve the same objectives and complement each other if necessary.

- **Traffic control measures can be applied pro-actively and future effects of the traffic control measures can be taken into account** since MPC is predictive. The benefits of a predictive traffic controller are twofold: First, future demands, incoming shock waves, and other disturbances can be taken into account (as far as they are predictable). These disturbances can be predicted based on traffic states in the links directly upstream and downstream from the entrances and exits of the considered network. Second, the dynamics of the process can be taken into account. In many traffic situations it is better to hold back the traffic for a short time in order to achieve a better performance in the future. Examples of such situations are: a shock wave on a freeway where the inflow has to be limited in order to remove the high density region of the shock wave; on-ramp jams where the on-ramp flow has to be limited sufficiently to resolve the jam.

- **User-defined control objectives can be used** since MPC can optimize control inputs according to an externally supplied objective function. The objectives of the traffic control are formulated as a function of the predicted behavior of the traffic network. The overall objective is typically a weighted combination of several objective functions, where the weights represent the trade-off between the objectives. The objectives and the weights can be defined by the user. In traffic these objectives may include terms that aim at efficiency, safety, network reliability, low fuel consumption, low air and noise pollution. In this thesis we have used an objective function representing the total time spent (TTS) by the drivers in the network and a small control variation penalty term. In Chapter 7 we have also added a term that penalizes the travel time prediction errors, to ensure reliability of the dynamic route guidance.

- **Constraints on traffic control signals and traffic states can be implemented** since MPC can handle constraints. In traffic control several constraints on the traffic state or control signal may be present. E.g., we have considered the following constraints: maximum on-ramp queue length, minimum or maximum ramp metering rate, minimum or maximum main-stream metering rate, minimum or maximum speed limit, maximum speed limit drop that a driver may encounter.

- **Unpredictable demands and incoming shock waves can be handled** since MPC has a feedback structure. Feedback in control theory is used to reduce the sensitivity to unpredictable disturbances. In traffic such disturbances are the traffic demand (which may be partially predictable), upstream propagating shock waves, or the stochasticity of the behavior of traffic flow. Feedback in MPC is realized in the rolling horizon framework by updating the system state from measurements every controller step. In this way the unpredictable evolution of the system state due to disturbances and the model mismatch is taken into account.

- **Changing networks parameters due to weather influence or incidents, etc. can be handled** since MPC is adaptive. As the MPC controller uses a rolling horizon framework, the prediction model can also be updated in every controller time step or every N controller time steps. Such an update may be necessary if the behavior of the traffic network has changed significantly compared to the prediction model. E.g., when an incident has occurred, the capacity of the link where the incident has occurred is reduced, which should be taken into account by the controller. Examples of other causes that change the traffic behavior significantly are road works and weather influences. We have not considered adaptivity in this thesis. As adaptivity is powerful property of MPC we included the examination in the topic for further research (see also Section 9.4).

The developed MPC framework was applied to several traffic scenarios and all of them resulted in a significant improvement of the TTS compared to the uncontrolled case. The main reason for the improvements was the resolution of the capacity drop in the high-density areas. While the capacity drop is only one of the possible causes for network performance degradation, the MPC approach can be applied similarly to scenarios where other causes play a role, such as blocking or route choice.

However, a critical note about the numerical value of these improvements should be made here. As we explained in Section 1.1.4 the TTS is strongly related to the outflow of the network. Since traffic control cannot improve the outflow equally for any scenario, the improvement also depends on the chosen or given demand scenario. This means that it is meaningful to compare the performances for the controlled and uncontrolled cases for one given scenario, but it is not meaningful to compare performances for different demand scenarios. Furthermore, it is not meaningful to compare improvements for different networks, simulation lengths, or traffic simulation models, even if the same traffic control measures are used. For this reason, it would be useful for the comparison of alternative traffic control methods to define a number of standard benchmark problems.

9.2 Conclusions per chapter

Now we present the conclusions of each chapter in detail.

- In Chapter 3 we have given an overview of existing traffic flow models, where the models were classified according to *application area*, *level of detail*, and *deterministic versus stochastic* or *continuous or discrete* process representation. The choice for a traffic model should be based on efficiency and accuracy considerations and on the typical phenomena that a certain class of traffic flow models can or cannot represent.

  We have also introduced the METANET model, which is a macroscopic, deterministic and discrete model suitable for modeling freeway traffic. To this model we have made the following extensions:

  - We have formulated an explicit model for dynamic speed limits, such that the speed limit influences the traffic only if the speed in the unlimited case would be higher than the displayed speed limit.
  - We have formulated a model for main-stream metering, which is similar to ramp metering, except that the traffic dynamics upstream the main-stream metering device is not given in terms of a queue, but in terms of speed, flow and density.
  - We have formulated a model for the main-stream origins which have different dynamics than on-ramps. The main difference is that a main-stream origin has a different lay-out which allows less inflow in congested situation.

  - We have separated the anticipation constant into two constants that represent the anticipation behavior at the head and the tail of shock waves. This gives a better reproduction of shock waves and the capacity drop.
  - We have added a new formulation for the downstream boundary condition, which expresses free flow downstream conditions except for some upstream propagating shock waves.

  The extended METANET model was used in the simulations in Chapters 6–8.

- In Chapter 4 we have presented the MPC framework. Guidelines for tuning the prediction horizon and the control horizon were discussed. The length of the prediction horizon is a trade-off between complexity and the requirement that it should be long enough to reproduce all important process dynamics. The length of the control horizon is a trade-off between complexity and performance.

  We have also discussed the advantages (feedback, easy tuning, prediction, multiple inputs, multiple outputs, easy constraint formulation, modularity, and adaptivity) and the disadvantages (complexity, precise model and precise disturbance prediction necessary, stability difficult to prove, optimality not guaranteed) and discussed solutions to some of the disadvantages.

- In Chapter 5 we have discussed necessary conditions for effective traffic control, under the assumptions that

  - the total time spent is to be minimized,
  - the reason for the performance degradation is the capacity drop or the blocking of traffic not traveling over the real bottleneck location.

  The conditions necessary for effective control include

  - the presence of the capacity drop or blocking in the real traffic situation,
  - if model-based predictive control is applied, the ability of the traffic model to reproduce these phenomena, with a sufficiently high accuracy,
  - the possibility to sufficiently reduce the inflow of the congested area,
  - the network boundaries should be chosen such that the vehicles that are delayed by traffic control are inside the network,
  - the network boundaries should be chosen such that the roads downstream can accommodate the improved traffic flows,
  - the presence of traffic demands for which control is useful.

From these general conditions specific conditions are derived for speed limits and ramp metering.

For speed limits the traffic flow should be between the capacity flow and the dropped flow, which results in a metastable state where the unstable shock wave can be converted into a wider but stable disturbance with a higher outflow.

For ramp metering the analysis of freeway and ramp demands shows that the region for which ramp metering can improve the total time spent is relatively small compared to the region where congestion occurs. The main reason for this is that usually the ramp metering rate is bounded from below, and as a result the inflow of the congested area cannot be restricted sufficiently. Ramp metering will be effective if there is a ramp jam and the condition

$$q_{\mathrm{fw,dem}} < q_{\mathrm{drop}} - q_{\mathrm{r,min}}$$

is satisfied, where $q_{\mathrm{fw,dem}}$ is the freeway demand, q_{drop} the outflow of the ramp jam (after the capacity drop), and $q_{\mathrm{r,min}}$ the minimum ramp flow.

- In Chapter 6 we have examined several set-ups with speed limits and other control measures. We have applied the MPC framework to several traffic problems that can benefit from the use of speed limits. For all problems the main purpose of the control was to find the control signals that minimize the total time that vehicles spend in the network (i.e, TTS).

  In general, we can say that by using dynamic speed limits not only as a speed limitation, but also as a flow limitation – possibly in combination with other flow limiting measures, such as ramp metering –, more traffic jams can be prevented or resolved, as in this way not only the speed and the flow, but also the density can be controlled better.

  In Section 6.1 we have dealt with the integrated control of ramp metering and speed limits, where the speed limits can prevent a traffic breakdown when ramp metering only is insufficient. Since the main effect of the speed limits in this section is to limit the flow when necessary, in Section 6.2 this set-up was compared with a set-up where the speed limits are replaced by main-stream metering.

  Speed limits proved to be useful when ramp metering was unable to keep the on-ramp segment of the freeway congestion free. The cases ‘ramp metering only’ and ‘coordinated ramp metering and speed limits’ were compared for a typical demand scenario. Compared to the ‘no control’ case the TTS improvement in the ‘ramp metering only’ case was 5.3 % and in the ‘coordinated ramp metering and speed limits’ case 14.3 %.

  Since the main effect of the speed limits in such situations is that they hold back the traffic, we have compared the ‘coordinated ramp metering and speed limits’ scenario with the ‘coordinated ramp metering and main-stream metering’ scenario (where the speed limits are replaced by a main-stream metering device). The comparison was made for several bounds on the main-stream metering signal. If the bounds were chosen such that the maximal flow limitation is equal to the maximal flow limitation of the speed limits, the improvement was close to the improvement achieved by the speed limits. If the bounds were chosen such that the flow limitation can be stronger, the improvement of the TTS was even better. The interpretation of these results is that the choice between speed limits and main-stream metering should be made based on the demands on the on-ramp and the freeway. If the speed limits can limit the flow sufficiently then speed limits should be used. If not, main-stream metering should be used, which can limit the flow much more.

The preference for speed limits is motivated by the advantages of speed limits compared to main-stream metering. First, the maximum flow for main-stream metering is limited to approximately 75 % of the nominal capacity of the freeway when main-stream metering is on. This can cause oscillatory behavior when only a lighter flow limitation is necessary. Second, main-stream metering limits the flow only at one location which may cause shock waves. Opposed to this, speed limits can limit the flow more gradually, and since there are often installed more speed limit signs on freeway stretch they can prevent shock waves.

In Section 6.3 we have applied speed limits to reduce or eliminate shock waves on freeways. We have shown that dynamic speed limits are very suitable to prevent or eliminate shock waves on freeways.

Note that for speed limits systems it is important to make a distinction between approaches that aim at homogenizing the traffic flow and that aim at the resolution of shock waves or jams. While in theory the homogenizing approach is promising, field studies show that the achievable improvement is negligible [72, 33] The main disadvantage is that these systems cannot resolve congestion after breakdown has occurred. The approaches that aim resolution of shock waves or jams use speed limits that are low enough to limit the inflow of the congested area while the homogenizing approach uses speed limits that are above the critical speed. The approach discussed in Section 6.3 aims at the removal of shock waves.

The MPC framework was applied to a benchmark network consisting of a link of 12 km, where 6 segments of 1 km are controlled by speed limits. The controller was evaluated for two different downstream density scenarios for the shock wave entering from the downstream end of the link. Simulations were run with and without safety constraints, and with continuous versus discrete speed limits. In all cases the coordinated speed limits eliminated the shock wave, and the TTS was improved on the average by 19 %.

- In Chapter 7 we have introduced a new route guidance concept that makes it possible to use DRIPs as a traffic control measure (instead of merely informing), while providing accurate travel time predictions at the same time.

  We have considered the problem of MPC traffic control with ramp metering and dynamic route guidance as the traffic control measures.

  The first issue addressed in this chapter was the integration of dynamic route guidance and ramp metering. The approach we have chosen is to use the dynamic route information panel (DRIP) as both a control tool *and* an information provider to the drivers, and ramp metering as a control tool to redistribute the delays over the on-ramp and the freeway. The drivers' reaction to the travel times shown on the DRIPs is modeled by the logit model. The travel times shown on the DRIPS are optimized travel times, which are chosen such that the reactions of the drivers and the control actions of ramp metering are taken into account. This results in one optimization that optimizes both the ramp metering and the travel times shown on the DRIPs at the same time such that on the one hand, the total time spent in the network is reduced by optimally rerouting traffic over the available alternative routes in the network, but on the other hand, the difference between the travel times shown on the DRIPs and the travel times actually realized by the drivers is also kept as small as possible.

  The second issue addressed in this chapter is whether the proposed approach leads to an improvement of the traffic system in congested situations such as described in the case study. The simulations show that rerouting of traffic and on-ramp metering using MPC leads to an improvement in performance of 28.8 % for the case study.

- In Chapter 8 we have proposed a new, extended model for urban traffic that is based on Kashani's model, but that has the following additional features: horizontal queues, a shorter time step, and turning direction dependent queues, which results in a more accurate description of the urban traffic. The urban model was combined with the METANET freeway traffic flow model, and with an interface that described the interaction between the urban and the freeway model. This resulted in an overall model for mixed urban and freeway traffic networks that is suitable for MPC. Next, we used this model as a basis for an MPC approach for coordinated traffic control of mixed urban and freeway traffic networks. The model and the control approach are illustrated via a case study, for which MPC control resulted in a reduction of 4-8 % of the TTS with respect to fixed-time control.

- In Appendix A we describe the results of a joint project with the Traffic Management group of Prof. H.J. van Zuylen in which we have developed a fuzzy decision support system (FDSS-TC) for traffic control centers. This system is part of a larger traffic decision support system that assists operators of traffic control centers when selecting the most appropriate traffic control measures to efficiently manage non-recurrent congestion. The FDSS-TC uses a case base and fuzzy interpolation to generate a ranked list of combinations of control measures and their estimated performance. The predictions made by the case-based reasoning system can be made more precise by adding new cases. An important feature of the system is that the performance function is not fixed but consists of a weighted combination of several partial performance measures. In addition, the weights of this combination can be changed on-line depending on the current traffic management policy and on other considerations. Since the case base can be generated off-line, the FDSS-TC reduces the time that is needed to determine the optimal traffic control for a given situation by limiting the number of combinations of control measures for which on-line traffic simulations should be performed in the traffic control center. At a later stage the system can be extended with a fuzzy module that incorporates expert knowledge, and with an adaptive learning module.

9.3 Contributions to the state of the art

The contributions of this thesis to the state of the art can be summarized as follows:

- In Chapter 3 we have extended the METANET model with the modeling of: dynamic speed limits, main-stream metering, main-stream origin, differentiation between the anticipation behavior at the head and the tail of shock waves, a new formulation of the downstream boundary condition.

- In Chapter 4 we have applied the MPC framework to traffic systems and presented heuristic tuning rules for traffic control problems.

- In Chapter 5 we have discussed the necessary conditions for successful traffic control in case of ramp metering, and dynamic speed limits.

- In Chapter 6 we have examined several set-ups with speed limits and other control measures, such as the integrated control of speed limits and ramp metering, and the integrated control of main-stream metering (replacing the speed limits) and ramp metering.

- Also in Chapter 6 we have applied speed limits against shock waves. The control concept is different from homogenization: it aims at resolving the high density region of the shock wave by flow limitation, and at restoring the dropped flow to the capacity flow. We have also presented a method to find discrete speed limit values, and introduce constraints that ensure the safe operation of speed limits.

- In Chapter 7 we have introduced a new route guidance concept, that makes it possible to use DRIPs as a traffic control measure (instead of merely informing), while providing accurate travel time predictions at the same time.

- In Chapter 8 we have developed an urban traffic model and an interface in order to combine it with the freeway model METANET, such that the overall model is suitable for MPC.

- In Appendix A we have developed a prototype decision support tool for operators in traffic control centers, which is based on case-based reasoning and fuzzy interpolation.

9.4 Further research

In this section we present topics for further research. The topics are grouped into the categories *modeling, control, the investigation of a wider range of scenarios,* and the *necessary conditions* for successful control.

Topics related to **modeling:**

- **Validation of modeling assumptions.** In Chapter 3 we made modeling assumptions for the extended METANET model: speed limits, main-stream metering, anticipation constants, downstream boundary condition. These modeling assumptions will be validated with real data.

