Technical report CSE02-015

Optimal coordination of variable speed limits to suppress shock waves[*]

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

Optimal Coordination of Variable Speed Limits to Suppress Shock Waves

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

Delft Center for Systems and Control, Delft University of Technology
Mekelweg 2, 2628 CD Delft, The Netherlands
phone: +31-15-278.20.87, fax: +31-15-278.66.79
email: {a.hegyi,b.deschutter,j.hellendoorn}@dcsc.tudelft.nl

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, which 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 even eliminate shock waves. To predict the evolution of the traffic flows in the network, which MPC requires, 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 hence 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.

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, it has many advantages to consider the effect of the measures on the network level instead. 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 cannot be observed instantaneously, so the prediction 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. 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 [16] 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 (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), 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. For this reason and because there are always (unpredictable) disturbances present in a traffic network, *feedback control* 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 (for practical implementation) that the driver should not encounter a decrease in the displayed 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 the literature, basically two views on the use of speed limits can be found. The first emphasizes the homogenizing effect, see [1, 2, 20, 18, 19], whereas the second is more focused on the preventing traffic breakdown by reducing the flow by means of speed limits, see [5, 10, 11]. 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 homogenizing 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).

The traffic breakdown prevention approach focuses more on preventing too high densities, and also allows speed limits that are lower than the critical speed. The results in [20] 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 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 [14], sliding-mode control [10, 11], and optimal control [1, 2]. In this paper we demonstrate the feasibility of on-line optimization. The advantage of this approach is that it is able to adapt to changing traffic conditions.

Most of the models used in literature represent the speed limits by a factor that downscales the fundamental diagram, see for example [1, 2, 10]. This can give too optimistic results (see Section 3.2.2),
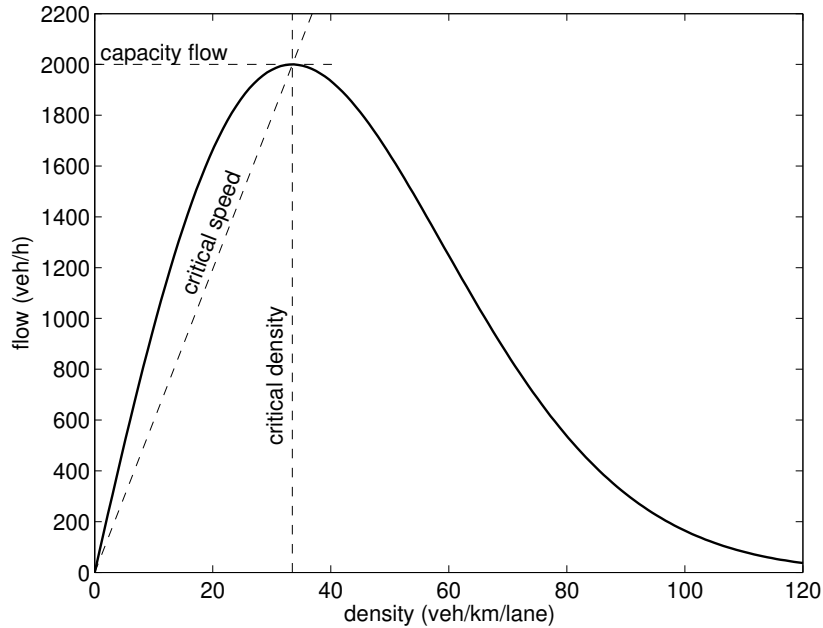
Figure 1: A typical example of the fundamental diagram. The critical speed is the speed that corresponds to maximum flow. The slope of the line connecting the origin and a point on the fundamental diagram represents the speed corresponding to that point.

