Delft University of Technology

Fac. of Information Technology and Systems **Control Systems Engineering**

Technical report CSE02-013

Optimal Coordination of Variable Speed Limits to Suppress Shock Waves*

A. Hegyi, B. De Schutter, and J. Hellendoorn

To cite this work, please refer to the published version:

A. Hegyi, B. De Schutter, and J. Hellendoorn, "Optimal coordination of variable speed limits to suppress shock waves," *Proceedings of the 7th TRAIL Congress 2002 – TRAILblazing into the Future – Selected Papers* (P. H. L. Bovy, ed.), Rotterdam, The Netherlands, pp. 197–220, Nov. 2002.

Control Systems Engineering
Faculty of Information Technology and Systems
Delft University of Technology
Delft, The Netherlands
Current URL: https://www.dcsc.tudelft.nl

* This report can also be downloaded via https://dspub.eu/02-013

Optimal Coordination of Variable Speed Limits to Suppress Shock Waves

TRAIL Research School, Delft, November 2002

Authors

Ir. Andreas Hegyi, Dr. ir. B. De Schutter, Prof. dr. ir. J. Hellendoorn
Faculty of ITS, Delft University of Technology, Control Systems Engineering

Contents

Abstract

1 Introduction 1

2 Problem description 4

3 Approach 5

3.1 Model Predictive Control 5

3.2 Prediction model 5
3.2.1 Original METANET model 5
3.2.2 Extensions 7

3.3 Objective function 8

3.4 Constraints 8

3.5 Tuning of N_{p} and N_{c} 9

4 A benchmark problem 10

4.1 Set-up 10

4.2 Results 11

5 Conclusions and future research 15

Acknowledgements 16

References 17

Abstract

We present a model predictive control (MPC) approach to optimally coordinate variable speed limits for highway traffic. A safety constraint is formulated that prevents drivers from encountering speed limit drops larger than, say, 10 km/h is incorporated in the controller. The control objective is to minimize the total time that vehicles spend in the network. This approach results in dynamic speed limits that reduce or eliminate shock waves. For the prediction of the evolution of the traffic flows in the network, which is needed for MPC, we use an adapted version of the METANET model that takes the variable speed limits into account. The performance of the discrete-valued and safety constrained controllers is compared with the performance of the continuous-valued unconstrained controller. It is found that both types of controllers result in a network with less congestion, a higher outflow, and a lower total time spent. For our benchmark problem, the performance of the discrete controller with safety constraints is comparable to the continuous controller without constraints.

Keywords

variable speed limits, coordinated control, safe speed limits, shock wave reduction, model predictive control

1 Introduction

As the number of vehicles and the need for transportation grow, cities around the world face considerable traffic congestion problems: almost every weekday morning and evening during rush hours the saturation point of the highways and the main roads in and around the city is attained. Traffic jams do not only cause considerable costs due to unproductive time losses, but they also augment the possibility of accidents, and they have a negative impact on the environment and on the quality of life. On the short term the most effective measures in the battle against traffic congestion seem to be a selective construction of new roads and a better control of traffic by dynamic traffic management measures. We will concentrate on the latter option. In practice, dynamic traffic management usually operates based on local data only. However, considering the effect of the measures on the network level has in general many advantages compared to local control. So, a network-wide *coordination* of control measures, based on global data is necessary. Since the effect of a control measure on more distant locations might only be visible after some time, a *prediction* of the network evolution is also necessary to achieve optimal network control. The approach presented in this paper contains both elements: network-wide coordination and prediction. In this paper we consider a special case of traffic control measures: variable speed limits to reduce or eliminate shock waves. Also in this case prediction and coordination is 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 can not be observed instantaneously,[1] so the prediction should be made at least up to the point where the improvement can be observed. Besides prediction and coordination the speed limit control problem has other characteristics imposes certain requirements to the control strategy.

1. There is a direct relation between the outflow of a network and the total time spent (TTS) in the network, assuming that the traffic demand is fixed. Papageorgiou Papageorgiou et al. (1998) showed that in a traffic network an increase of outflow of 5 % may result in an decrease of the total time spent in the network of 20 %. This effect can be explained by the fact that the number of vehicles in the network is equal to the accumulated net inflow of the network (where the net inflow is the difference between the inflow and the outflow). But the outflow is lower when the traffic is congested[2], so the queue grows faster, and consequently congestion will last longer, and the outflow will be low for a longer time (the time that the queue needs to dissolve). This is why one should try to prevent or postpone a breakdown as much as possible. We can conclude that any control method that resolves (reduces) congestion will at best achieve a flow improvement of approximately 5–10 %, but this improvement can decrease the TTS significantly. This also means that the control strategy requires great precision, and since there are always (unpredictable) disturbances present in a traffic network *feedback control*

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

[2] The congestion after a breakdown usually has an outflow that is (only 5–10 %!) lower than the available capacity; this is the so called capacity-drop phenomenon.

is required.

2. 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-valued control signals.