- **Comparison with other models.** Another topic for further research is the study of other macroscopic traffic flow models, such as the gas kinetic model of Helbing [65] or Hoogendoorn and Bovy [76], and the combined urban-freeway model METACOR for MPC traffic control applications.

- **Simulation with calibrated models.** Whichever traffic flow model is chosen, the model should be calibrated and validated with real data. In the future, the effectiveness of MPC for the benchmark scenarios of this thesis will be studied with model parameters extracted from real, measured data. It is expected that MPC will improve performance for any set of model parameters that result in a model that can reproduce the capacity drop phenomenon. Furthermore, the necessity of on-line calibration will be examined.

- **Modeling and control of other measures.** Other control measures will be examined that can potentially improve the traffic flows such as peak lanes, reversible lanes, and the ‘keep your lane’ directive.

- **Modeling and control of capacity drop at other bottlenecks.** Besides on-ramps and shock waves there are other freeway bottlenecks known that may cause a capacity drop, such as off-ramps, merges, diverges, bridges, tunnels, curves, and grades. The explicit modeling of these bottlenecks may improve the quality of the predictions and therefor result in better control performance.

- **Explicitly taking into account incidents.** A significant part of the traffic jams is caused by incidents. Taking into account the relation between safety and efficiency may improve the performance of traffic control.

  There is a twofold relation between safety and efficiency. If traffic is safer then there are less incidents, and consequently the traffic flow is higher. On the other hand, a more efficient traffic flow is usually achieved by a more stable flow (where traffic control prevents e.g., breakdowns), which can be expected to result in less incidents. Taking into account both effects in the MPC prediction model results in traffic control that is safe and efficient, or if necessary (if safety and efficiency are conflicting) in traffic control that finds a trade-off between safety and efficiency. Probabilistic modeling may be useful to model these effects.

Topics related to **control:**

- **Real-world testing of the MPC approach.** Altough the results in this thesis are very promising, the ultimate proof a traffic control approach is the testing in a real-world situation. Further investigations towards the real-world applicability are necessary.

- **Faster tuning.** In Chapter 4 we gave some heuristic arguments for the tuning of N_c, N_p and the weights ξ_i. However, the tuning process still involves trial-and-error. It is difficult to find exact tuning rules for a non-linear process, but better tuning rules would be certainly useful. E.g., from the simulations it seems that the sensitivity of the performance to a change in N_c is quite small if N_c is greater than a given lower bound. Estimating this lower bound is a question for future research.

- **Further examination of speed limit rounding.** The heuristic rounding presented in Section 6.3 performed satisfactorily for the benchmark scenarios. However, the performance degradation caused by rounding may depend on the traffic demand scenario, network topology or other traffic control measures used in the same network. Further examination of the trade-off between efficiency and optimality for rounding versus full discrete optimization is necessary.

- **Other objective functions.** In this thesis we used as main objective the TTS, the prediction error made by the DRIPs, and a control variation penalty. However there are also other options possible, such as:

  - the functions presented in Chapter 4,
  - the total delay per kilometer in a network,
  - a term that expresses air and noise pollution,
  - a term that expresses safety,
  - a term that penalizes too low speeds, in order to guarantee travel times on certain links or routes,
  - a term that expresses fuel consumption.

  These terms can be included in the overall objective function and the effect of different weighting strategies can be examined.

- **Investigating the effects of imprecise expected demand.** In the benchmark scenarios we assumed the disturbances (demands and downstream shock waves) to be known. In reality the average (historical) demand is typically known, and the daily variation of the demand and incoming shock waves are unknown. It is interesting to examine the effect on the performance of realistic unpredictable disturbances.

  A possibility to predict in the short-term (in the order of minutes) the daily variations of the demands and incoming shock waves is to take the traffic state in the links upstream the entries and downstream the exits into account. In this way the future inputs of the controlled network can be predicted better and the controller can act pro-actively. An interesting question is to what extent (freeway length) is it necessary to know the traffic states to achieve a satisfactory performance? The approach could be similar to that in [122], where the consequences of the partially unknown rainfall loads are examined for the MPC control of a sewer network.

- **Investigating adaptivity.** Adaptivity is easily implemented in the MPC framework. In case of weather influences, roadworks, or incidents the traffic behavior may change significantly and a re-parametrization of the prediction model may be necessary. This can be achieved by on-line calibration or direct intervention in the prediction model. E.g., if there is an incident and some lanes are closed, this change can be directly introduced in the model.

- **Investigating the effects of a model mismatch.** For the benchmark scenarios we assumed that the controller model exactly matches the process. In reality there is always a mismatch between the two. Bellemans [11] studied the effects of a model-mismatch for a ramp-metering set-up. Bellemans suggests to update the model parameter every 30 min. to minimize the model mismatch. However, MPC is in general known to be robust to model mismatch. Further examination of this robustness can give more information about the sensitivity to the mismatch of certain parameters and the necessary update interval. To study the effects of the model mismatch the process model can be replaced by another macroscopic or microscopic traffic model.

- **Investigating the effects of unmeasurable states, incomplete and noisy measurements.** Another source of performance degradation can be the unmesurable states, or incomplete or noisy measurements. In all of these cases the states have to be estimated, which in general will introduce an estimation error. The investigation of the sensitivity to the esitmation error will give more information about the real-world applicability, because traffic measurement are known to contain errors and data is missing regularly.

- **Comparing with alternative control methods.** The MPC control of ramp metering, route guidance and speed limits used in the benchmark scenarios in Chapters 6 and 7 can be replaced by other existing control approaches. E.g., MPC for ramp metering can be replaced by ALINEA or the RWS strategy; MPC for route guidance can be replaced by the predictive feedback approach of Wang*et al.* [165], and the MPC for the speed limits can be replaced by the approach of Alessandri*et al.* [2] can be used. The performances of these approaches can be compared with the MPC controller.

- **Effect of switching scheme for ramp-metering.** As pointed out in Section 2.1 the capacity of some ramp metering devices is around 75 % of the road capacity. Therefore, the on/off switching scheme of the ramp metering device is relevant and should be incorporated into the controller design procedure.

- **Network reliability.** In a dense traffic network the relationships between the different parts of the network may be strong. A few incidents on crucial location may block large parts of the network. Therefore, traffic performance is also characterized by network reliability. Network reliability can be interpreted in different ways, such as disturbance rejection (what disturbance can be handle without serious performance degradation) or as the speed of recovery after a breakdown. The relation between dynamic traffic control and network reliability is an important question for this topic.

- **Combining traffic control and user equilibrium process.** It is generally assumed that drivers' route choice tends towards the user equilibrium. When dynamic traffic management is applied it may result in traffic control measures that structurally change the travel times on certain routes. As a reaction to this, drivers may change their route choice on long term (longer than the time scope of the dynamic traffic control). The traffic control measures are again adjusted to the new route choices, and so on. This may cause instabilities [154]. This is typically solved by a bi-level optimization problem resulting in an anticipative control framework, where at the top level the traffic control problem is solved, and at the bottom level the user equilibrium assignment is solved. However, this approach assumes that the user equilibrium is always realized, while in practice this is not plausible because the route choice process is slower than the traffic control process. A better approach may be to force the controller to an (average) behavior that results in the (or better, a) desired assignment. The inclusion of a term in the objective function that expresses the quality of the resulting assignment from the control action is an option.

- **Efficient implementation / efficient algorithms.** For larger networks, more control measures, longer control horizons and prediction horizons the computation complexity may become a too high for real-time control. Therefore, there is a need for a more efficient implementation of the prediction model and the controller. An important subtopic here is the examination and comparison of different optimization methods to solve the MPC problem.

Topics related to the investigation of the effectiveness of MPC for a **wider range of scenarios:**

- **Off-ramp blocking.** Since upstream of many on-ramps an off-ramp is located, and on-ramp queues often block these off-ramps, it would be interesting to also include off-ramps in the benchmark network of Section 6.1. In this way the effect of integrated control of speed limits and ramp-metering on off-ramp blocking could be investigated. In practice the blocking of off-ramp traffic is a typical source of network performance degradation; the theoretical improvement that can be achieved by preventing off-ramp blocking is analyzed by Papageorgiou and Kotsialos [130]. Besides off-ramps also other locations where blocking can occur are interesting for further research.

- **Shock waves emerging from an on-ramp.** Since in practice on-ramps are a typical source of shock waves, the joint control of the on-ramp jam and the shock waves emerging from the on-ramp would be an interesting problem. The network setup could be similar to that in Section 6.1 but with a longer freeway stretch with more speed limits to be able to eliminate the shock waves.

- **A simple urban-freeway scenario.** The disadvantage of the benchmark scenarios presented in Chapter 8 is that by their complexity it is difficult to assess the meaning of the improvement achieved by MPC. For better comparison it would be useful to test the approach on a simpler network where the achievable improvement can be expressed analytically.

- **Preventing rat running.** Rat running can be a problem when there exists an alternative route through urban areas parallel to a freeway route. When the freeway is congested drivers on the freeway may chose the urban route in order to reduce their travel time. This can have a serious impact on local traffic (efficiency, safety, and noise and air pollution). To discourage drivers to take the urban route, a delay could be introduced on the urban route by e.g., traffic light settings, that is at least as long as the delay caused by the freeway congestion. If there are more traffic lights on the route the delay could be distributed over the traffic lights such the negative impact on the local traffic is minimized.

- **Larger networks.** Another interesting and relevant topic for further research is the study of larger networks, such as ring roads around cities or areas where several cities are connected by freeways, and where traffic typically travels through the whole network and the interrelations are strong. An important aspect of the study of larger networks will be the trade-off between efficiency and optimality.
- **MPC controllers for sub-networks.** The computational complexity for an MPC controller for large networks is expected to be too high for real-time control. One way to reduce complexity is to define (partially overlapping) sub-networks with for each sub-network a separate MPC controller. Coordination between the sub-networks can be handled by multi-agent techniques (e.g., communicate future demands or available capacities) or hierarchical control (give set-points for flows, speeds, etc.). If the control measures are in overlapping regions of the sub-networks, special care should be taken to resolve possible conflicts between control signals.

Topics related to the **necessary conditions:**

- **Consideration of other bottlenecks.** The presence of other types of bottlenecks (such as bridges, tunnels, off-ramps, merges, diverges, curves, grades), may also determine the conditions for successful traffic control. Examination of the location and capacity (or capacity drop) can also be included in the considerations of the necessary conditions.
- **Extension to networks.** In Chapter 5 we have examined only simple cases. The extension of the considerations to networks with given origin-destination matrices and route choice would give more insight in achievable improvement in more complex networks (including urban networks). Also blocking, route guidance (e.g., to reduce the effect of blocking), or multiple control measures, such as ramp metering and dynamic speed limits, will be considered in a network context.
- **Incidents and weather conditions.** Incidents and weather conditions are reasons that justify dynamic traffic management, because they both may change the capacity (or other properties) of certain parts of the network. However, we have not considered a dynamically changing network. More insight into the possibility of improving traffic conditions under given changes in the network would be useful.

- **Study the capacity drop and metastability phenomena.** One of the conditions in Chapter 5 was that the controller model should be able to reproduce the capacity drop and metastability phenomena. To the author's best knowledge the only way to determine whether a traffic model is able to reproduce these phenomena is by simulation. The development of an analytical tool to determine whether a given model can reproduce these phenomena would be very useful.

Appendix A

FDSS-TC: A Fuzzy Decision Support Systems for Traffic Control Centers

In this Appendix[1] we present as an alternative to the MPC approach a decision support system for operators in traffic control centers. The main advantage of a decision support over the MPC approach is that the traffic operator make his own choice between the proposed control scenarios while for the MPC approach there is no direct facility for operator intervention. However, since the feedback loop including a traffic operator is much slower than the feedback loop of the MPC controller, and since the control scenarios are precise, it can be expected that decision support systems will not achieve the optimality that could be achieved by MPC. Nevertheless, we present the decision support approach in this Appendix, because in practice a ‘sufficient performance level’ or a ‘level of service’ is often accepted instead of the optimal performance.

The fuzzy decision support system presented in this Appendix is a part of a larger traffic decision support system (TDSS) that can assist the operators of traffic control centers when they have to reduce non-recurrent congestion using a network-wide approach. The kernel of the system is a fuzzy case base that has been constructed using simulated scenarios. By using the case base and fuzzy interpolation, the decision support system generates a ranked list of combinations of traffic control measures. The best combinations can then be examined in more detail by other modules of the TDSS that evaluate or predict their performance using macroscopic or microscopic traffic simulation. At a later stage the fuzzy decision system can be complemented with an adaptive learning feature, with a set of fuzzy rules that incorporate heuristic knowledge of experienced traffic operators, and with a hierarchical decision structure (to address scalability problems).

[1] The material presented in this appendix is a result of a joint project with the Traffic Management group of Prof. H.J. van Zuylen.

A.1 Introduction

Contemporary traffic control centers use dynamic traffic management measures such as ramp metering, DRIPs (dynamic route information panels) or VMSs (variable message signs) to control traffic flows on highways and urban ring roads. The DRIPs can be used to display queue length information or indications of congestion, traffic jams and alternative routes. VMSs can be used to show dynamic speed limits per lane, advisory speeds, or lane closures. Recurrent congestion can usually be managed satisfactorily because traffic operators have gained sufficient experience to select the appropriate combination of available control measures. However, operators in traffic control centers often face a difficult task when non-recurrent, non-predictable congestion occurs (e.g., as a consequence of an incident or due to unexpected weather conditions). In such situations, local measures are usually not sufficient and often an intervention at the network level is required to manage congestion and to return to a normal traffic situation.

The effects of non-recurrent congestion can be attenuated by redirecting the traffic flows in a larger part of the network. The operator of the traffic control center then has to assess the severity of the situation, predict the most probable evolution of the state of the network, and select the most appropriate measures. This is a complex task, which requires expert knowledge and much experience, which can often only be obtained after extensive training. As a result, the approaches used by human traffic operators are in general neither structured nor uniform. Therefore, our aim is to provide a decision support tool to assist the operators in their decisions when they have to take measures to deal with non-recurrent, non-predictable congestion. This decision support system should help the operators to react in a uniform and structured way to unusual situations. Since we want to create a decision support system that allows for an easy and smooth interaction with human operators, and that uses a decision process that is both intuitive and can be explained in linguistic terms, we have opted for a decision support system based on a fuzzy knowledge base. Furthermore, in order to increase the acceptance of the decision support system by the traffic operators, it is designed as an advisory and analysis tool that assists the operators (instead of trying to replace them).

In short, the system works as follows. Given the current state of the network and the optimization criterion (such as minimal total travel time, maximal throughput, or a weighted combination of several criteria), the fuzzy decision support system generates a ranked list of the best control measures and presents them to the human operator of the traffic control center. If necessary, the effect of these measures on the current traffic situation can be simulated by an external simulation unit. The resulting output of the overall system is a characterization of the actions that can be taken and their predicted effectiveness in the current situation. The system described in this appendix operates in a multi-level control framework. At the lowest level we have semi-autonomous local traffic controllers for, e.g., traffic lights or ramp metering. At a higher level the operation of several local traffic controllers is coordinated or synchronized by supervisory controllers. The role of the fuzzy decision support system in this set-up is to suggest whether a particular local traffic controller or control measure should be activated or not.