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, see [3], 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 this paper is as follows. In Section 2 the problem and the basic idea of the solution to 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., [7]) that some types 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 [12] 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 will 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 that 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 the stable, metastable, and unstable traffic flow states observed by Kerner and Rheborn [7]. 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 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.) From the resulting optimal control signal only the first sample $k+1$ is applied to the process. In the next time step $k+1$, a new optimization is performed (with a prediction horizon that is shifted one time step further) and of the resulting control signal again only the first sample is applied, and so on. This scheme, called rolling horizon, allows for updating the state (from measurements), or even for updating the model in every iteration step. Updating the state results in a controller that has a low sensitivity to prediction errors, and updating the model results in an adaptive control system, which could be useful in situations where the model significantly changes, such as in case of incidents or changing weather conditions. For more information on MPC see [4, 13] 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, see [8, 17]. The modifications are introduced to model shock waves better and to include the effect of speed limits. Note that the MPC approach is generic and will find the optimal speed limits independently of the model that is used (e.g., independently of the way in which 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).
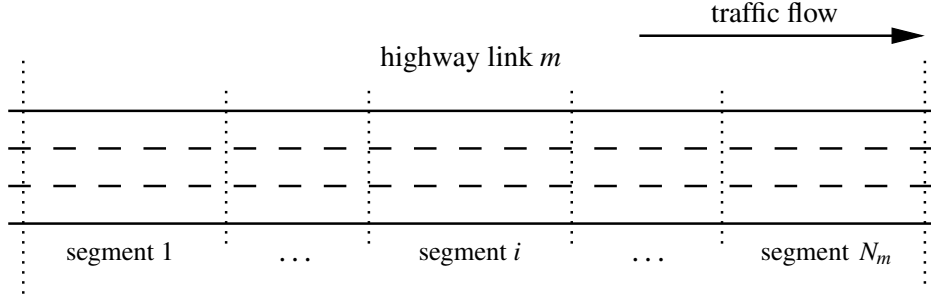
Figure 2: In the METANET model, a motorway link is divided into segments.

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 $T = 10\,\text{s}$).

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):

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

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):

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

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:

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

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

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

with a_m a dimensionless 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 $d_o(k)$ (just as in [8, 9, 15], we assume that the demand is independent of any control actions taken in the network), minus the outflow $q_o(k)$:

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

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.

3.2.2 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 a mainstream origin, which has a different nature than an on-ramp origin. The third extension describes the different effects of a positive or negative downstream density gradient on the speed (cf. the anticipation term in 2).

In some publications, the effect of the speed limit is expressed by scaling down the desired speed-density diagram; see [1, 2, 10]. 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 (see Figure 1). 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 target speed is the minimum of the following two quantities: the target speed based on the experienced density, and the target speed caused by the speed limit displayed on the variable message sign (VMS):

$$V\left(\rho_{m,i}(k), v_{\text{ctrl},m,i}(k)\right) = \min\left((1+\alpha)v_{\text{ctrl},m,i}(k), v_{\text{free},m}\exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_{\text{crit},m}}\right)^{a_m}\right]\right) , \tag{5}$$

where $v_{\text{ctrl},m,i}(k)$ is the speed limit imposed on segment i, link m, at time k, and $(1+\alpha)$ expresses the non-compliance, i.e., the factor that the target speed is higher (lower) than the displayed speed limit. This is due to the fact that some drivers may want to drive faster (or slower) than the indicated speed limit. The parameter α can be determined from empirical data. For segments where a speed limit is present (5) is used in (2) instead of (3).

To express the different natures of a mainstream origin link o and 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 in 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 in 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:

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

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 the parameter ν. In (2), ν is a global parameter and has the same value for all segments. However, here we take different values for $v_{m,i}(k)$, depending on whether the downstream density is higher or lower than the density in the actual segment:

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

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:

$$v_{m,0}(k) = v_{m,1}(k) \ . \tag{6}$$

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 in 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 in 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 for the three situations are formulated as follows:

$$v_{\mathrm{ctrl},m,i}(l-1) - v_{\mathrm{ctrl},m,i}(l) \leq v_{\mathrm{maxdiff}} \quad \text{for all } (m,i,l) \text{ such that}$$
$$(m,i) \in I_{\mathrm{speed}} \text{ and}$$
$$l \in \{k,\ldots,k+N_{\mathrm{c}}-1\},$$

$$v_{\mathrm{ctrl},m,i}(l) - v_{\mathrm{ctrl},m,i+1}(l) \leq v_{\mathrm{maxdiff}} \quad \text{for all } (m,i,l) \text{ such that}$$
$$(m,i) \in I_{\mathrm{speed}} \text{ and}$$
$$(m,i+1) \in I_{\mathrm{speed}} \text{ and}$$
$$l \in \{k,\ldots,k+N_{\mathrm{c}}-1\},$$