3. For safety it is often required that the driver should not encounter a decrease in the speed limit larger than a prespecified amount. The controller should be able to take this kind of constraints into account.

The control strategy presented in this paper addresses these issues. In literature basically two views on the use of speed limits can be found. The first emphasizes the homogenization effect Alessandri et al. (1998, 1999); Febbraro et al. (2001); Hoogen and Smulders (1994); Smulders (1990, 1996), whereas the second is more focused on the prevention of traffic breakdown by flow reduction accomplished by speed limits. Chien et al. (1997); Lenz et al. (1999, 2001). The idea of homogenization is that speed limits reduce the speed differences between vehicles, by which a higher (and safer) flow can be achieved. The homogenization approach typically uses speed limits that are close to the critical speed (i.e. the speed that corresponds to the maximal flow; see Figure 1).

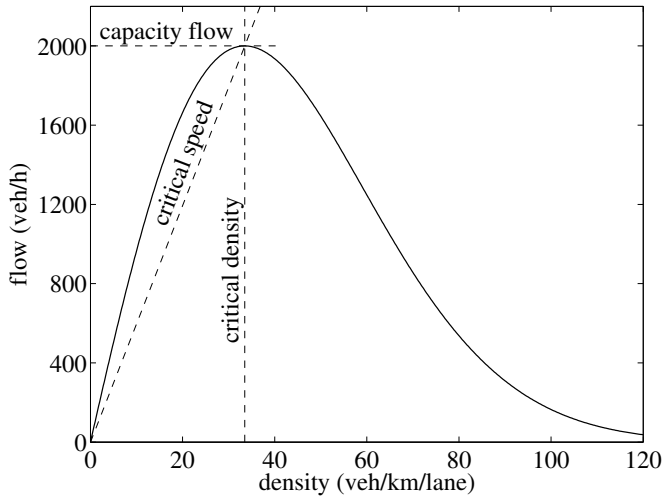

Figure 1: A typical example of the fundamental diagram. The critical speed is the speed that corresponds to maximum flow. The slope of the line connecting the origin and a point on the fundamental diagram represents the speed corresponding to that point.

The traffic breakdown prevention approach focuses more on preventing too high densities, and it also allows lower than critical speed limits. The results in Hoogen and Smulders (1994) indicate that the effect of homogenization on motorway performance is small; however, a positive safety effect can be expected. To the authors' best knowledge there are currently no published results available of experiments in connection with using speed limits to prevent traffic breakdown. Currently, the Dutch Ministry of Transport, Public Works and Water Management is preparing an experiment in the DYVERS project where the reduction of congestion by enforced dynamic speed limits is studied.

Several control methodologies are used in literature to find a control law for speed control, such as multi-layer control Papageorgiou (1983), sliding-mode control Lenz et al. (1999, 2001), or optimal control Alessandri et al. (1998, 1999). In Febbraro et al. (2001) the optimal control is approximated by a neural network in a rolling horizon framework.

Most of the models used in literature represent the speed limits by a factor that downscales the fundamental diagram Alessandri et al. (1998, 1999); Lenz et al. (1999). This can give too optimistic results (see Section 3.2.2), and therefore we use the METANET model which we extend with an equation that models the effect of a speed limit. We also introduce an equation to express the difference in the drivers' anticipation to increasing or decreasing downstream densities.

In a previous paper Breton et al. (2002) we demonstrated the effectiveness of continuous-valued speed limits against shock waves. Here we focus on discrete-valued speed limits and the constraints following from the safety considerations.

The organization of the paper is as follows. In Section 2 the problem and the basic idea of the solution of moving jams is described. In Section 3 the basic ingredients of model predictive control are introduced, and the prediction model including the extensions is presented. The proposed control method is applied to a benchmark problem in Section 4. Finally the conclusions and topics for future research are stated in Section 5.

2 Problem description

It is well known (see, e.g., Kerner and Rehborn (1996)) 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 motorway upstream of the jammed area will have to pass through 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. Lighthill and Whitham Lighthill and Whitham (1955) 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 paper we will use the term "shock wave" for any wave (the moving jammed areas) and not distinguish between waves and shock waves, because in practice any wave is undesired. To suppress shock waves speed limits can be used in the following way. On 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 it would be in the uncontrolled situation) that propagates downstream. This low density wave meets the shock wave and compensates its high density, which reduces or eliminates the shock wave. 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 higher delays than in the uncontrolled case. This can be explained in terms of stable, metastable, and unstable traffic flow states observed by Kerner and Rheborn Kerner and Rehborn (1996). Stable means that any (no matter how large) disturbance will vanish without intervention. Metastable means that small disturbances will vanish, but large disturbances will create a shock wave. Unstable means that any (no matter how small) disturbance will trigger a shock wave. For the application of speed limits against shock waves, the metastable state is a necessary condition, 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 to limit the flow without creating large disturbances. In the following sections we demonstrate how the proper speed limits can be found.