Several authors have described decision support systems for traffic management, such as FRED (Freeway Real-Time Expert System Demonstration) [141, 142, 173], or the Santa Monica Smart Corridor Demonstration Project [6, 144]. However, these architectures do not use fuzzy logic in their decision process. Since we want a system with an intuitive operation process that is able to generate decisions in cases that are not explicitly covered by the knowledge base, we have opted for a fuzzy system. Other fuzzy decision support systems for traffic control have been developed in [26, 96, 113]. The TRYS system described in [26, 113] is an agent-based system for urban freeway control. The network is divided in possibly overlapping regions and to each region an agent is assigned. The agent has to detect and diagnose traffic problems in its region and subsequently suggest possible control measures to a higher level coordinator, that then decides which action will actually be taken. The decision process in the TRYS system is based on knowledge frames, and some of these frames use fuzzy logic. The paper [96] describes a fuzzy logic control architecture that can be applied in existing traffic control systems on a multi-lane highway with VMSs. This system uses fuzzy logic to incorporate the experience of human traffic operators.

The main aim of the system presented in this appendix is to make the process of on-line, real-time selection of the most appropriate traffic control measures more efficient. To that extend we use fuzzy interpolation (based on a case base) to select a limited number of best combinations of traffic control measures for a given traffic situation. In that way we can limit the number of possible combinations of traffic control measures that have to be simulated on-line or that have to be further assessed by the traffic operators.

This appendix is organized as follows. First, we give a short introduction to fuzzy sets. Next, we describe the overall traffic decision support system of which the fuzzy decision support system is a subsystem. Next, we describe the set-up and operation of the fuzzy decision support system and a small prototype we have developed to assess the technical feasibility of the proposed approach. Finally, we propose possible extensions of the current system.

A.2 Fuzzy set theory

In recent years, fuzzy set theory has found a large number of applications and has thus become one of the more successful methods to deal with complexity, uncertainty and imprecision in various systems and processes.

Conventional, crisp sets are characterized by the property that an object either belongs to the set or not. However, many concepts, such as congestion, do not lend themselves very well for a representation by crisp sets. Indeed, assume — for the sake of simplicity

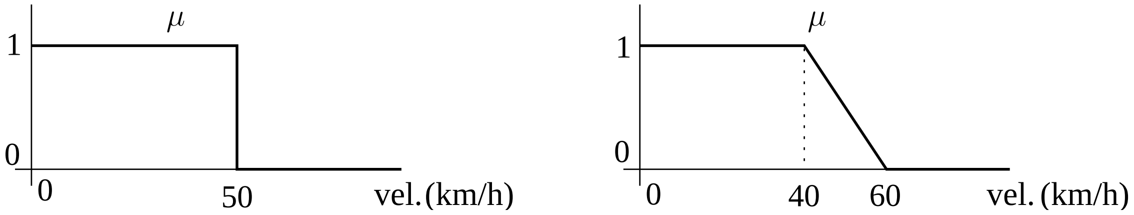


Figure A.1: Congestion defined using a crisp set (on the left) and a fuzzy set (on the right). The membership function μ then expresses the degree of congestion.

— that we use the average traffic velocity to determine whether there is congestion or not on a given highway segment. In that case we could select a threshold value, say 50 km/h, and say that there is congestion if the average velocity is below 50 km/h, and no congestion if the average velocity is above 50 km/h (cf. Figure A.1 on the left). This would imply that for an average speed of 51 km/h there is no congestion, whereas for 49 km/h there is congestion. However, such a small difference in the average velocity does not make a significant difference in the driver's perception of congestion.

So a definition in which we have a gradual transition from congestion to non-congestion seems to be more plausible than the sudden transition at 50 km/h. A fuzzy set exhibits such a gradual transition from membership to non-membership. As a consequence, fuzzy sets are a much better model for concepts such as "congestion". A fuzzy set is characterized by a *membership function* that expresses the degree to which an object belongs to the set. Consider, e.g., the situation on the right in Figure A.1 where the bold piecewise-linear curve represents the membership function. Using this membership function we say that for average velocities below 40 km/h congestion is definitely said to occur; above 60 km/h we say that there is no congestion; and in the region between 40 km/h and 60 km/h, the degree of congestion decreases gradually from 1 to 0.

The range of a membership function is always (a subset of) the interval [0,1]. The position and the shape of the membership function depends on the particular application and context. Commonly used types of membership functions are triangular, trapezoidal, bell-shaped, and singleton functions. Singleton membership functions correspond to crisp singleton sets, since a singleton membership function $s(\cdot)$ is 0 everywhere except for the center point c, where the function value of s is 1; so $s(x) = 0$ if $x \neq c$ and $s(c) = 1$.

When describing the behavior of a system or a process, associating a linguistic term with a fuzzy set makes a link to a linguistic description, which corresponds more closely to the human way of reasoning and thinking than a mathematical model. This leads to the concept of linguistic variable, i.e., a variable that instead of numbers can take on words as its value. A linguistic value is assigned a meaning depending on the context and is represented by a fuzzy set. E.g., traffic density could be classified as "uncongested", "regular", "dense" or "congested".

For more information on fuzzy logic and fuzzy set theory the interested reader is

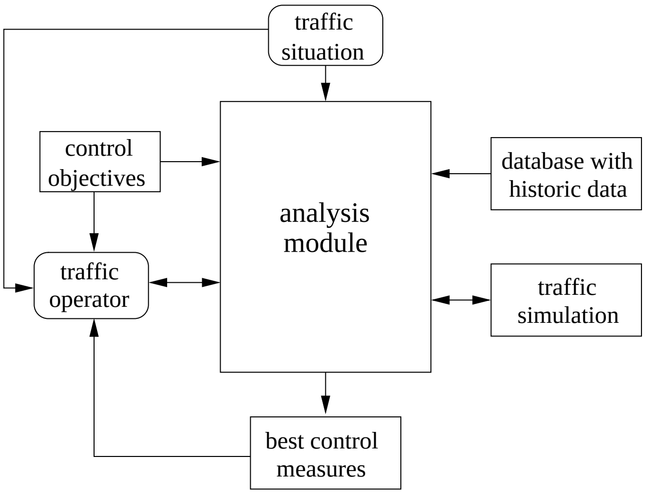


Figure A.2: The overall traffic decision support system (TDSS). The fuzzy decision support system (FDSS-TC) presented in this appendix is a part of the analysis module.

referred to [86, 116, 135] and the references therein.

A.3 Overall framework

The system we are developing is a part of a larger traffic decision support system (TDSS) [87] that is currently being developed by the Dutch Ministry of Transport, Public Works, and Water Management. The structure of this system is depicted in Figure A.2.

The inputs for the TDSS are indicators of the current traffic situation, such as traffic densities, average speeds, traffic demand, time of day, weather conditions, incidents, etc. Furthermore, the traffic operator can provide or adjust additional parameters and specify which control objective should be used. Based on the measurements, historic data and traffic simulation, the system predicts the future traffic situation (more specifically, the TDSS uses the METANET macroscopic flow model [109] to make a forecast of the traffic situation). In that way we can also predict the performance of the traffic control measures (such as DRIP messages, ramp metering, or lane closures) that will be applied. Since in general a large number of traffic control measures (and combinations of them) are possible, it is not tractable to evaluate all possible combinations of traffic control measures using macroscopic or microscopic traffic simulation. Therefore, in practice only a limited number of combinations can be simulated. The aim of the subsystem we are developing is to limit the number of possible combinations of control measures that should be simulated by using an intelligent decision support system to rank the possible combinations of control measures and to present the operator with a limited number of possibilities that deserve further examination (via a quick assessment based on the operator's experience or by a real-time traffic simulation program). Afterward, the operator can select the most appropriate control strategy. Once the operators are familiar with the system and get more confidence in the output, the best control strategy presented by the decision support system can be implemented automatically, without further human intervention.

A.4 The fuzzy decision support system

A.4.1 Structure

The fuzzy decision support system (FDSS-TC) selects optimal combinations of traffic control measures for a given situation by using a weighted performance index J, defined as

$$J = \sum_{k=1}^{N} w_k J_{\text{sub},k}$$

where the $J_{\text{sub},k}$'s are partial performance indexes such as predicted queue lengths, total travel times, waiting times, fuel consumption, etc. The weights w_k are not necessarily fixed, but can be changed on-line by the user (i.e., the operator in the traffic control center) depending on the current traffic management policies and other considerations.

Let S_{cm} be the set of possible traffic control measures, such as lane closures, ramp metering, DRIP messages, etc. In general, we can combine several traffic control measures. However, not all combinations are possible or allowed. Therefore, we define a set $\mathcal{S}_{\text{cm}} \subset \mathcal{P}(S_{\text{cm}})$ of allowed combinations of traffic control measures, where $\mathcal{P}(S)$ represents the power set (i.e. the set of all subsets) of a set S.

The kernel of the FDSS-TC is a case base in which several scenarios are stored together with the corresponding partial performance index values. Each scenario or case is characterized by:

- the traffic situation (traffic densities, queue lengths, average speeds, traffic demand, etc.), which we assume to be representable by a vector b_i belonging to a multi-dimensional space $\mathcal{B}$;

- the traffic control measures to be taken based on the current traffic situation, i.e., an element C_i of the set $\mathcal{S}_{\text{cm}}$;

- the predicted effect of C_i on the traffic conditions for traffic situation b_i, i.e., the values of the partial performance indexes $J_{\text{sub},k}(b_i, C_i)$.

Case i is represented in the case base by the tuple

$$(b_i, C_i, J_{\text{sub},1}(b_i, C_i), \ldots, J_{\text{sub},N}(b_i, C_i))$$

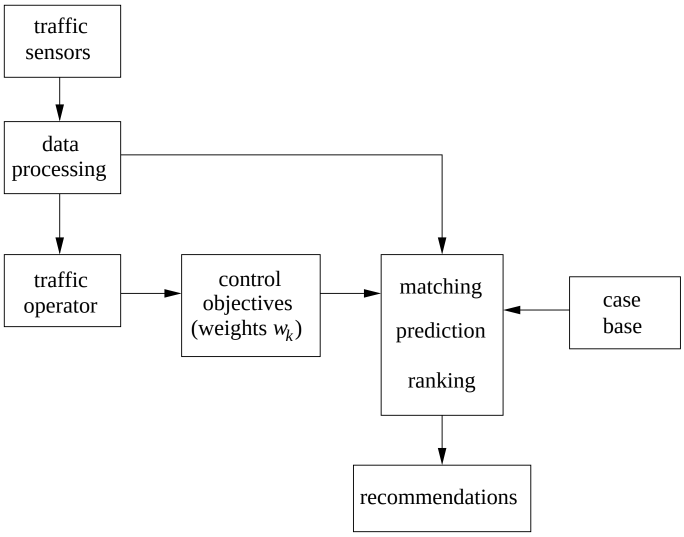


Figure A.3: A detailed block diagram of the FDSS-TC.

. Hence, given the weights w_k, we can compute the total performance $J(b_i, C_i)$ of the set of control measures C_i in traffic situation b_i:

$$J(b_i, C_i) = \sum_{k=1}^{N} w_k J_{\text{sub},k}(b_i, C_i) \quad . \tag{A.1}$$

Remark A.4.1 An important difference between our approach and conventional case-based reasoning is that in conventional case-based reasoning one usually has a fixed solution (for our application this would be a combination of traffic control measures) for each case in the case base. So in the conventional case-based reasoning approach only the traffic situation would be used to characterize a case. However, since we consider an objective function J that is a weighted combination of various performance indicators and since the weights w_k are not fixed but variable, we cannot directly relate an optimal solution to each case (or traffic situation) and therefore we also have to include the control measures and the values of the partial performance indexes $J_{\text{sub},i}$ in the characterization of the cases. □

The core of the fuzzy decision process involves three steps: matching, prediction and ranking, as shown in Figure A.3.

A.4.2 Matching

When presented with a new traffic situation that does not appear in the case base, we have to select the cases for which the traffic situation corresponds best to the given traffic situation. This is done by using a similarity function based on membership functions that describes the degree of similarity between two traffic situations[2]. The similarity between the current traffic vector b_{current} and the traffic situation b_i of case i is characterized by $\mu_i(b_{\text{current}})$ where μ_i is the membership function that corresponds to case i (which will be defined below in Section A.4.5). Note that the range of μ_i is $[0, 1]$. So the similarity ranges from 0 for no similarity at all to 1 for a perfect match.

A.4.3 Prediction

Suppose that we want to predict the performance of the set of control measures C in the current traffic situation. First, we use the similarity measure introduced in the previous section to select the K cases (K is a user-defined integer parameter) for which the traffic situation corresponds best to the current situation and in which the set of control measures $C_i = C$ is present. Assume, without loss of generality, that the K closest cases correspond to the vectors b_1, b_2, $\ldots$, $b_K \in \mathcal{B}$. Note that we have $C_1 = C_2 = \ldots = C_K = C$. Recall that $J(b_i, C)$ expresses the total performance J of the set of control measures C (= C_i) in case i (cf. (A.1)). Then we estimate the performance of C in the current traffic situation as

$$\hat{J}(b_{\text{current}}, C) = \frac{\displaystyle\sum_{i=1}^{K} \mu_i(b_{\text{current}})\, J(b_i, C)}{\displaystyle\sum_{i=1}^{K} \mu_i(b_{\text{current}})} \quad .$$

A.4.4 Ranking

The best M combinations of control measures are then selected and presented to the operator (where M is again a user-defined integer parameter). By choosing M much smaller than the total number of combinations in $\mathcal{S}_{\text{cm}}$ we can significantly reduce the timed needed in the subsequent analysis process by removing from the decision process those combinations for which the performance will probably not be satisfactory.

A.4.5 Membership functions

For each case i we define a membership function μ_i. Recall that this membership function is used to express the degree of similarity between the current traffic situation and

[2] We could also have taken the inverse of the Euclidean distance function to characterize the degree of similarity. However, in that case we have to take into account that the units for different coordinates of the traffic situation may differ. So a rather arbitrary weighting has to be introduced.

the traffic situation in case i. There are several possible membership functions such as trapezoidal, bell-shaped, triangular. We have opted for the last option.

We consider each coordinate of the space $\mathcal{B}$ separately when defining the membership functions. The overall membership function μ_i for case i is then defined as the product of the membership functions $\mu_{i,j}$ for the separate coordinates:

$$\mu_i(x) = \prod_{j=1}^{m_{\mathcal{B}}} \mu_{i,j}(x_j)$$

where $m_{\mathcal{B}}$ is the dimension of the space $\mathcal{B}$.

For coordinates x_j that can only take on discrete values such as the incident status (0 – no incident, 1 – incident), we use singleton membership functions:

$$\mu_{i,j}(x_j) = \begin{cases} 1 & \text{if } x_j = b_{i,j} \\ 0 & \text{otherwise} \end{cases}$$

where $b_{i,j} = (b_i)_j$. Note that by using singleton membership functions for discrete-valued coordinates, the similarity between a situation with an incident and a case with no incident will always be 0, so that a case with no incident will never be used to determine the performance of control measures in an incident situation[3].

For the real-valued coordinates x_j, we use triangular membership functions that can be parameterized using a width factor $\nu \in [0, \infty]$ (as shown in Figure A.4) and that are defined as follows. Assume that there are n cases b_1, b_2, …, b_n in the case base. Let $\Delta_{i,j} = b_{i,j} - b_{i-1,j}$. The membership function $\mu_{i,j}$ for the real-valued coordinate x_j has $b_{i,j}$ as its center point and is defined as

$$\mu_{i,j}(x_j) = \max\Bigg(0, \min\Bigg(\frac{x_j - b_{i,j} + \nu\Delta_{i,j}}{\nu\Delta_{i,j}}, \frac{\nu\Delta_{i+1,j} + b_{i,j} - x_j}{\nu\Delta_{i+1,j}}\Bigg)\Bigg)$$

for $i = 2, \ldots, n-1$. So $\mu_{i,j}(x_j)$ is the piecewise affine curve that connects the points $(-\infty, 0)$, $(b_{i,j} - \nu\Delta_{i,j}, 0)$, $(b_{i,j}, 1)$, $(b_{i,j} + \nu\Delta_{i+1,j}, 0)$ and $(\infty, 0)$. The leftmost and rightmost membership functions $\mu_{1,j}$ and $\mu_{n,j}$ are defined as

$$\mu_{1,j}(x_j) = \max\left(0, \min\left(1, \frac{\nu\Delta_{2,j} + b_{1,j} - x_j}{\nu\Delta_{2,j}}\right)\right)$$

$$\mu_{n,j}(x_j) = \max\left(0, \min\left(1, \frac{x_j - bn, j + \nu\Delta_{n,j}}{\nu\Delta_{n,j}}\right)\right) \quad .$$

[3] Note that this is a major difference from using distance measures to determine the degree of similarity (cf. 2).