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

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

$$v_{\mathrm{ctrl},m,i}(l) \geq v_{\mathrm{ctrlmin}} \quad \text{for all } (m,i) \in I_{\mathrm{speed}} \text{ and } l \in \{k,\ldots,k+N_{\mathrm{c}}-1\}.$$

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 [13]). However, these rules cannot be straightforwardly applied to the traffic flow control framework presented above.

The prediction horizon N_{p} should be larger than the maximum travel time between the control inputs and the exit (under the presence of a shock wave), because the vehicles that are influenced by the current control measure only have an effect on the network performance when they exit the network. Furthermore, a control action may still 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. 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 the benchmark set-up consisting of a motorway link equipped with variable speed signs.
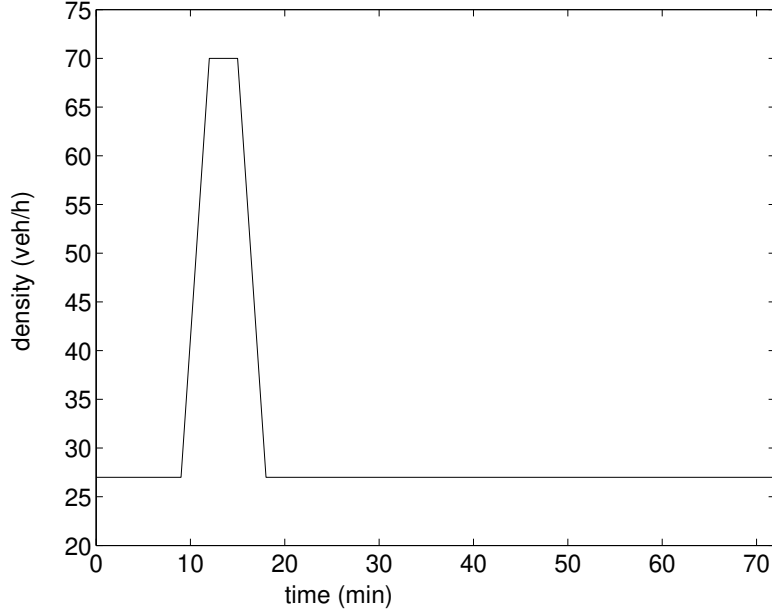
Figure 3: The downstream density scenario considered in the experiments.

4.1 Set-up

The benchmark set-up consists of one origin, one motorway 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 segment 12 are uncontrolled, while segments 6 up to 11 are equipped with a variable message sign where speed limits can be set. We choose to include the five uncontrolled upstream segments to be sure that boundary condition of equation (6) does not play a dominant role. Link L_1 ends in destination D_1. We use the same network parameters as in [8]: $T = 10\,\text{s}$, $\tau = 18\,\text{s}$, $\kappa = 40\,\text{veh/lane/km}$, $\rho_{\text{max}} = 180\,\text{veh/lane/km}$, $\rho_{\text{crit}} = 33.5\,\text{veh/lane/km}$, $a_m = 1.867$ and $v_{\text{free}} = 102\,\text{km/h}$.

Furthermore, we take $v_{\text{high}} = 65\,\text{km}^2/\text{h}$, $v_{\text{low}} = 30\,\text{km}^2/\text{h}$, $\alpha = 0.05$ 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 back-propagating wave in the segments, see Figures 3 and 4. It is assumed that the upstream demand and downstream density is known, or predicted by an external algorithm. 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 continuous and discrete-valued controls with or without safety constraints are examined. In the discrete control case, the control values $v_{\text{ctrl},m,i}$ are in the set $\{50, 60, \ldots, 110\}$.