3 Approach

3.1 Model Predictive Control

We use a model predictive control (MPC) scheme to solve the problem of optimal coordination of speed limits. In MPC, at each time step k the optimal control signal is computed (by numerical optimization) over a prediction horizon N_{p}. A control horizon N_{c} ($< N_{\mathrm{p}}$) is selected to reduce the number of variables and to improve the stability of the system. After the control horizon has been passed the control signal is usually taken to be constant. In addition, a rolling horizon strategy is used, which means that at each time step only the first sample of the optimal control signal is applied to the system; afterwards the time axis is shifted one sample step, the model is updated, and the procedure is restarted. The rolling horizon approach results in an on-line predictive and adaptive control scheme that allows us to take changes in the system or in the system parameters into account by regularly updating the model of the system or the predicted demands as new measurements from the traffic sensors become available. For more information on MPC see Camacho and Bordons (1995); Maciejowski (2002) and the references therein.

3.2 Prediction model

The MPC procedure includes a prediction of the network evolution as a function of the current state and a given control input. For this prediction we use a slightly modified version of the (destination-independent) METANET model Kotsialos et al. (1999); Papageorgiou et al. (1990b). The modifications are introduced for better modeling of shock waves and the effect of speed limits. Note that the MPC approach is generic and will find the optimal speed limits independent from the model that is used (e.g. the way that speed limits enter the model), so the modifications are not necessary for the effectivity of MPC. For the sake of brevity, we describe only those parts of the model that are relevant for interpreting and understanding the simulation results of our benchmark network (see Section 4).

3.2.1 Original METANET model

The METANET model represents a network as a directed graph with the links corresponding to highway stretches. Each motorway link has uniform characteristics, i.e., no on-ramps or off-ramps and no major changes in geometry. Each link m is divided into N_m segments of length L_m (see Figure 2). Each segment i of link m is characterized by the *traffic density* $\rho_{m,i}(k)$ (veh/lane/km), the *mean speed* $v_{m,i}(k)$ (km/h), and the *traffic volume* or *flow* $q_{m,i}(k)$ (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 $T = 10\,\mathrm{s}$).

The following equations describe the evolution of the network over time. 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):

$$q_{m,i}(k) = \rho_{m,i}(k)\, v_{m,i}(k)\, \lambda_m \quad . \tag{1}$$

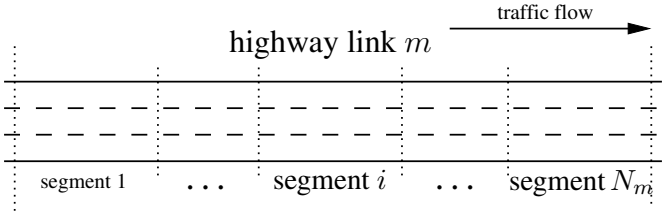

Figure 2: In the METANET model a motorway link is divided into segments.

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):

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

The mean speed equals the previous mean speed 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:

$$\begin{aligned} v_{m,i}(k+1) = v_{m,i}(k) &+ \frac{T}{\tau}\Big(V\big(\rho_{m,i}(k)\big) - v_{m,i}(k)\Big) + \\ &\frac{T}{L_m} v_{m,i}(k)\big(v_{m,i-1}(k) - v_{m,i}(k)\big) - \\ &\frac{\nu T}{\tau L_m}\frac{\rho_{m,i+1}(k) - \rho_{m,i}(k)}{\rho_{m,i}(k) + \kappa} \ , \end{aligned} \tag{2}$$

where τ, ν and κ are model parameters, and with

$$V\big(\rho_{m,i}(k)\big) = v_{\text{free},m} \ \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_{\text{crit},m}}\right)^{a_m}\right] \ , \tag{3}$$

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 flowing freely, and the critical density $\rho_{\text{crit},m}$ is the density at which the traffic flow becomes unstable. Origins are modeled with a simple queue model. The length of the queue $w_o(k)$ equals the previous queue length plus the demand[3] $d_o(k)$, minus the outflow $q_o(k)$:

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

The outflow depends on the traffic conditions on the motorway and the capacity of the origin. The flow $q_o(k)$ is the minimum of the demand and the maximal flow that can enter the motorway given the mainstream conditions:

$$q_o(k) = \min\left[d_o(k) + \frac{w_o(k)}{T}, Q_o\frac{\rho_{\max} - \rho_{\mu,1}(k)}{\rho_{\max} - \rho_{\text{crit},\mu}}\right], \tag{4}$$

where Q_o is the on-ramp capacity (veh/h) under free-flow conditions, $\rho_{\max}$ is the maximum density, and μ the index of the link to which the on-ramp is connected.

[3]Just as in Kotsialos et al. (1999, 2001); Papageorgiou et al. (1990a) we assume that the demand is independent of any control actions taken in the network.