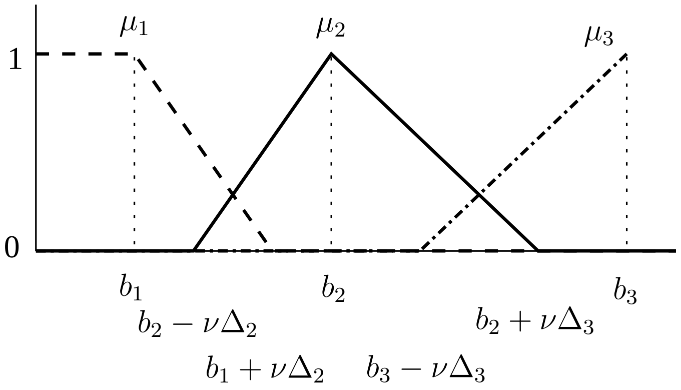


Figure A.4: The membership functions for real-valued coordinates used in the prototype are triangular functions parameterized using a width factor $\nu \in [0, \infty]$.

So $\mu_{1,j}$ is 1 to the left of the first center point coordinate $b_{1,j}$ and $\mu_{n,j}$ is 1 to the right of the last center point coordinate $b_{n,j}$.

The parameter ν defines the width or degree of overlapping between the membership functions. The value $\nu = 0.5$ corresponds to non-overlapping membership functions that still cover the whole coordinate axis, so that in every point that is not halfway between two center points at least one membership function is nonzero. For $\nu \to 0$ all non-border membership functions are 0 everywhere except in their center point where the function value is 1 (note that this corresponds to the singleton membership functions we have used for the discrete-valued coordinates). So the choice $\nu \to 0$ would result in a crisp case base (i.e. without fuzzy interpolation). The choice $\nu \to \infty$ would correspond to membership functions that are identically 1 over the whole input range. If $\nu = 1$ then in any point of the input space that is not a center point and that lies between the first and the last center point, exactly two membership functions are nonzero. The designer of the system can change the value of ν. Also note that due to the modular approach used in the prototype system we can easily replace the triangular membership functions by trapezoidal or bell-shaped membership functions.

A.5 Prototype of the FDSS-TC

In order to *assess* the *technical feasibility* of the approach proposed above, we have created a small prototype of the FDSS-TC for a simple set-up consisting of a highway that at one point splits into two branches — a longer one of 13 km and a shorter one of 11 km, — which join each other again at the end, as shown in Figure A.5. Both branches have two lanes for each direction. This network is part of the larger peri-urban network around the city of Amsterdam in the Netherlands. The longer branch is the A22 highway that includes the Velser tunnel; the shorter branch is part of the A9 highway and includes

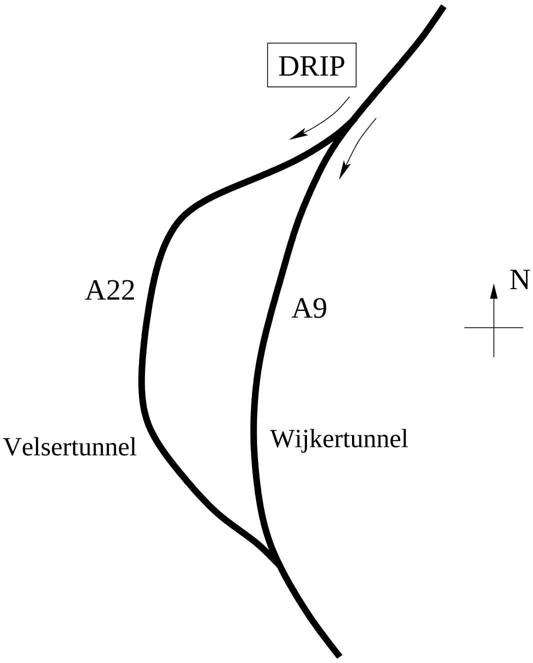


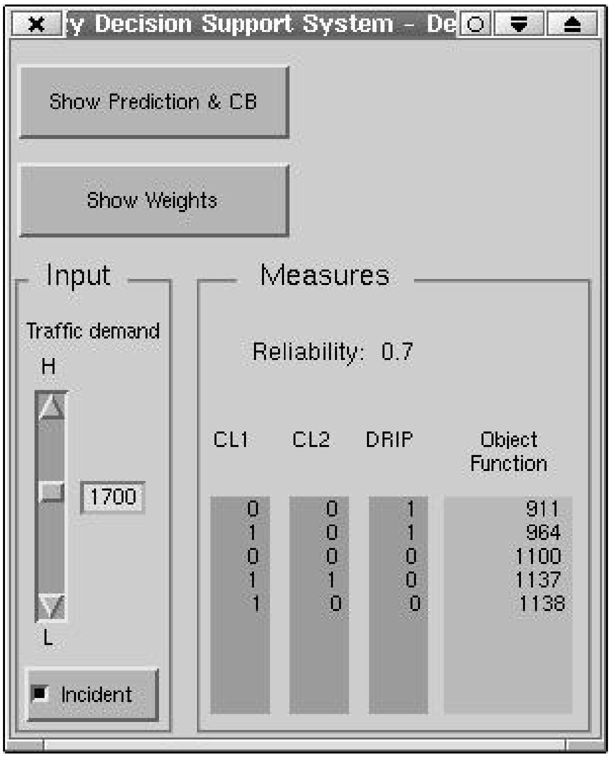
Figure A.5: The road configuration considered in the prototype system.

the Wijker tunnel. The A22 is mostly used for traffic having local origin or destination whereas the A9 is mostly used for long distance traffic. We only consider traffic going from the north to the south. The two alternative routes that can be followed by the drivers are indicated by the arrows. Near the point where the highway splits there is a DRIP that can display queue information. The prototype of the FDSS-TC has been implemented in the mathematical software package Matlab (which includes a programming language and the possibility to create graphical user interfaces).

Since at this stage of the project we only wanted to assess the technical feasibility of the system, we have only considered a limited number of inputs, control measures and cases. In practical situations, the number of inputs for the FDSS-TC and the number of control measures and cases will of course be much larger; this will be a topic for future work. Note, however, that since our system has been programmed in modular way, the number of inputs, possible control measures and cases can be extended very easily (see also Section A.6 for a method to deal with scalability problems).

There are two inputs for the decision support system: traffic demand and occurrence of incidents on the A9; and three possible control measures:

- c_1: closure of lane 1 on the A9 (upstream of the incident),
- c_2: closure of lane 2 on the A9 (upstream of the incident),
- c_3: display a DRIP message.



Traffic demand: 1700

Reliability: 0.7

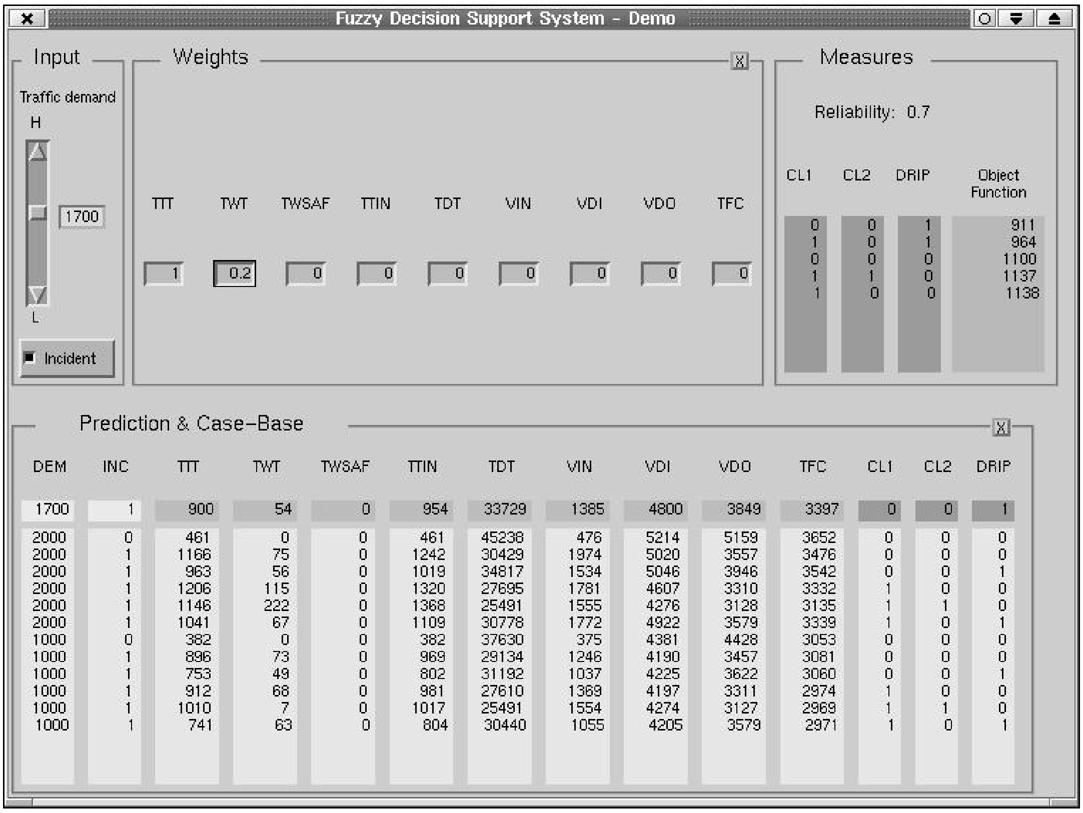
CL1	CL2	DRIP	Object Function
0	0	1	911
1	0	1	964
0	0	0	1100
1	1	0	1137
1	0	0	1138

Figure A.6: A screenshot of the prototype of the decision support system in the operator view (with control measures CL1: close lane 1, CL2: close lane 2, and DRIP: display a DRIP message).

The set $\mathcal{S}_{\mathrm{cm}}$ of allowed control measures equals $\{\emptyset, \{c_1\}, \{c_3\}, \{c_1, c_2\}, \{c_1, c_3\}\}$. The case base has been constructed using ten METANET [109] simulations. Due to the small number of inputs and cases we have selected the value $K = 2$ for the number of cases among which the fuzzy interpolation takes place. For the width factor ν of the membership functions we have selected the default value $\nu = 1$.

Figures A.6 and A.7 show some screenshots of the system prototype. The interface window that is presented to the operators has two modes: operator or basic mode, and expert or full mode. In the basic mode, shown in Figure A.6, the operator enters the parameters that describe the current traffic situation on the left; on the right she will then see a ranked list of the various possible combinations of control measures, and an indication of the reliability, i.e., the maximal degree of similarity between the current traffic situation and that of the cases in the fuzzy case base. The most promising combination(s) of control measures can then be examined in more detail (e.g., by microscopic or macroscopic traffic simulation). In the Weights subscreen of the full mode view, shown in Figure A.7, the user can specify the weights w_k for the various subcomponents[4] $J_{\mathrm{sub},k}$ of the objec-

[4] The partial performance measures have been extracted from the METANET simulations that have been



Fuzzy Decision Support System - Demo

Input

Traffic demand

H

1700

L

■ Incident

Weights

TTT	TWT	TWSAF	TTIN	TDT	VIN	VDI	VDO	TFC
1	0.2	0	0	0	0	0	0	0

Measures

Reliability: 0.7

CL1	CL2	DRIP	Object Function
0	0	1	911
1	0	1	964
0	0	0	1100
1	1	0	1137
1	0	0	1138

Prediction & Case-Base

DEM	INC	TTT	TWT	TWSAF	TTIN	TDT	VIN	VDI	VDO	TFC	CL1	CL2	DRIP
1700	1	900	54	0	954	33729	1385	4800	3849	3397	0	0	1
2000	0	461	0	0	461	45238	476	5214	5159	3652	0	0	0
2000	1	1166	75	0	1242	30429	1974	5020	3557	3476	0	0	0
2000	1	963	56	0	1019	34817	1534	5046	3946	3542	0	0	1
2000	1	1206	115	0	1320	27695	1781	4607	3310	3332	1	0	0
2000	1	1146	222	0	1368	25491	1555	4276	3128	3135	1	1	0
2000	1	1041	67	0	1109	30778	1772	4922	3579	3339	1	0	1
1000	0	382	0	0	382	37630	375	4381	4428	3053	0	0	0
1000	1	896	73	0	969	29134	1246	4190	3457	3081	0	0	0
1000	1	753	49	0	802	31192	1037	4225	3622	3060	0	0	1
1000	1	912	68	0	981	27610	1369	4197	3311	2974	1	0	0
1000	1	1010	7	0	1017	25491	1554	4274	3127	2969	1	1	0
1000	1	741	63	0	804	30440	1055	4205	3579	2971	1	0	1

Figure A.7: A screenshot of the prototype of the decision support system in the expert view.

tive function such as the total travel time (TTT), total waiting time (TWT), total waiting store-and-forward (TWSAF), total time in net (TTIN), total distance traveled (TDT), vehicles in net (VIN), vehicles driven in (VDI), vehicles driven out (VDO), and total fuel consumption (TFC). In the Prediction & Case-Base subscreen the values for each subcomponent $J_{\mathrm{sub},k}$ of the objective function are then displayed for the current inputs and for each scenario in the case base. In this way the effects of the choice of the weights and the effects of the various control measures can be examined in more detail. However, this level of detail is usually not needed for daily operation. That is why we have chosen for a system with two modes (operator mode and expert mode).

used to generate the cases for our simple prototype system. Due to the modular approach we have used, other partial performance measures can easily be included.

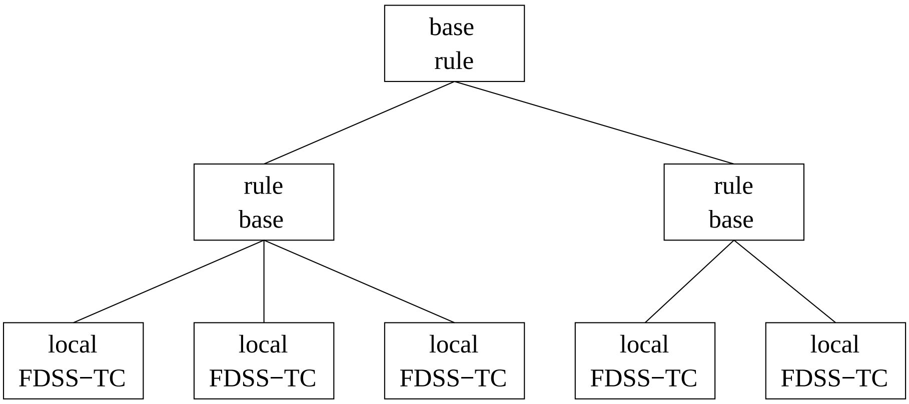


Figure A.8: A multi-level decision support architecture.