3.2.2 Extensions

Since the original METANET model does not describe the effect of speed limits, we have slightly modified the equation for the desired speed (3) to incorporate speed limits. The second extension regards the modeling of the different nature of a mainstream origin as opposed to an on-ramp origin. The third extension considers the different effect of the downstream density gradient on the speed (cf. the anticipation term in (2)) when this gradient is positive or negative. In some publications the effect of the speed limit is expressed by scaling down the desired speed-density diagram Alessandri et al. (1998, 1999); Lenz et al. (1999). This changes the whole speed-density diagram, also for the states where the speed would otherwise be lower than the value of the speed limit. This means, e.g., that if the free flow speed is 120 km/h and the displayed speed limit is 100 km/h then the speed and flow of the traffic are reduced even when the vehicles are traveling at 80 km/h. Furthermore, scaling down the desired speed also reduces the capacity, while there is no reason to assume that a speed limit *above* the critical speed (speeds where the flow has not reached capacity yet) would reduce the capacity of the road. These assumptions are rather unrealistic, and they exaggerate the effect of speed limits. However, to get a more realistic model for the effects of the speed limits, we assume that the desired speed is the minimum of the following two quantities: the desired speed based on the experienced density, and the desired speed caused by the speed limit displayed on the variable message sign (VMS):

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

where $v_{\text{ctrl},m,i}(k)$ is the speed limit imposed on segment i, link m, at time k. To express the different nature of a mainstream origin link o compared to a regular on-ramp (the queue at a mainstream origin is in fact an abstraction of the sections upstream of the origin of the part of the motorway network that we are modeling), we use a modified version of (4) with another flow constraint, because the inflow of a segment (and thus the outflow of the mainstream 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 flow equals the flow that follows from the speed-flow relationship from (1) and (3) 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

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

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

$$q_{\text{lim},\mu,1}(k) = \begin{cases} \lambda_\mu\, v_{\text{lim},\mu,1}(k)\,\rho_{\text{crit},\mu}\left[-a_\mu \ln\left(\dfrac{v_{\text{lim},\mu,1}(k)}{v_{\text{free},m}}\right)\right]^{\frac{1}{a_\mu}} \\ \qquad\qquad\qquad\qquad \text{if } v_{\text{lim},\mu,1}(k) < V(\rho_{\text{crit},\mu}) \\ q_{\text{cap},\mu} \qquad\qquad\qquad\qquad \text{if } v_{\text{lim},\mu,1}(k) \geq V(\rho_{\text{crit},\mu}), \end{cases}$$

where $v_{\text{lim},\mu,1}(k) = \min(v_{\text{ctrl},\mu,1}(k), v_{\mu,1}(k))$ is the speed that limits the flow, and $q_{\text{cap},\mu} = \lambda_\mu V(\rho_{\text{crit},\mu})\rho_{\text{crit},\mu}$ is the capacity flow. Since the effect of a higher downstream density is usually stronger than the effect of a lower downstream density, we distinguish between these two cases. The sensitivity of the speed to the downstream density is expressed by parameter ν. In (2) ν is a global parameter and has the same value for all segments. However, here we take different values for $\nu_{m,i}(k)$ depending on whether the downstream density is higher or lower than the density in the actual segment:

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

In addition, when there is no entering link (but a mainstream origin) we assume that the speed of the (virtual) entering link equals the speed of the first segment:

$$v_{m,0}(k) = v_{m,1}(k) \ . \tag{5}$$

This is a good approximation of the speed behavior when there are enough (e.g., three or more) uncontrolled upstream segments.

3.3 Objective function

We consider the following objective function:

$$J(k) = T \sum_{l=k}^{k+N_{\text{p}}-1} \left\{ \sum_{(m,i)\in I_{\text{all}}} \rho_{m,i}(l) L_m \lambda_m + \sum_{o\in O_{\text{all}}} w_o(l) \right\} + a_{\text{speed}} \sum_{l=k}^{k+N_{\text{c}}-1} \sum_{(m,i)\in I_{\text{speed}}} \left(\frac{v_{\text{ctrl},m,i}(l) - v_{\text{ctrl},m,i}(l-1)}{v_{\text{free},m}} \right)^2 ,$$

where I_{all} and O_{all} are the sets of indices of all pairs of segments and links and of all origins respectively, and I_{speed} is the set of pairs of indices (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 a_{speed}.

3.4 Constraints

In general, for the safe operation of a speed control system, it is required that the maximum decrease of speed limits that a driver can encounter (v_{maxdiff}) is limited. There are three situations where a driver can encounter a different speed limit value: (1) when the speed limit changes on a given segment (and there are more speed limit signs on the same segment), (2) when a driver enters a new segment, (3) when the driver enters a new segment and the speed limit changes. The maximum speed difference constraints in the three situations are formulated as follows:

$$v_{\text{ctrl},m,i}(l-1) - v_{\text{ctrl},m,i}(l) \leq v_{\text{maxdiff}} \quad \text{for all } (m,i,l) \quad \text{such that}$$