A.6 Extensions

In the prototype FDSS-TC of the previous section we have considered a rather simple configuration with a limited number of inputs, control measures and cases since our main intention was to assess the technical feasibility of such a system. In our future work we will consider more complex configurations. However, if the number of inputs and possible control measures becomes too large, it will not be tractable anymore to construct a case base that contains enough cases to cover the entire input/control space in an adequate way [98]. In that case we propose to construct a multi-level decision support system with at the lowest level several local FDSS-TCs (each covering a smaller subregion of the overall configuration or covering a small subset of possible traffic situations, and thus having a limited number of inputs, cases and control measures) and at the higher levels rule bases which select the local FDSS-TC to be used, as shown in Figure A.8. In this way we can address the scalability problems.

The current knowledge base of the FDSS-TC is mainly based on simulations. Once the system operates in a real traffic control center, we can include actual situations and the effects of control measures that have actually been applied to the traffic system in our case base. In that way we get an adaptive system that learns during operation. Such a system is described in [145]. We then get a process that consists of the cyclic application of the following steps:

1. Retrieve the most similar cases (in our case the similarity can be determined using the membership function as has been explained above).

2. Use these cases to solve the problem (in our case: to generate the ranking of the combinations of control measures using fuzzy interpolation).

3. Revise the proposed solution (in our case: see how the traffic system reacts to the proposed solution, i.e., determine or measure its performance).

4. Retain the parts of this experiences to be used for future application.

Another issue is that the current FDSS-TC system is in fact a fuzzy interpolator. For operator acceptance it may be important to provide suggestions for traffic control scenarios of which the reasoning path can be tracked. To achieve this, operator knowledge should be incorporated in an expert system with reasoning capabilities. Such an approach is followed in [113, 70].

A.7 Conclusions and further research

We have presented a fuzzy decision support system (FDSS-TC) for traffic control centers. This system is part of a larger traffic decision support system that assists operators of traffic control centers when selecting the most appropriate traffic control measures to efficiently manage non-recurrent congestion. The FDSS-TC uses a case base and fuzzy interpolation to generate a ranked list of combinations of control measures and their estimated performance. Since the scenarios in the case base are generated by METANET, the quality of the ranking basically depends on the quality of the simulations. The predictions made by the case-based reasoning system can be made more precise by adding new cases. An important feature of the system is that the performance function is not fixed but consists of a weighted combination of several partial performance measures. In addition, the weights of this combination can be changed on-line depending on the current traffic management policy and on other considerations. Since the case base can be generated off-line, the FDSS-TC reduces the time that is needed to determine the optimal traffic control for a given situation by limiting the number of combinations of control measures for which on-line traffic simulations should be performed in the traffic control center. At a later stage the system can be extended with a fuzzy module that incorporates expert knowledge, and with an adaptive learning module.

Currently we have demonstrated the technical feasibility of the system. In the next stage of the project we will examine the performance of the system (for a larger network than the one described in this appendix), see how the parameters of the system have to tuned to improve the performance, and compare this performance with other traffic control strategies using both simulations and field experiments. The quality of the FDSS-TC depends on the quality of the simulations that generated the cases. In this context an important question is — assuming that the quality of the simulation is good — how many cases do we need for a good performance. Another interesting question is how many inputs are needed in a larger traffic network to be able to make adequate decisions. In our network there was only one input link that characterized the traffic state, but in a larger network not only the demands on the input are important, but also the states (speed, density) on the internal links. Moreover, we have not considered the dynamic aspects of the system. The the time-of-day and day-of-week can carry important information about the expected traffic demands. This information could also be used to make better decisions. If the number of inputs and control measures increases, the number of cases also has to increase, which might lead to tractability problems. These problems can be addressed by using a multi-level decision support architecture. The design of such an architecture will also be a topic for future research. Finally, note that this project has had a follow-up. We refer the interested reader to [79, 32, 78].

Samenvatting

Model predictive control voor het integreren van verkeersbeheersingsmaatregelen

Door het gebruik van dynamische verkeersbeheersingsmaatregelen kan de beschikbare wegcapaciteit beter benut worden. Deze maatregelen moeten echter zodanig aangestuurd worden, dat – indien er interactie is tussen de maatregelen – ze allen hetzelfde gemeenschappelijke doel dienen. Dit impliceert een regeling waarin alle maatregelen onderling worden gecoördineerd.

Door de toenemende hoeveelheid verkeer en het groeiend aantal files is de interactie tussen de maatregelen zodanig toegenomen dat lokale regelingen vaak niet meer volstaan. Een voorbeeld van dergelijke interactie is een verkeersstroom die door twee files heen gaat, zoals geïllustreerd in figuur A.9. Als door toeritdosering file 1 opgelost wordt, zullen de weggebruikers uit die file eerder in file 2 terechtkomen, wat het oplossen van die file juist bemoeilijkt. In het gecoördineerde geval zou eerst file 2 en daarna pas file 1 worden opgelost, wat tot een betere doorstroming zou leiden.

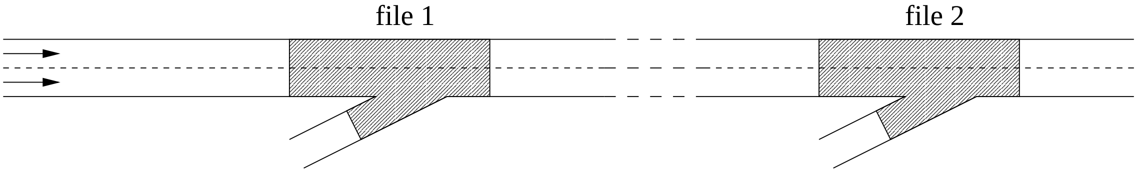
Voor een gecoördineerd verkeersregelsysteem is de selectie van het optimale regelscenario een complex probleem. Ten eerste, kan het aantal beschikbare verkeersmaatregelen groot zijn, waardoor het aantal keuzemogelijkheden exponentieel groeit. Ten tweede, is er in het algemeen een grote hoeveelheid historische en gemeten data beschikbaar op grond waarvan het regelscenario geselecteerd moet worden.

Vanwege deze redenen keuze voor een geautomatiseerd, gecentraliseerd, netwerkgeoriënteerd regelsysteem voor de hand liggend. Een dergelijk systeem resulteert in een betere voorspelling van de prestatie van het verkeersnetwerk ten gevolge van het gekozen regelscenario, en een betere benutting van de beschikbare data. Als gevolg hiervan kan de optimale regelscenario nauwkeuriger bepaald worden.

Een dergelijk netwerk-geörienteerd regelsysteem heeft naast de *integratie* van de maatregelen een tweede belangrijk aspect: de *voorspelling*. Het voordeel van voorspellen is tweevoudig:

- ten eerste kunnen sommige problemen voorkomen worden, door die voortijdig te signaleren en adequate maatregelen te treffen,

- ten tweede opent voorspelling de mogelijkheid om maatregelen te treffen die nadelig lijken op het moment van toepassing (bijvoorbeeld: het verkeer afremmen) maar op langere termijn de afwikkeling van het verkeer aanzienlijk verbeteren.



Figuur A.9: Een voorbeeld van verkeer dat door twee files heen moet. De volgorde van het oplossen van de files is zeer belangrijk. Daarom is er coördinatie nodig tussen de verkeersbeheersingsmaatregelen.

Een belangrijk aspect van het verkeersregelprobleem is de formulering van het regeldoel. De sleutelbegrippen efficiëntie, veiligheid, betrouwbaarheid, brandstofverbruik en milieubelasting spelen bij de formulering van het regeldoel een belangrijke rol. Daarbij zal vaak niet één enkel van deze begrippen het doel zijn maar een combinatie ervan, waarbij de afweging tussen de mogelijk conflicterende doelen een beleidsmatige keuze is, gebaseerd op door de politiek gestelde prioriteiten. Verder kunnen er ook *beperkingen* aanwezig zijn in het verkeersregelprobleem, zoals een maximaal toegestane lengte van de wachtrij op een toerit, of combinaties van snelheidslimieten die vanwege veiligheidsredenen niet toegestaan zijn. Een verkeersregelsysteem moet dus ook met zulke, door de het beleid en de politiek gekozen regeldoelen en gegeven beperkingen, kunnen omgaan.

Uitgaande van de bovenstaande beschouwingen, definiëren we het verkeersregelprobleem als volgt:

Gegeven

- de netwerkstructuur,
- de huidige toestand van het verkeer,
- de voorspellingen van de verkeerssituatie aan de randen van het netwerk, zoals de verkeersvraag en inkomende schokgolven,
- de beschikbare verkeersbeheersingsmaatregelen,
- de beperkingen,
- het beleidsmatig gestelde regeldoel,

vind het regelscenario dat het regeldoel optimaal bereikt.

Het centrale thema in dit proefschrift is het oplossen van dit verkeersregelprobleem met behulp van model predictive control.

Een onderdeel van het oplossen van dit probleem is het voorspellen van de effecten van een gegeven regelscenario met behulp van een verkeersstroommodel. Daarom geven we in hoofdstuk 3 een kort overzicht van de bestaande verkeersstroommodellen. Daarnaast geven we een gedetailleerde beschrijving van het macroscopische verkeersstroommodel METANET, dat we in latere hoofdstukken gebruiken voor de simulatie van een aantal case studies. Verder voegen we in hoofdstuk 3 een aantal uitbreidingen aan het METANET-model toe, om tot een meer realistische modellering te komen van schokgolven, dynamische snelheidslimieten, hoofdrijbaandosering, en de randvoorwaarden van een verkeersscenario.

In hoofdstuk 4 behandelen we de formele beschrijving van model predictive control, en de vraagstukken gerelateerd aan de toepassing van deze techniek op verkeersregelproblemen. Model predictive control is zeer geschikt voor verkeersregelproblemen, omdat het alle elementen bevat die nodig zijn voor een succesvolle verkeersregeling: integratie van maatregelen, voorspelling, een door de gebruiker gedefinieerde doelfunctie, en het omgaan met beperkingen. Een mogelijk nadeel van deze methode is echter de hoge rekencomplexiteit, die een probleem kan worden bij het real-time regelen van grotere netwerken.

Niet *alle* verkeersproblemen kunnen opgelost worden door verkeersregelingen. Daarom is het belangrijk om te beschrijven onder welke voorwaarden verkeersregelingen effectief zullen zijn. In hoofdstuk 5 geven we een eerste aanzet voor de formulering van zulke voorwaarden. Uitgangspunt in dat hoofdstuk is dat het voornaamste regeldoel het minimaliseren van de totale reistijd van alle verkeersdeelnemers is. Vanuit dit oogpunt heeft regelen alleen zin als de totale reistijd niet optimaal is, ofwel indien er

- vanwege een file een capaciteitsval[5] optreedt,
- vanwege een file voertuigstromen worden geblokkeerd die niet de werkelijke bottleneck hoeven te passeren,
- andere capaciteit-reducerende gebeurtenissen optreden, zoals incidenten of wegwerkzaamheden, waardoor de routekeuze van de weggebruikers niet optimaal is.

In hoofdstuk 5 beschouwen we alleen de eerste twee situaties. Verdere voorwaarde voor een succesvolle regeling is dat het de instroom in een file moet kunnen beperken tot onder het niveau van de uitstroom van de file, om de file op te kunnen lossen. Vanwege de capaciteitsval is dit niveau significant lager dan de capaciteit van de bottleneck onder free-flow omstandigheden. Daarom kan voor sommige combinaties van verkeersvraag en verkeersbeheersingsmaatregelen de instroom niet voldoende beperkt worden, en is het noodzakelijk om voorafgaand aan de keuze en de toepassing van de maatregelen het typische verkeersaanbod, de capaciteitsval en de maximale intensiteitsbeperking van de verkeersmaatregelen op deze voorwaarde te toetsen. In hoofdstuk 5 bespreken we deze voorwaarde voor twee vaak voorkomende verkeerssituaties: schokgolven op snelwegen, en files op een snelweg vlakbij toeritten. In het geval van schokgolven op snelwegen zal de instroom doorgaans voldoende beperkt kunnen worden door dynamische snelheidslimieten om de schokgolven te elimineren. Echter, in het geval van files op een snelweg

[5] Capaciteitsval is het verschijnsel dat de uitstroom uit een file aanzienlijk lager is dan de capaciteit van de weg onder *free flow* omstandigheden.

bij toeritten zal de toeritdosering echter maar voor een beperkte combinatie van hoofdstroom en toeritstroom effectief zijn, omdat de stroom op de hoofdrijbaan niet beperkt kan worden.

In hoofdstuk 6 behandelen we twee case studies waarin we de mogelijkheden van dynamische snelheidslimieten in verkeersregelingen bestuderen. De eerste case beschrijft het scenario waar de beperking van de hoeveelheid verkeer op de toerit alleen niet toereikend is om de file op te lossen. De extra beperking van de hoofdstroom door dynamische snelheidslimieten kan doorslaggevend zijn om de file bij de toerit op te kunnen lossen. De tweede case is een toepassing van dynamische snelheidslimieten voor het elimineren van schokgolven die vaak ontstaan bij bottlenecks. Deze schokgolven zijn korte files die op de snelweg stroomopwaarts propageren en lang kunnen blijven bestaan. Verder is bij deze schokgolven de reistijd langer en de kans op ongelukken groter, wat ongewenst is. In hoofdstuk 6 tonen we aan deze schokgolven geëlimineerd kunnen worden door dynamische snelheidslimieten, zonder dat er nieuwe schokgolven worden opgewekt.

In netwerken waar routekeuze mogelijk is, is routegeleiding een maatregel die de efficiëntie van het verkeersnetwerk kan vergroten. In deze context heeft routegeleiding twee functies:

- *informeren:* de bestuurders zo nauwkeurig mogelijk informeren over de verwachtte reistijden op de alternatieve routes,
- *regelen:* het verkeer optimaal over de alternatieve routes verdelen.

Er is echter er een conflict tussen deze twee functies, omdat nauwkeurig informeren tot een verdeling van het verkeer kan leiden welke niet optimaal is. Of andersom geformuleerd, voor een optimale verdeling zouden mogelijk incorrecte reistijden getoond moeten worden op de dynamische routegeleidingspanelen. Een bijkomend probleem is dat het vaak onmogelijk is om 100 % correcte reistijdvoorspellingen te geven vanwege de onvoorspelbare verstoringen in het verkeer, en vanwege de onvoorspelbare invloed van andere verkeersmaatregelen, zoals toeritdosering. Om deze problemen gezamenlijk aan te pakken wordt er in hoofdstuk 7 een regelprobleem geformuleerd waarin een afweging wordt gemaakt tussen een de precisie van de reistijdvoorspelling en het optimaal functioneren van het verkeersnetwerk. Het resultaat hiervan is dat de regeling enerzijds het netwerk optimaliseert met behulp van routegeleiding en toeritdosering, en anderzijds de voorspellingsfout beperkt houdt.

Veel verkeersproblemen ontstaan bij de overgang tussen snelwegen en stedelijke gebieden, zoals sluipverkeer, lange wachtrijen bij toeritten die stedelijke kruispunten blokkeren, of snelwegverkeer dat de snelweg niet af kan, vanwege een te lage capaciteit van het stedelijk netwerk. Om dit soort problemen op te lossen is er coördinatie nodig tussen de maatregelen op de snelweg en de maatregelen in de stad. In hoofdstuk 8 wordt een gecombineerd model ontwikkeld voor zowel stedelijke als snelwegen, dat geschikt is als voorspellingsmodel in een model predictive control regelaar. De effectiviteit van de regeling met het gecombineerde model wordt in hoofdstuk 8 gedemonstreerd voor een aantal spits- en filescenario's.

Samenvattend kunnen we zeggen dat model predictive control een zeer succesvolle aanpak is voor uiteenlopende verkeersregelproblemen, omdat alle essentiële aspecten van verkeersregelproblemen in dit framework te formuleren zijn. Hoewel we in dit proefschrift geen praktijkproeven of gecalibreerde simulatiestudies presenteren, zijn de resultaten van de synthetische studies veelbelovend. Daarom wordt verder onderzoek richting de praktische toepassing van model predictive control ten zeerste aanbevolen.