$$
\begin{aligned}
& & & (m,i) \in I_{\text{speed}} \text{ and} \\
& & & l \in [k, \ldots, k+N_{\text{c}}-1], \\
v_{\text{ctrl},m,i}(l) - v_{\text{ctrl},m,i+1}(l) \leq & v_{\text{maxdiff}} & \text{for all } & (m,i,l) \quad \text{such that} \\
& & & (m,i) \in I_{\text{speed}} \text{ and} \\
& & & (m,i+1) \in I_{\text{speed}} \text{ and} \\
& & & l \in [k, \ldots, k+N_{\text{c}}-1], \\
v_{\text{ctrl},m,i}(l-1) - v_{\text{ctrl},m,i+1}(l) \leq & v_{\text{maxdiff}} & \text{for all } & (m,i,l) \quad \text{such that} \\
& & & (m,i) \in I_{\text{speed}} \text{ and} \\
& & & (m,i+1) \in I_{\text{speed}} \text{ and} \\
& & & l \in [k, \ldots, k+N_{\text{c}}-1].
\end{aligned}
$$

In addition to the safety constraints the speed limits are often subject to a minimum value v_{ctrlmin}:

$$v_{\text{ctrl},m,i} \geq v_{\text{ctrlmin}} \quad \text{for all } (m,i) \in I_{\text{speed}}.$$

3.5 Tuning of N_{p} and N_{c}

In conventional MPC heuristic tuning rules have been developed to select appropriate values for N_{p} and N_{c} (see Maciejowski (2002)). However, these rules cannot be straightforwardly applied the traffic flow control framework presented above.
For the prediction horizon N_{p} should be larger than the maximum travel time between the control inputs and the exit (under presence of a shock wave), because the vehicles that are influenced by the current control measure have only an effect on the network performance when they exit the network. 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. So based on this heuristic reasoning we select N_{p} to be about the typical travel time in the network when a shock wave is present. For the control horizon N_{c} we select a value that represents a trade-off between the computational effort and the performance.

4 A benchmark problem

In order to illustrate the control framework presented above we will now apply it to benchmark set-up consisting of a motorway link equipped with variable speed signs.

4.1 Set-up

The benchmark set-up consists of one origin, one freeway link, and one destination, as in Figure 2 with $N_1 = 12$. The mainstream origin O_1 has two lanes with a capacity of 2000 veh/h each. The motorway link L_1 follows with two lanes, and is 12 km long consisting of twelve segments of 1 km each. Segments 1 up to 5 and 12 are uncontrolled, segments 6 up to 11 are equipped with a variable message sign where speed limits can be set. The choice for the five uncontrolled upstream segments was made to be sure that boundary condition of equation (5) does not play a dominant role. Link L_1 ends in destination D_1. We use the same network parameters as in Kotsialos et al. (1999): $T = 10\,\text{s}$, $\tau = 18\,\text{s}$, $\kappa = 40\,\text{veh/lane/km}$, $\rho_{\max} = 180\,\text{veh/lane/km}$, $\rho_{\text{crit}} = 33.5\,\text{veh/lane/km}$, $a_m = 1.867$ and $v_{\text{free}} = 102\,\text{km/h}$.

Furthermore, we take $\nu_{\text{high}} = 65\,\text{km}^2\text{/h}$, $\nu_{\text{low}} = 30\,\text{km}^2\text{/h}$, and $a_{\text{speed}} = 2$. For the variable speed limits we have assumed that they can change only every minute, and that they cannot be less than $v_{\text{ctrlmin}} = 50$ km/h. This is imposed as a hard constraint in the optimization problem. If there is a safety constraint then $v_{\text{maxdiff}} = 10$ 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 backpropagating wave in the segments (see Figures 3 4).

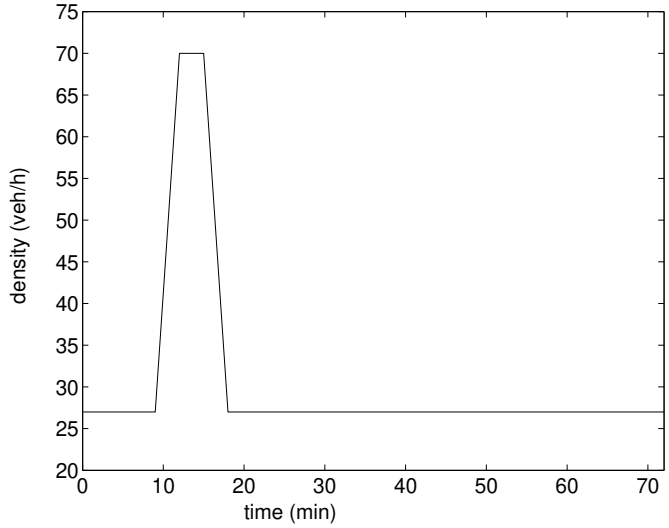

Figure 3: The downstream density scenario considered in the experiments.

For the above scenario the tuning of N_{p} and N_{c} will be demonstrated, and the performance (TTS) of the continuous and discrete-valued controls with or without safety

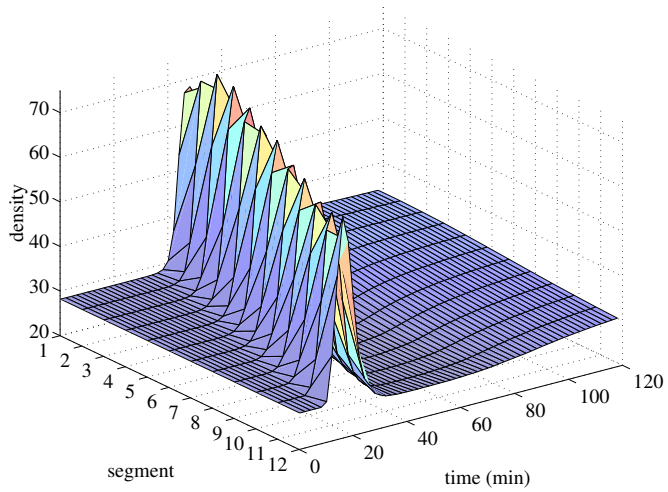

Figure 4: The shock wave propagates through the link in the no control case.

constraints are examined. In the discrete control case the control values $v_{\mathrm{ctrl},m,i}$ are in the set $\{50, 60, 70, 80, 90, 100, 110\}$.

The solution of the continuous-valued speed control problem is calculated by the Matlab implementation of the SQP (sequential quadratic programming) algorithm "fmincon". The discrete-valued control signals the discrete-valued signal is a rounded version of the continuous optimization result. Three different types of discretization are examined: The first (round) rounds the continuous control values to the nearest discrete value, the second (ceil) to the nearest discrete value that is higher than the continuous value, and the third (floor) to the nearest discrete value that is lower than the continuous 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, simulated annealing or genetic algorithms, but since for this set-up and input the discretization method results in comparable performance as the continuous version it is not necessary to do so.

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 way of rounding is not the same as rounding the continuous signal of the whole prediction horizon at once, because here the different traffic behavior caused by the discretization is already taken into account in the next MPC iteration.

The improvements of the discrete-valued control are compared to the improvement achieved by the continuous valued control case without constraints, and the effect of introducing the safety constraints is examined.

4.2 Results

The results of the simulations of the no control and the control with continuous speed limits without constraints are displayed in Figures 4 and 5. In the controlled case the shock wave disappears after approximately 2 hours, 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, segment 11 has higher values than the critical speed and is not effective as

argued before (see Figure 6). The active speed limits start to limit the flow at $t = 4$ min and create a low density wave traveling downstream (the small dip in Figure 5). 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.

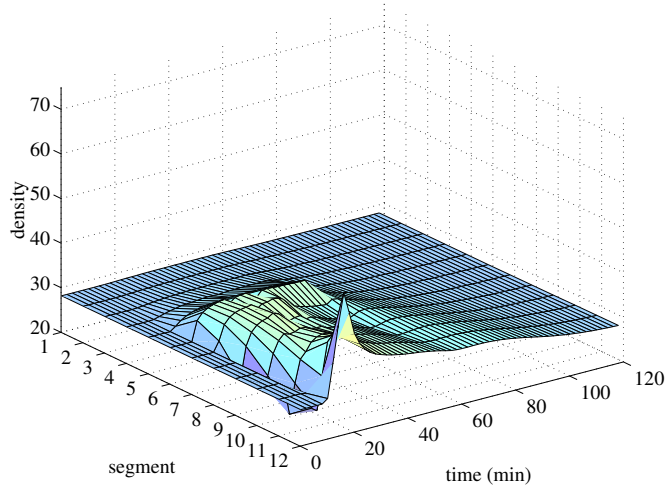

Figure 5: In the coordinated control case the shock wave disappears after approximately 2 hours (bottom).

The speed limits persist until the shock wave (to be precise, the high density region) is completely dissolved. The speed limits in Figure 6 start to increase after $t = 17$ min and return gradually to a high value that is not limiting the flow anymore.

The TTS was 1862.0 veh.hours in the no control case and 1458.0 veh.hours in the controlled (continuous, unconstrained) case, which is an improvement of 21.7 %.

The relative improvement of the performance as function of N_{p} and N_{c} is shown in Figure 7. The performance depends stronger on N_{p}, but for $N_{\mathrm{p}} \geq 10$ min (which is somewhat larger than the maximum travel time from segment 6 to the exit as argued in Section 3.5) the graphs become nearly flat. We chose for further analysis $N_{\mathrm{p}} = 11$ and $N_{\mathrm{c}} = 8$.

The result of the several types of discretization is shown in Table 1. The performance loss caused by the discretized speed limits is small in the "round" and "ceil" cases, but large for "floor". The explanation of the performance degradation in case of "floor" is topic for future research. The inclusion of the safety constraints, the results are comparable to Table 1, which are not shown here. The performance improvement for $N_{\mathrm{p}} = 11$, $N_{\mathrm{c}} = 8$ in the constrained case is 21.4 %, compared to 21.7 % in the unconstrained case. Figure 8 shows the values of the optimal speed limits discrete (ceil) case with safety constraints and $N_{\mathrm{p}} = 11$, $N_{\mathrm{c}} = 8$.