Summary

By using dynamic traffic control measures the available road capacity can be utilized better. However, these measures need to be controlled such that – if there is interaction between the measures – they serve all the same common goal. This implies that the control measures need to be coordinated.

By the increasing traffic volumes and the increasing number of traffic jams, the interaction between the control measures has become stronger, and local controllers are often not satisfactory anymore. An example of such an interaction is a traffic stream that passes two traffic jams, as illustrated in Figure A.10. If jam 1 is resolved by ramp metering, the vehicles will arrive earlier in jam 2, which will make the resolution of that jam more difficult. In case of coordination control, jam 2 would be resolved before jam 1, which would result in a higher traffic flow.

It is a complex problem to select the optimal control scenario for a coordinated traffic control system. First, the number of the available traffic control measures can be large, resulting in an exponential growth of number of choices. Second, there is a large amount of historical and measured data available, which should be a basis for the selection of the control scenario.

For these reasons, the choice for an automatic, centralized, network-oriented control system are clear is easily made. Such a system results in a better prediction of the performance of the traffic network as function of the selected control scenario, and a better utilization of the available data. Consequently the optimal control scenario can be determined better.

Such a control system has besides the *integration* of control measures a second important aspect: the *prediction*. The advantage of prediction is twofold:

- first, some problems can be prevented by recognizing them in an early stage, and taking appropriate measures,
- second, prediction opens the possibility to apply measures that may seem adverse at the moment of application, but that improve the traffic flow considerably on the long term.

An important aspect of the traffic control problem is the formulation of the control objective. Keywords, such as efficiency, safety, reliability, fuel consumption, and environmental effects play an important role in the formulation of the control objective. Usually not a single one of these keywords is the objective, but a combination of them, where the trade-off between conflicting objectives is a matter of policy and politics. Furthermore,

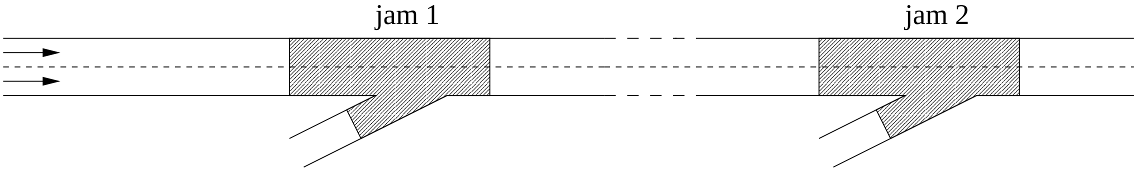


Figure A.10: An example of traffic that passes two jams. The order of the resolution of the jams is very important. Therefore, coordination is necessary between the control measures.

there can be constraints present in a traffic control problem, such as a maximal queue length at a metered on-ramp, or combinations of dynamic speed limits that are forbidden because of safety considerations. Therefore, a traffic control system should also be able to accept objectives defined by policy and the politics, and to handle given constraints.

Given the considerations above, we define the traffic control problem as follows:

> Given
>
> - the network structure,
> - the current state of the traffic,
> - the predictions of the traffic states at the edges of the network, such as traffic demands and incoming shock waves,
> - the available traffic control measures,
> - the constraints,
> - the policy-defined control objective,
>
> find the control scenario that optimally realizes the control objective.

The central topic in this thesis is the solution of this traffic control problem with model predictive control.

A part of the solution of this problem is the prediction of the effects of a given control scenario with the use of a traffic flow model. Therefore, we give in Chapter 3 a short overview of the existing traffic flow models. In addition we give an extended description of the macroscopic traffic flow model METANET, which we use for the simulation of a number of case studies in subsequent chapters. Furthermore, we add a number of extensions to the METANET model to achieve a more realistic modeling of shock waves, dynamic speed limits, main-stream metering, and the boundary conditions of a traffic scenario.

In Chapter 4 we discuss the formal description of model predictive control, and the issues related to the application of this technique to traffic control problems. Model predictive control is very suitable for traffic control problems, because it contains all necessary elements for successful traffic control: integration of control measures, prediction, a user-defined objective function, and constraint handling. A possible disadvantage of this method is the high computational complexity, that could become a problem for the real-time control of larger traffic networks.

Not all traffic problems can be solved by traffic control. Therefore it is important to describe under which conditions traffic control will be effective. In Chapter 5 we present the first steps towards the formulation of these conditions. The basic assumption in that chapter is that the main control objective is the minimization of the total travel time of all vehicles. From this perspective, traffic control is only useful when the total travel time is not optimal, which occurs when

- there is a capacity drop[6] caused by a jam,
- a traffic jam blocks traffic streams that do not need to pass the real bottleneck,
- other capacity-reducing events occur, such as incidents or road works, that cause the route choice of the drivers to be sub-optimal.

In Chapter 5 we only consider the first two cases. A further condition for a successful controller is that it should be able to limit the inflow to a traffic jam to a level below the outflow of the jam in order to be able to resolve the jam. Because of the capacity drop this level is significantly lower than the capacity of the bottleneck under free flow conditions. As a consequence, for some combinations of traffic demands and control measures the inflow cannot be limited sufficiently. Therefore, it is necessary to check the typical traffic demand, the capacity drop, and the maximal flow limitation due to traffic control, and verify this condition before the selection and application of the traffic control measures. In Chapter 5 we discuss this condition for two frequently occurring traffic situations: shock waves on freeways, and freeway jams at on-ramps. In the case of shock waves on freeways it will be typically possible to limit the inflow sufficiently by dynamic speed limits to eliminate the shock waves. However, in the case of freeway jams at on-ramps, ramp metering will be effective only for a limited combination of main-stream demand and ramp demand, because the main-stream flow cannot be limited.

In Chapter 6 we discuss two case studies where the possibilities of dynamic speed limits are examined. The first case the scenario where the limitation of the on-ramp flow alone is insufficient to resolve the jam. The extra limitation of the main-stream flow due to dynamic speed limits can be decisive to be able to resolve the jam at the on-ramp. The second case is the application of dynamic speed limits to eliminate shock waves, that frequently occur at bottlenecks. These shock waves are short jams that propagate upstream and can remain existent for a long time. In addition, these shock waves increase travel time and increase the probability of accidents, which is undesired. In Chapter 6 we

[6] The capacity drop is the phenomenon that the outflow of a traffic jam is considerably lower than the capacity of the freeway under free flow conditions.

demonstrate that these shock waves can be eliminated by dynamic speed limits, without creating new shock waves.

In networks where route choice is possible, route guidance is a traffic control measure that can improve the efficiency of the traffic network. In this context route guidance has two functions:

- *informing:* inform the drivers as accurate as possible about the expected travel times over the alternative routes,

- *controlling:* distribute the traffic optimally over the alternative routes.

However, there is a conflict between these two functions, because informing accurately may result in a traffic distribution that is not optimal. Or saying it the other way around: to achieve an optimal distribution, possibly incorrect travel times need to be shown on the dynamic route information panels. An additional problem is that it is often impossible to provide 100 % correct travel time predictions due to unpredictable disturbances in the traffic, and due to the unpredictable influence of other traffic control measures, such as ramp metering. To solve these problems altogether in Chapter 7 a traffic control problem is formulated that makes a trade-off between the accuracy of the travel time prediction and the optimality of the network. This results in a controller that optimizes the network performance with the use of route guidance and ramp metering, on the one hand, and keeps the travel time prediction error limited, on the other hand.

Traffic problems frequently occur at the interface between freeways and urban areas. Examples of such problems are: rat running, long queues at on-ramps block urban intersections, and congested freeway traffic due to insufficient capacity of the urban network. To solve these kind of problems coordination is necessary between the freeway and urban traffic control measures. In Chapter 8 a mixed urban-freeway model is developed, that is suitable as a prediction model in a model predictive control-based controller. The effectivity of the control with the combined model is demonstrated in Chapter 8 for a number of peak hour and traffic jam scenarios.

In conclusion we can say that model predictive control is a successful approach to a range of traffic problems, since all essential aspects of traffic control problems can be formulated in this framework. Although we do not present real-life experiments or calibrated studies in this thesis, the results of the synthetic studies are very promising. Therefore, further research towards the practical application of model predictive control is strongly recommended.

Bibliography

[1] A. Alessandri, A. Di Febbraro, A. Ferrara, and E. Punta, “Optimal control of freeways via speed signalling and ramp metering,” *Control Engineering Practice*, vol. 6, pp. 771–780, 1998.

[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*, vol. 48, no. 6, pp. 2042–2052, Nov. 1999.

[3] S. Algers, E. Bernauer, M. Boero, L. Breheret, C. Di Taranto, M. Dougherty, K. Fox, and J.-F. Gabard, “SMARTEST - Final report for publication,” Tech. rep., ITS, University of Leeds, http://www.its.leeds.ac.uk/smartest, 2000.

[4] F. Allgöwer, T.A. Badgwell, J.S. Qin, J.B. Rawlings, and S.J. Wright, “Nonlinear predictive control and moving horizon estimation – An introductory overview,” in *Advances in Control: Highlights of ECC ’99* (P.M. Frank, ed.), pp. 391–449, Springer, 1999.

[5] M. André and U. Hammarström, “Driving speeds in europe for pollutant emissions estimation,” *Transportation Research Part D*, vol. 5, no. 5, pp. 321–335, 2000.

[6] Arti Gupta Ashkan Karimi, “Incident management system for santa monica smart corridor,” in *Compendium of technical papers; 63rd Annual Meeting of the Institute of Transportation Engineers*, pp. 180–185, 1993.

[7] J. Barceló and J. L. Ferrer, “AIMSUN2: Advanced interactive microscopic simulator for urban networks, user’s manual,” Tech. rep., Universitat Politècnica de Catalunya, 1997.

[8] R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg, “Metastable states in cellular automata for traffic flow,” *The European Physical Journal B*, vol. 5, no. 3, pp. 793–800, 1998.

[9] T. Basar and G.J. Olsder, *Dynamic Noncooperative Game Theory*. Academic Press, 1995.

[10] T. Bellemans, *Traffic Control on Motorways*. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, May 2003.

[11] T. Bellemans, B. De Schutter, and B. De Moor, "Model predictive control with repeated model fitting for ramp metering," in *Proceedings of the the IEEE 5th International Conference on Intelligent Transportation Systems (ITSC 2002)*, Singapore, pp. 236–241, Sept. 2002. (CDROM).

[12] P.T. Boggs and J.W. Tolle, "Sequential Quadratic Programming," *Acta Numerica*, vol. 4, pp. 1–51, 1995.

[13] H. Botma, "State-of-the-art report "Traffic Flow Models"," Tech. rep. R-78-40, SWOV, 1978.

[14] D. Branston, "Models of single lane time headway distributions," *Transportation Science*, vol. 10, pp. 125–148, 1976.

[15] P. Breton, A. Hegyi, B. De Schutter, and H. Hellendoorn, "Shock wave elimination/reduction by optimal coordination of variable speed limits," in *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems*, Singapore, Sept. 3–6 2002.

[16] D.J. Buckley, "A semi-poisson model of traffic flow," *Transportation Science*, vol. 2, no. 2, pp. 107–132, 1968.

[17] I. Veling (Traffic Test BV), "Project evaluatie effecten verkeersbeheersingsmaatregelen eva (Project evaluation effects of traffic control measures eva)," Tech. rep. TT00-34/lba3210, AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management, Rotterdam, July 2000.

[18] E.F. Camacho and C. Bordons, *Model Predictive Control in the Process Industry*. Berlin, Germany: Springer-Verlag, 1995.

[19] Cambridge Systematics, Inc., *Twin Cities Ramp Meter Evaluation – Final Report*. Cambridge Systematics, Inc., Oakland, California, Feb. 2001. Prepared for the Minnesota Department of Transportation.

[20] M. J. Cassidy and R. L. Bertini, "Some traffic features at freeway bottlenecks," *Transportation Research Part B*, vol. B33, pp. 25 – 42, 1999.

[21] C.-C. Chien, Y. Zhang, and P. A. Ioannou, "Traffic density control for automated highway systems," *Automatica*, vol. 33, no. 7, pp. 1273–1285, 1997.

[22] C. C. Chien, Y. Zhang, A. Stotsky, and P. Ioannou, "Roadway traffic controller design for automated highway systems," in *Proceedings of the 33rdIEEE Conference on Decision and Control*, Lake Buena Vista, pp. 2425–2430, IEEE, Dec. 1994.

[23] J.S. Cramer, *An Introduction to the Logit Model for Economists*. London, UK: Timberlake Consultants Press, 2001.

[24] M. Cremer, *Der Verkehrsfluss auf Schnellstrassen (Traffic flow on freeways)*, vol. 3 of *Fachberichte Messen, Steuern, Regeln*. Berlin: Springer-Verlag, 1979. In German.

[25] M. Cremer, "On the calculation of individual travel times by macroscopic models," in *Proceedings of the 1995 Vehicle Navigation and Information Systems Conference*, Washington, D.C., pp. 187–193, 1995.

[26] J. Cuena, J. Hernández, and M. Molina, "Knowledge-based models for adaptive traffic management systems," *Transportation Research Part C*, vol. 3, no. 5, pp. 311–337, 1995.

[27] C.F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B*, vol. 28B, no. 4, pp. 269–287, 1994.

[28] C.F. Daganzo, "The cell transmission model, part II: Network traffic," *Transportation Research Part B*, vol. 29B, no. 2, pp. 79–93, 1995.

[29] C.F. Daganzo, "A finite difference approximation of the kinematic wave model of traffic flow," *Transportation Research Part B*, vol. 29B, no. 4, p. 261276, 1995.

[30] C.F. Daganzo, "Requiem for second-order fluid approximations of traffic flow," *Transportation Research Part B*, vol. 29B, no. 4, pp. 277–286, 1995.

[31] L. Davis, ed., *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.

[32] B. De Schutter, S.P. Hoogendoorn, H. Schuurman, and S. Stramigioli, "A multiagent case-based traffic control scenario evaluation system," Tech. rep. CSE02-018, Control Systems Engineering, Fac. of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands, Nov. 2002. Accepted for the 6th IEEE International Conference on Intelligent Transportation Systems (ITSC'03), Shanghai, China, Oct. 2003.

[33] DHV Milieu en Infrastructuur BV, "Eindevaliatie SlimRijden– verkeerskundigeen weggebruikersevaluatie," Tech. rep., AVV Traffic Research Centre, Dutch Ministry of Transport, Public Works and Water Management, 2003.

[34] A. Di Febbraro, T. Parisini, S. Sacone, and R. Zoppoli, “Neural approximations for feedback optimal control of freeway systems,” *IEEE Transactions on Vehicular Technology*, vol. 50, no. 1, pp. 302–312, Jan. 2001.

[35] C. Diakaki, M. Papageorgiou, and T. McLean, “Integrated traffic-responsive urban corridor control strategy in Glasgow, Scotland,” *Transportation Research Record*, vol. 1727, 2000.

[36] R.W. Eglese, “Simulated annealing: A tool for operations research,” *European Journal of Operational Research*, vol. 46, pp. 271–281, 1990.

[37] N. Elloumi, H. Haj-Salem, and M. Papageorgiou, “A motorway-to-motorway contol – Results of a simulation study,” in *Transportation Systems 1997 – A Proceedings volume from the 8th IFAC/IFIP/IFORS Symposium* (A. Pouliezos M. Papageorgiou, ed.), vol. 3, Chania, Greece, pp. 1025–1030, June16–18 1997.

[38] Rand EUROPE, “Validatie metanet een rapport voor rijkswaterstaat avv (validation metanet a report for rijkswaterstaat), traffic research centre,” Tech. rep. 0061, Rand EUROPE, Leiden, The Netherlands, July 2001. In Dutch.

[39] C.E. García, D.M. Prett, and M. Morari, “Model predictive control: Theory and practice — A survey,” *Automatica*, vol. 25, no. 3, pp. 335–348, May 1989.

[40] N. H. Gartner, “Development of demand-responsive strategies for urban traffic control,” in *Proceedings of the 11th IFIP Conf. Syst. Modelling and Optimiz.* (P. Thoft-Christensen, ed.), pp. 166–174, New York: Springer-Verlag, 1984.

[41] F. Glover, “Tabu search: A tutorial,” *Interfaces*, vol. 20, no. 4, pp. 74–94, 1990.

[42] F. Glover and M. Laguna, *Tabu Search*. Boston: Kluwer Academic Publishers, 1997.

[43] F. Glover, E. Taillard, and D. de Werra, “A user’s guide to tabu search,” *Annals of Operations Research*, vol. 41, pp. 3–28, 1993.

[44] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, Massachusetts: Addison-Wesley, 1989.

[45] Arti Gupta, Victor J. Maslanka, and Gary S. Spring, “Development of prototype knowledge-based expert system for managing congestion on massachusets turnpike,” *Transportation Research Record*, no. 1358, pp. 60–66, 1992.

[46] H. Haj-Salem and M. Papageorgiou, “Ramp metering impact on urban corridor traffic: Field results,” *Transportation Research Part A*, vol. 29, no. 4, pp. 303–319, 1995.

[47] F.L. Hall and K. Agyemang-Duah, “Freeway capacity drop and the definition of capacity,” *Transportation Research Record*, no. 1320, pp. 99–109, 1991.

[48] E. J. Hardman, “Motorway speed control strategies using SISTM,” in *Road Traffic Monitoring and Control*, no. 422 in Conference Publication, pp. 169–172, IEE, Apr.23-25 1996.

[49] Masroor Hasan, Mithilesh Jha, and Moshe Ben-Akiva, “Evaluation of ramp control algorithms using microscopic traffic simulation,” *Transportation Research Part C*, vol. C10, pp. 229–256, 2002.

[50] A. Hegyi, B. De Schutter, R. Babuška, S. Hoogendoorn, H. van Zuylen, and H. Schuurman, “A fuzzy decision support system for traffic control centers,” in *Proceedings of the TRAIL 6th Annual Congress 2000 — Transport, Infrastructure and Logistics*, Part 2, The Hague/Scheveningen, The Netherlands, Dec. 2000.

[51] A. Hegyi, B. De Schutter, and H. Hellendoorn, “Model predictive control for optimal coordination of ramp metering and variable speed control,” in *Proceedings of the 1st European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems (EUNITE 2001)*, Tenerife, Spain, pp. 260–272, Dec. 2001.

[52] A. Hegyi, B. De Schutter, and H. Hellendoorn, “Model predictive control for optimal coordination of ramp metering and variable speed control,” in *First International NAISO Conference on Neuro Fuzzy Technologies*, Havana, Cuba, Jan.16-19 2002.

[53] A. Hegyi, B. De Schutter, and H. Hellendoorn, “Optimal coordination of variable speed limits to suppress shock waves,” Tech. rep. CSE02-015, Control Systems Engineering, Fac. of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands, Aug. 2003. Accepted for publication in *Transportation Research Record*.

[54] A. Hegyi, B. De Schutter, and J. Hellendoorn, “Optimal coordination of variable speed limits to suppress shock waves,” in *Proceedings 7th TRAIL Congress* (P.H.L. Bovy, ed.), TRAIL Conference Proceedings, Rotterdam, The Netherlands, pp. 197 – 220, TRAIL, Nov.26 2002. ISBN: 90-407-2365-6.

[55] A. Hegyi, B. De Schutter, and J. Hellendoorn, “MPC-based optimal coordination of variable speed limits to suppress shock waves in freeway traffic,” in *Proceedings of the American Control Conference*, Denver, Colorado, USA, June 4–6 2003. Accepted for publication.

[56] A. Hegyi, B. De Schutter, and J. Hellendoorn, “Optimal coordination of variable speed limits to suppress shock waves,” *Transportation Research Record*, 2003. Accepted for publication.

[57] A. Hegyi, B. De Schutter, and J. Hellendoorn, “Optimal coordination of variable speed limits to suppress shock waves,” Tech. rep. 03-003, Delft Center for Systems and Control, Delft, The Netherlands, July 2003. Accepted for the 2003 IEEE Conference on Decision & Control (CDC 2003), Maui, Hawaii, Dec. 2003.

[58] A. Hegyi, B. De Schutter, and J. Hellendoorn, “Optimal coordination of variable speed limits to suppress shock waves — Addendum,” Tech. rep. CSE02-015A, Control Systems Engineering, Fac. of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands, Mar. 2003. See http://www.dcsc.tudelft.nl/˜bdeschutter/pub.

[59] A. Hegyi, B. De Schutter, S. Hoogendoorn, and H.J. van Zuylen R. Babuška, “Fuzzy decision support system for traffic control centers,” in *ESIT 2000*, pp. 389–395, September 2000. ISBN: 3-89653-797-0.

[60] A. Hegyi, B. De Schutter S. Hoogendoorn, R. Babuška, and H. Schuurman H.J. van Zuylen, “A fuzzy decision support system for traffic control centers,” in *2001 IEEE Intelligent Transportation Systems Proceedings* (B. Stone, ed.), Oakland (CA), USA, pp. pp. 358–363, Aug.25-29 2001.

[61] A. Hegyi, D. Ngo Duy, B. De Schutter, J. Hellendoorn, S.P. Hoogendoorn, and S. Stramigioli, “Suppressing shock waves on the A1 in The Netherlands — Model calibration and model-based predictive control,” Tech. rep. CSE03-001, Control Systems Engineering, Fac. of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands, Mar. 2003. Accepted for the 6th IEEE International Conference on Intelligent Transportation Systems (ITSC’03), Shanghai, China, Oct. 2003.

[62] 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*, Anchorage, Alaska, pp. 3600–3605, May 2002.

[63] D. Helbing, “High-fidelity macroscopic traffic equations,” *Physica A*, vol. 219, pp. 391–407, 1995.

[64] D. Helbing, “Derivation and empirical validation of a refined traffic flow model.,” *Physica A*, vol. 233, pp. 253–282, 1996.

[65] D. Helbing, *Verkehrsdynamik — Neue physikalische Modellierungskonzepte*. Berlin: Springer-Verlag, 1997.

[66] D. Helbing, “Derivation, properties, and simulation of a gas-kinetic-based, nonlocal traffic model,” *Physical Review E*, vol. 59, no. 1, pp. 239–253, 1999.

[67] D. Helbing, A. Hennecke, and M. Treiber, “Phase diagram of traffic states in the presence of inhomogeneities,” *Physical Review Letters*, vol. 82, no. 21, pp. 4360–4363, May 1999.

[68] D. Helbing and M. Schreckenberg, “Cellular automata simulating experimental properties of traffic flow,” *arXiv:cond-mat/9812300v2*, 21 Mar 1999.

[69] B. Hellinga and M. Van-Aerde, “Examining the potential of using ramp metering as a component of ATMS,” *Transportation Research Record*, no. 1494, pp. 169–172, 1995.

[70] J. Hernández, J. Cuena, and M. Molina, “Real-time traffic management through knowledge-based models: The TRYS approach,” *ERUDIT tutorial on Intelligent Traffic Management Models, Helsinki Finland*, Aug. 3 1999. http://www.erudit.de/erudit/events/tc-c/tut990803.htm.

[71] F.-S. Ho and P. Ioannou, “Traffic flow modelling and control using artificial neural networks,” *IEEE Control Systems Magazine*, vol. 16, no. 5, pp. 16–26, 1996.

[72] E. van den Hoogen and S. Smulders, “Control by variable speed signs: results of the dutch experiment,” in *7th International Conference on Road Traffic Monitoring and Control*, IEE Conference Publication No. 391, London, England, pp. 145–149, Apr. 26-28, 1994.

[73] S. P. Hoogendoorn and P.H.L. Bovy, “Modelling multiple user-class traffic flow,” *Transportation Research B*, vol. 34, no. 2, pp. 123–146, 2000.

[74] S.P. Hoogendoorn, “Optimal control of dynamic route information panels,” in *4th World Congress on Intelligent Transport Systems*, IFAC Transportation Systems, Chania, Greece, pp. 399–404, 1997.

[75] S.P. Hoogendoorn, *Multiclass Continuum Modelling of Multiclass Traffic Flow*. PhD thesis, Delft University of Technology, TRAIL Thesis Series, Delft, The Netherlands, Dec. 1999.

[76] S.P. Hoogendoorn and P.H.L. Bovy, “Continuum modelling of multiclass traffic flow,” *Transportation Research Part B*, vol. 34, pp. 123–146, 2000.

[77] S.P. Hoogendoorn and P.H.L. Bovy, “State-of-the-art of vehicular traffic flow modelling,” *Journal of Systems Control Engineer — Proceedings of the Institution of Mechanical Engineers, Part I*, vol. 215, no. 14, pp. 283–303, 2001.

[78] 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*, vol. 3, no. 1, pp. 21–38, 2003.

[79] S.P. Hoogendoorn, H. Schuurman, and B. De Schutter, “Real-time traffic management scenario evaluation,” in *Proceedings of the 10th IFAC Symposium on Control in Transportation Systems (CTS 2003)*, Tokyo, Japan, pp. 343–348, Aug. 2003.

[80] L. Jacobson, K. Henry, and O. Mehyar, “Real-time metering algorithm for centralized control,” *Transportation Research Record*, no. 1232, pp. 17–26, 1989.

[81] Abdessadek Karimi, “Integrated traffic control – on the integration of dynamic route guidance and ramp metering & prototyping in Matlab.,” Tech. rep. A03.009.857, AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management, 2003.

[82] H.R. Kashani and G.N. Saridis, “Intelligent control for urban traffic systems,” *Automatica*, vol. 19, no. 2, pp. 191–197, Mar. 1983.

[83] B. Kerner and H. Rehborn, “Experimental properties of phase transitions in traffic flow,” *Physical Review Letters*, vol. 79, no. 20, pp. 4030–4033, 17Nov. 1997.

[84] B. S. Kerner, “Empirical features of congested patterns at highway bottlenecks,” in *Proceedings of the 81st Annual Meeting of the Transportation Research Board*, Washington, D.C., 2002.

[85] B. S. Kerner and H. Rehborn, “Experimental features and characteristics of traffic jams,” *Physical Review E*, vol. 53, no. 2, pp. R1297–R1300, February 1996.

[86] G.J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, 1995.

[87] W.J.J. Knibbe, “Macroscopic traffic simulation for on-line forecasting,” in *Proceedings of the 9th IFAC Symposium on Control in Transportation Systems 2000*, Braunschweig, Germany, pp. 34–39, June 2000.

[88] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, “Traffic flow modelling of large-scale motorway networks using the macroscopic modelling tool METANET,” in *Recent Advances in Traffic Flow Modelling and Control, Proceedings of the expert seminar on recent advances in traffic flow modelling and*

control (P.H.L. Bovy and S.P. Hoogendoorn, eds.), TRAIL Conference Preceedings Series No.P99/2, The Netherlands TRAIL Research School, Sept. 20, 1999.

[89] 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*, vol. 3, no. 4, pp. 282–292, Dec. 2002.

[90] A. Kotsialos, M. Papageorgiou, H. Haj-Salem, S. Manfriedi, J. van Schuppen, J. Taylor, and M. Westerman, "DACCORD – Development and application of coordinated control of corridors," Tech. rep. Deliverable D06.1, CWI, Dec. 1997. Project TR 1017, Telematics Applications Programme, TRANSPORT.

[91] 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*, vol. 10, pp. 65–84, 2002.

[92] A. Kotsialos, M. Papageorgiou, and A. Messmer, "Integrated optimal control of motorway traffic networks," in *Proceedings of the 18th American Control Conference*, pp. 2183–2187, 1999.

[93] A. Kotsialos, M. Papageorgiou, and A. Messmer, "Optimal coordinated and integrated motorway network traffic control," in *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT)*, Jerusalem, Israel, pp. 621–644, July 1999.

[94] A. Kotsialos, M. Papageorgiou, and F. Middelham, "Optimal coordinated ramp metering with advanced motorway optimal control," in *Proceedings of the 80th Annual Meeting of the Transportation Research Board*, no. 01-3125, Washington, D.C., 2001.

[95] 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*, no. 1689, pp. 60–67, 1999.

[96] B. Krause and C. von Altrock, "A complete fuzzy logic control approach for existing traffic control systems," in *Mobility for Everyone, Proceedings of the 4th World Congress on Intelligent Transportation Systems*, Berlin, Germany, Oct. 1997. Paper no. 2045.

[97] R. D. Kühne, "Freeway control using a dynamic traffic flow model and vehicle reidentification techniques," *Transportation Research Record*, vol. 1320, pp. 251–259, 1991.

[98] A.C.B. de Langen, “Fuzzy Decision Support System: uitbreiding naar een groter wegennet (Fuzzy Decision Support System: extension to a larger road network),” Tech. rep., AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management and Delft University of Technology, Faculty of Civil Engineering and Geosciences, 2001. In Dutch.

[99] H. Lenz, R. Sollacher, and M. Lang, “Standing waves and the influence of speed limits,” in *Proceedings of the European Control Conference 2001*, Porto, Portugal, pp. 1228–1232, 2001.

[100] H. Lenz, R. Sollacher, and M. Lang, “Nonlinear speed-control for a continuum theory of traffic flow,” in *14th World Congress of IFAC*, vol. Q, Beijing, China, pp. 67–72, Jan.1999.

[101] P. Lertworawanich and L. Elefteriadou, “A methodology for estimating capacity at ramp weaves based on gap acceptance and linear optimization,” *Transportation Research Part B*, vol. 37, no. 5, pp. 459–83, 2003.

[102] J.B. Lesort and J.P. Lebacque, “The Godunov scheme and what is means for first order traffic flwo models,” in *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, pp. 647–677, July 1996.

[103] P. Y. Li, R. Horowitz, L. Alvarez, J. Frankel, and A. M. Robertson, “Traffic flow stabilization,” in *Proceedings of the American Control Conference*, Seattle, Washington, pp. 144–149, June 1995.

[104] M. J. Lighthill and G. B. Whitham, “On kinematic waves, II. a theory of traffic flow on long crowded roads,” *Proc. of the Royal Society*, vol. 229A, May 1955.

[105] H.K. Lo, E. Chang, and Y.C. Chan, “Dynamic network traffic control,” *Transportation Research Part A*, vol. 35, pp. 721–744, 2001.

[106] S. Logghe, *Dynamic Modeling of Heterogeneous vehicular traffic*. PhD thesis, Katholieke Universiteit Leuven, 2003.

[107] M. Lorenz and L. Elefteriadou, “A probabilistic approach to defining freeway capacity and breakdown,” in *Fourth International Symposium on Highway Capacity*, pp. 84–95, 2000.

[108] J.M. Maciejowski, *Predictive Control with Constraints*. Harlow, England: Prentice Hall, 2002.

[109] A. Messmer and Papageorgiou, “METANET: a macroscopic simulation program for motorway networks,” *Traffic Engineering and Control*, vol. 31, pp. 466–470, 1990.

[110] F. Middelham, “A synthetic study of the network effects of ramp metering.” AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management, 1999.

[111] F. Middelham, “State of practice in dynamic traffic management in The Netherlands,” in *Proceedings of the 10th IFAC Symposium on Control in Transportation Systems (CTS 2003),* Tokyo, Japan, Aug. 2003.