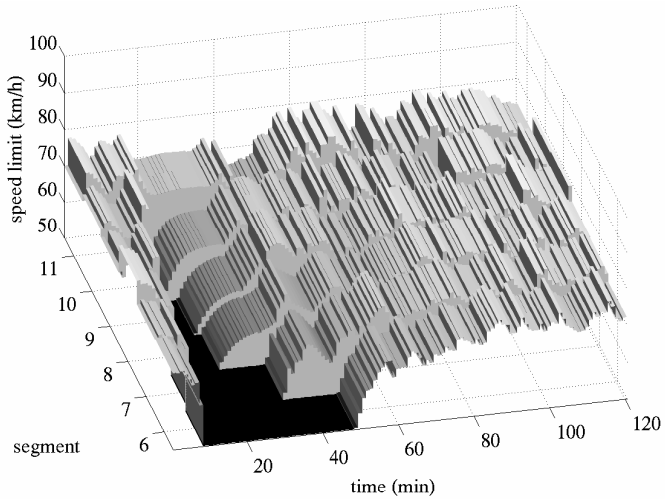

Figure 6: The speed for the continuous case without safety constraints and $N_{\mathrm{p}} = 11$, $N_{\mathrm{c}} = 8$ (top). For the purpose of visibility the direction of travel is opposite to Figure 4.

Table 1: The relative improvement of the performance (Total Time Spent) for several combinations of N_{p} and N_{c}, and for the continuous-valued speed limits and the three discrete-valued speed limits: round, ceil, and floor; without safety constraints.

Horizon		Relative improvement (%)			
N_{p}	N_{c}	continuous	round	ceil	floor
9	4	21.1	20.6	21.5	1.4
9	6	20.9	18.3	21.3	5.8
9	8	21.1	15.9	21.6	10.1
11	4	21.5	20.1	21.5	1.2
11	6	21.6	21.0	21.7	2.2
11	8	21.7	21.5	21.7	6.6
12	4	21.6	18.8	21.6	-0.1
12	6	21.7	21.5	21.7	2.0
12	8	21.7	21.5	21.7	12.7

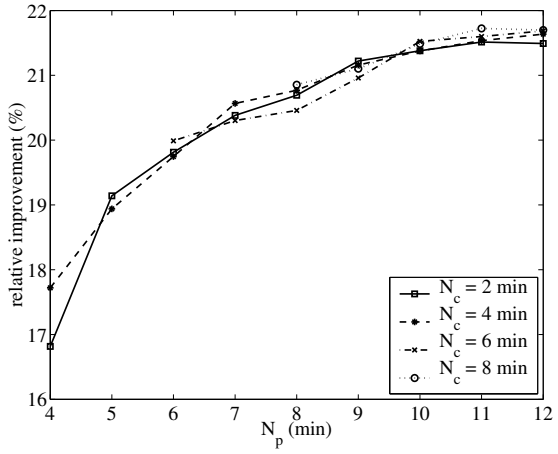

Figure 7: The relative improvement of the performance (Total Time Spent) in the continuous-valued, unconstrained case compared to the no control case as a function of N_{p} for several values of N_{c}. The sensitivity to N_{p} is much higher than to N_{c}.

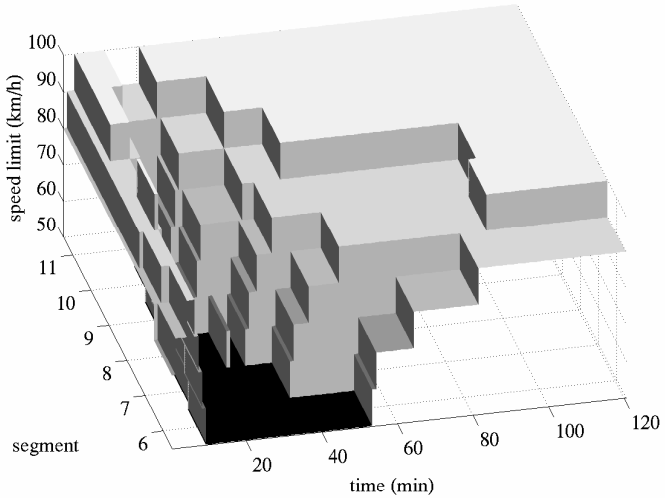

Figure 8: The speed limits for the discrete (ceil) case with safety constraints and $N_{\mathrm{p}} = 11$, $N_{\mathrm{c}} = 8$. For the purpose of visibility the direction of travel is opposite to Figure 4.

5 Conclusions and future research

We have applied model predictive control to optimally coordinate variable speed limits. The purpose of the control was to find the control signals that minimize the total time that vehicles spend in the network. We have applied the developed control framework to a benchmark network consisting of a link of 12 km, where 6 links are controlled by speed limits. It was shown that coordinated control with continuous-valued speed limits (base case) is effective against shock waves. The performance loss caused by discrete-valued speed limits and the inclusion of safety constraints was examined. The performance of the discrete-valued, safety constrained speed limits was comparable with the base case if the discrete-valued speed limits are generated by "round" or "ceil". In all of these cases the coordination of speed limits eliminated the shock wave entering from the downstream end of the link. The coordinated case resulted in a network where the outflow was sooner restored to capacity, and in a decrease of the total time spent with 21 %.