[112] F. Middelham, T. C. Wang, R. Koeijvoets, and H. Taale, “Flexsyt-ii-,” Tech. rep., AVV Transport Research Centre, Dutch Ministry of Transport, Public Works and Water Management, The Netherlands, 1994.

[113] M. Molina, J. Hernández, and J. Cuena, “A structure of problem-solving methods for real-time decision support in traffic control,” *International Journal of Human-Computer Studies*, vol. 49, pp. 577–600, 1998.

[114] K. Nagel, “Particle hopping models and traffic flow theory,” *Physical Review E*, vol. 53, pp. 4655–4673, 1996.

[115] K. Nagel, “From particle hopping models to traffic flow theory,” *Transportation Research Record*, vol. 1644, pp. 1–9, 1998.

[116] H.T. Nguyen and E.A. Walker, *A First Course in Fuzzy Logic*. RCS Publications, 2nd ed., 1999.

[117] J. Nocedal and S.J. Wright, *Numerical Optimization*. Springer series in operation research, Springer, 1999.

[118] D. Nogduy and S. P. Hoogendoorn, “An automated calibration procedure for macroscopic traffic flow models,” in *Proceedings of the 10th IFAC Symposium on Control in Transportation Systems (CTS 2003)*, Tokyo, Japan, pp. 295–300, Aug. 2003.

[119] Pichai Pamanikabud and Prakob Vivitjinda, “Noise prediction for highways in Thailand,” *Transportation Research Part D*, vol. 7, no. 6, pp. 441–449, 2002.

[120] M. Papageorgiou, “A new approach to time-of-day control based on a dynamic freeway traffic model,” *Transportation Research Part B*, vol. 14B, pp. 349–360, 1980.

[121] M. Papageorgiou, *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*, vol. 50 of *Lecture Notes in Control and Information Sciences*. Springer Verlag, Berlin, 1983.

[122] M. Papageorgiou, "Certainty equivalent open-loop feedback control applied to multireservoir networks," *IEEE Transactions on Automatic Control*, vol. 33, no. 4, pp. 392–399, 1988.

[123] M. Papageorgiou, "Dynamic modeling, assignment, and route guidance in traffic networks," *Transportation Research Part B*, vol. 24B, no. 6, pp. 471–495, 1990.

[124] M. Papageorgiou, "Some remarks on macroscopic traffic flow modelling," *Transportation Research Part A*, vol. 32, no. 5, pp. 323–329, 1998.

[125] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, "Modelling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris: Part I: Modelling," *Transportation Research Part A*, vol. 24A, no. 5, pp. 345–359, 1990.

[126] M. Papageorgiou, J.-M. Blosseville, and H. Haj-Salem, "Modelling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris: Part II: coordinated on-ramp metering," *Transportation Research Part A*, vol. 24A, no. 5, pp. 361–370, 1990.

[127] M. Papageorgiou, J.M. Blosseville, and H. Hadj-Salem, "La fluidification des rocades de l'Ile de France: Un projet d'importance.," Tech. rep., Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania, Greece, 1998. Internal Report No. 1998-17.

[128] M. Papageorgiou, H. Hadj-Salem, and F. Middelham, "ALINEA local ramp metering — summary of field result," *Transportation Research Record*, no. No. 1603, pp. 90–98, 1998.

[129] M. Papageorgiou, H. Haj-Salem, and F. Middelham, "ALINEA local ramp metering: Summary of field result," in *Proceedings of the 76th Annual Meeting of the Transportation Research Board*, Washington, D.C., 1997.

[130] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: An overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271–280, Dec. 2002.

[131] M. Papageorgiou and A. Messmer, "Dynamic network traffic assignment and route guidance via feedback regulation," *Transportation Research Record*, vol. 1306, pp. 49–58, 1991.

[132] M. Papageorigiou, H. Hadj-Salem, and J.-M. Blosseville, "ALINEA: a local feedback control law for on-ramp metering," *Transportation Research Record*, no. 1320, pp. 58–64, 1991.

[133] S.L. Paveri-Fontana, “On Boltzmann-like treatments for traffic flow: A critical review of the basic model and an alternative proposal for dilute traffic analysis,” *Transportation Research Part B*, vol. 9, pp. 225–235, 1975.

[134] H.J. Payne, “Models of freeway traffic and control,” *Simulation Council Proceedings*, no. 1, pp. 51–61, 1971.

[135] W. Pedrycz and L.A. Zadeh, *Fuzzy Sets Engineering*. CRC Press, 1995.

[136] PTV, “Vissim – traffic flow simulation,” Tech. rep., PTV, Germany, 2003. http//:www.ptv.de.

[137] Quadstone, “Paramics v4.0 system overview,” Tech. rep., Quadstone Limited, Edinburgh, Scotland, 2002. http//:www.paramics-online.com.

[138] S.S. Rao, *Engineering optimization*. John Wiley & Sons, 1996.

[139] I. Rees, “Orbital decongestant,” *Highways*, vol. 63, no. 5, pp. 17–18, July/Aug. 1995.

[140] P.I. Richards, “Shock waves on the highway,” *Operations Research*, vol. 4, pp. 42–57, 1956.

[141] Stephen G. Ritchie, “A knowledge-based decision support architecture for advanced traffic management,” *Transportation Research Part A*, vol. 24A, no. 1, pp. 27–37, 1990.

[142] Stephen G. Ritchie and Neil A. Prosser, “Real-time expert system approach to freeway incident management,” *Transportation Research Record*, no. 1320, pp. 7–16, 1991.

[143] D.I. Robertson and R.D. Bretherton, “Optimizing networks of traffic signals in real time — The SCOOT method,” *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 11–15, Feb. 1991.

[144] David Roseman and Sun-Sun Tvedten, “Santa monica freeway smart corridor project operational multi-agency traffic management and expert system,” in *MOBILITY FOR EVERYONE 4th World Congress on Intelligent Transport Systems*, ITS America, 21-24 October 1997.

[145] A. W. Sadek and M. J. Demetsky, “Case-based reasoning for real-time traffic flow management,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 14, pp. 347–356, 1999.

[146] D.J. Sentinella and E.J. Hardman, “Review of motorway speed control systems in europe,” Tech. rep., Transport Research Laboratory, 1996. PR/TT/056/96, unpublished.

[147] S. Smulders, “Control of freeway traffic flow by variable speed signs,” *Transportation Research Part B*, vol. 24B, no. 2, pp. 111–132, 1990.

[148] S. Smulders, “Control by variable speed signs — the Dutch experiment,” in *Proceedings of the Sixth International Conference on Road Traffic Monitoring and Control*, IEE Conference Publication, London, pp. 99–103, IEE, Apr.28-30 1992.

[149] S. Smulders, *Control of Freeway Traffic Flow*. CWI Tract no. 80, CWI (Dutch institute for research in Mathematics and Computer Science), 1996. ISBN: 90-6196-451-2.

[150] S. A. Smulders and D. E. Helleman, “Variable speed control: State-of-the-art and synthesis,” in *Road Transport Information and Control*, no. 454 in Conference Publication, pp. 155–159, IEE, Apr.21-23 1998.

[151] A.R.M. Soeterboek, *Predictive Control — A Unified Approach*. New Jersey: Prentice Hall, Englewood Cliffs, 1992.

[152] H. Taale, “The combined traffic assignment and control problem: An overview of 25 years research.” Delft University of Technology, Transportation Planning and Traffic Engineering Section, 2000.

[153] 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, London, UK, pp. 106–110, Apr. 2000.

[154] H. Taale and H.J. van Zuylen, “Traffic control and route choice: Occurence of instabilities,” in *Five Years ”Crossroads of Theory and Practice”- TRAIL 5th Annual Congress 1999* (Piet H.L. Bovy, ed.), vol. 2 of *TRAIL Conference Proceedings*, Dec. 1999.

[155] C.J. Taylor, P.C. Young, A. Chotai, and J. Whittaker, “Nonminimal state space approach to multivariable ramp metering control of motorway bottlenecks,” *IEE Proceedings – Control Theory Applications*, vol. 146, no. 6, pp. 568–574, Nov. 1998.

[156] Technical University of Crete and A. Messmer, *METANET – A simulation program for motorway networks*. Technical University of Crete, Dynamic Systems and Simulation Laboratory and A. Messmer, Nov. 2001.

[157] H. Theil, “A multinomial extension of the linear logit model,” *International Economic Review*, vol. 10, pp. 251–259, 1969.

[158] J. van Toorenburg and R.J.P. van der Linden, “Predictive control in traffic-management, an investigation of opportunities for application of predictive control in traffic-management on networks of major routes with the assumption that a short-term traffic-forecast is available,” Tech. rep., Transpute & Dutch Ministry of Transport, Public Works and Water Management, The Netherlands, Mar. 1996.

[159] M. Treiber, A. Hennecke, and D. Helbing, “Congested traffic states in empirical observations and microscopic simulations,” *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, Aug. 2000.

[160] M. van den Berg, B. De Schutter, and A. Hegyi, “Model predictive control for mixed urban and freeway networks,” Tech. rep., Delft Center for Systems and Control, Delft, The Netherlands, July 2003. Accepted for publication in the *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, Washington, D.C., Jan. 2004.

[161] 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,” Tech. rep. 03-002, Delft Center for Systems and Control, Delft, The Netherlands, July 2003. Accepted for the 2003 IEEE Conference on Decision & Control (CDC 2003), Maui, Hawaii, Dec. 2003.

[162] J.A.C. van Toorenburg, “Toepasbaarheid van toeritdosering (applicability of ramp metering),” *Verkeerskunde*, vol. 40, no. 6, pp. 284–289, 1988.

[163] Francesco Viti, Stella Catalano Fiorenzo, Minwei Li, Charles Lindveld, and Henk van Zuylen, “An optimization problem with dyamice route-departure time choice and pricing,” in *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*, Washington, D.C., 2003.

[164] Y. Wang, M. Papageorgiou, and A. Messmer, “A predictive feedback routing control strategy for freeway network traffic,” in *Proceedings of the American Control Conference*, Anchorage, Alaska, pp. 3606–3611, May 2002.

[165] Y. Wang, M. Papageorgiou, and A. Messmer, “A predictive feedback routing control strategy for freeway network traffic,” in *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*, Washington, D.C, Jan. 2003.

[166] J. A. Wattleworth, “Peak-period analysis and control of a freeway system,” *Highway Research Record*, no. 157, pp. 1–21, 1965.

[167] Julia K. Wilkie, “Using variable speed limit signs to mitigate speed differentials upstream of reduced flow locations,” Tech. rep., Department of Civil Engineering, Texas A& M University, College Station, Texas 77843, Aug. 1997. Prepared for CVEN 677 Advanced Surface Transportation Systems.

[168] B. Williams, “Highway control,” *IEE Review*, pp. 191–194, Sept. 1996.

[169] D. E. Wolf, “Celllular automata for traffic simulations,” *Physica A*, vol. 263, pp. 438–451, 1999.

[170] H. Zackor, “Beurteilung verkehrsabhängiger Geschwindigkeitsbeschränkungen auf Autobahnen.,” *Strassenbau und Verkehrstechnik*, 1972.

[171] H. Zackor, “Self-sufficient control of speed on freeways,” in *Proceedings of the International Symposium on Traffic Control Systems*, vol. 2A, Berkeley, California, pp. 226–249, California University, Aug.6–9 1979.

[172] H.M. Zhang, “Structural properties of solutions arising from a nonequilibrium traffic flow theory,” *Transportation Research Part B*, vol. 34, pp. 583–603, 2000.

[173] Hongjun Zhang and Stephen G. Ritchie, “Real-time decision-support system for freeway management and control,” *Journal of Computing in Civil Engineering*, vol. 8, no. 1, pp. 35–51, January 1994.

[174] N. J. van der Zijpp and R. Hamerslag, “Improved kalman filtering approach for estimating origin-destination matrices for freeway corridors,” *Transportation Research Record*, no. 1443, pp. 54–64, 1994.

[175] N. van der Zijpp, *Dynamic Origin-Destination Matrix Estimation on Motorway Networks*. PhD thesis, Delft University of Technology, 1996.

[176] H. J. van Zuylen and G. H. J. van Schaik, “The development of an integrated technology policy for transport,” in *Proceedings of the 77th Annual Meeting of the Transportation Research Board*, 1998.

[177] H. J. van Zuylen and K. M. Weber, “Strategies for european innovation policy in the transport field,” *Technological Forecasting & Social Change*, no. 69, pp. 929–951, 2002.

[178] H.J. van Zuylen, “The assessment of economic benefits of dynamic traffic management,” in *European Transport Conference Proceedings*, Cambridge, England, p. 21, Association for European Transport, 2002.

[179] H.J. van Zuylen and T.H. Muller, “Regiolab Delft,” in *9th World Congress on Intelligent Transport Systems*, Chicago, Illinois, USA, pp. 9–, Oct.14–17 2002.

Curriculum Vitae

Andreas Hegyi was born on the 27th of March 1973 in Leeuwarden, The Netherlands. He received his secondary education in Budapest, Hungary and in Zwijndrecht, The Netherlands. Next, he studied Electrical Engineering in the period of 1991–1998 at the Delft University of Technology, Delft, The Netherlands, and received his M.Sc. degree in Electrical Engineering in 1998. The concluding thesis work was on modeling abductive reasoning in fuzzy logic.

Since 1999 Andreas is affiliated with the Delft Center for Systems and Control (formerly Control Systems Engineering group) at the Delft University of Technology, where he has worked on the development of a prototype fuzzy decision support system. In 2000 he started his Ph.D. research which was focused on traffic control with model predictive control. This research was supported by the AVV Transport Research Centre of the Dutch Ministry of Transport, Public Works and Water Management and the *Mobility of People and Transportation of Goods* spearhead program of the Delft University of Technology.

After finishing his Ph.D. he continued his research in the same group.

TRAIL Thesis Series

A series of The Netherlands TRAIL Research School for theses on transport, infrastructure and logistics.

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

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

Veenstra, A.W., *Quantitative Analysis of Shipping Markets*, T99/3, April 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

Hoogendoorn, S.P., *Multiclass Continuum Modelling of Multilane Traffic Flow*, T99/5, September 1999, 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

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

Subiono, *On Classes of Min-max-plus Systems and their Applications*, T2000/2, June 2000, 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

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

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

Roodbergen, K.J., *Layout and Routing Methods for Warehouses*, T2001/3, May 2001, TRAIL Thesis Series, 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

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

Rosmuller, N., *Safety analysis of Transport Corridors,* T2001/6, June 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

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

Wolters, M.J.J., *The Business of Modularity and the Modularity of Business,* T2002/1, February 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

Koppius, O.R., *Information Architecture and Electronic Market Performance,* T2002/3, May 2002, TRAIL Thesis Series, 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

Van Nes, R.*, Design of multimodal transport networks, a hierarchical approach*, T2002/5, September 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

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

Spek, S.C., van der, *Connectors. The Way beyond Transferring*, T2003/1, February 2003, TRAIL Thesis Series, Delft University Press, The Netherlands

Lindeijer, D.G., *Controlling Automated Traffic Agents*, T2003/2, February 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

Reeven, P.A., van, *Competition in Scheduled Transport*, T2003/4, April 2003, TRAIL Thesis Series, Eburon, The Netherlands

Peeters, L.W.P., *Cyclic Railway Timetable Optimization*, T2003/5, June 2003, TRAIL Thesis Series, 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

Lindveld, Ch..D.R., *Dynamic O-D matrix estimation: a behavioural approach*, T2003/7, September 2003, TRAIL Thesis Series, Eburon, The Netherlands

Weerdt, M.M., de, *Plan Merging in Multi-Agent Systems*, T2003/8, December 2003, TRAIL Thesis Series, The Netherlands

Langen, P.W, de, *The Performance of Seaport Clusters*, T2004/1, January 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