Topics for further research include: explanation of the performance degradation in case of "floor" discretization; comparison of the discrete MPC approach with other existing approaches; further examination of the trade-off between efficiency and optimality for rounding versus full discrete optimization; simulation other set-ups and scenarios; selecting other methods to model the effect of a speed limit; validating the new modeling assumptions regarding the speed limits and the mains stream on-ramp; further investigation of the effectiveness of MPC for optimal coordination of speed limits for a wider range of scenarios, networks, traffic flow models and/or model parameters; explicit inclusion of modeling errors and disturbances. Furthermore, including extra control measures in addition to speed limits (such as ramp metering, dynamic lane assignment, route information, reversible lanes, etc.) is also a topic for future research.

Acknowledgements

This research was supported by the Traffic Research Centre (AVV) of the Dutch Ministry of Transport, Public Works and Water Management.

The author would like to thank his thesis supervisors Prof.dr.ir J. Hellendoorn and dr.ir. B. De Schutter for their interesting and useful comments and suggestions on this paper.

References

Alessandri, A., Febbraro, A. D., Ferrara, A., and Punta, E. (1998). Optimal control of freeways via speed signalling and ramp metering. *Control Engineering Practice*, 6:771–780.

Alessandri, A., Febbraro, A. D., Ferrara, A., and Punta, E. (1999). Nonlinear optimization for freeway control using variable-speed signaling. *IEEE Transactions on Vehicular Technology*, 48(6):2042–2052.

Breton, P., Hegyi, A., Schutter, B. D., and Hellendoorn, H. (2002). Shock wave elimination/reduction by optimal coordination of variable speed limits. In *Proceedings of the 2002 IEEE Intelligent Transportation Systems*. Singapore. Accepted for publication.

Camacho, E. and Bordons, C. (1995). *Model Predictive Control in the Process Industry*. Springer-Verlag, Berlin, Germany.

Chien, C.-C., Zhang, Y., and Ioannou, P. A. (1997). Traffic density control for automated highway systems. *Automatica*, 33(7):1273–1285.

Febbraro, A. D., Parisini, T., Sacone, S., and Zoppoli, R. (2001). Neural approximations for feedback optimal control of freeway systems. *IEEE Transactions on Vehicular Technology*, 50(1):302–312.

Hoogen, E. v. d. and Smulders, S. (1994). 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, pages 145–149. London, England.

Kerner, B. S. and Rehborn, H. (1996). Experimental features and characteristics of traffic jams. *Physical Review E*, 53(2):R1297–R1300.

Kotsialos, A., Papageorgiou, M., and Messmer, A. (1999). Optimal coordinated and integrated motorway network traffic control. In *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT)*, pages 621–644. Jerusalem, Israel.

Kotsialos, A., Papageorgiou, M., and Middelham, F. (2001). Optimal coordinated ramp metering with advanced motorway optimal control. In *Proceedings of the 80th Annual Meeting of the Transportation Research Board*, number 01-3125. Washington, D.C.

Lenz, H., Sollacher, R., and Lang, M. (1999). Nonlinear speed-control for a continuum theory of traffic flow. In *14th World Congress of IFAC*, number Q-8e-10-2, pages 67–72.

Lenz, H., Sollacher, R., and Lang, M. (2001). Standing waves and the influence of speed limits. In *Proceedings of the European Control Conference 2001*, pages 1228–1232. Porto, Portugal.

Lighthill, M. J. and Whitham, G. B. (1955). On kinematic waves, II. a theory of traffic flow on long crowded roads. *Proc. of the Royal Society*, 229A.

Maciejowski, J. (2002). *Predictive Control with Constraints*. Prentice Hall, Harlow, England. ISBN 0-201-39823-0.

Papageorgiou, M. (1983). *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*, volume 50 of *Lecture Notes in Control and Information Sciences*. Springer Verlag, Berlin.

Papageorgiou, M., Blosseville, J., and Hadj-Salem, H. (1998). La fluidification des rocades de l'Ile de France: Un projet d'importance. Technical report, Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania, Greece. Internal Report No. 1998-17.

Papageorgiou, M., Blosseville, J.-M., and Hadj-Salem, H. (1990a). Modelling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris: Part I: Modelling. *Transportation Research Part A*, 24A(5):345–359.

Papageorgiou, M., Blosseville, J.-M., and Haj-Salem, H. (1990b). 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*, 24A(5):361–370.

Smulders, S. (1990). Control of freeway traffic flow by variable speed signs. *Transportation Research Part B*, 24B(2):111–132.

Smulders, S. (1996). *Control of Freeway Traffic Flow*. CWI Tract no. 80. CWI (Dutch institute for research in Mathematics and Computer Science). ISBN: 90-6196-451-2.