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
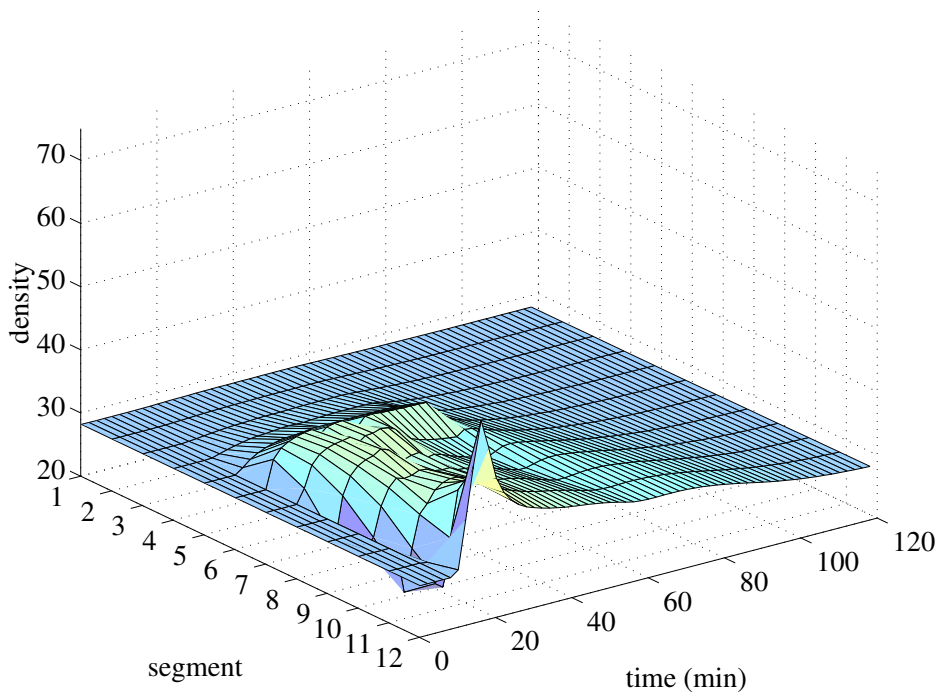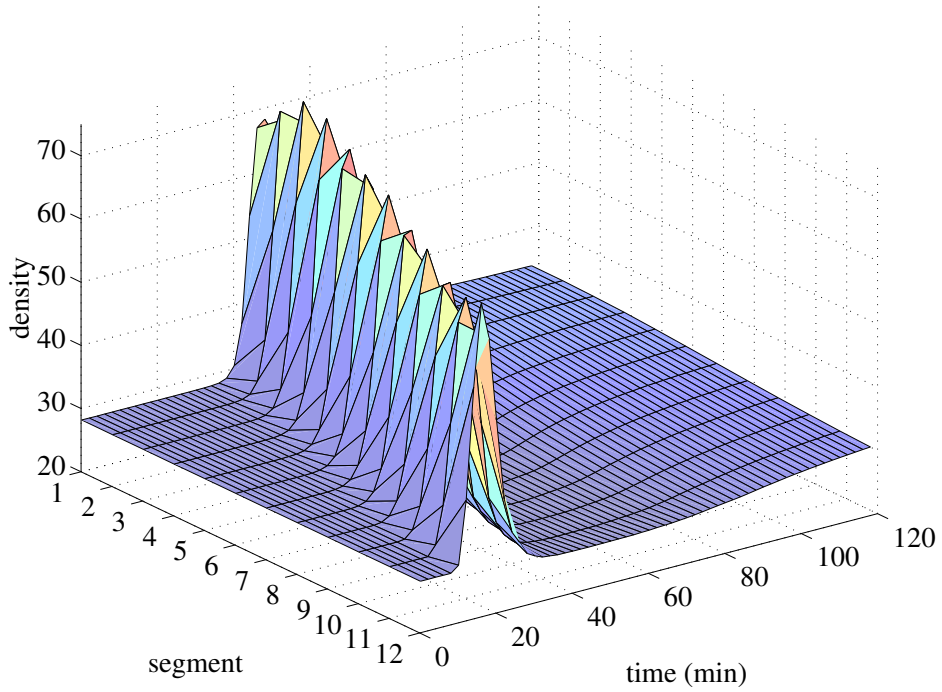
Figure 4: The shock wave propagates through the link in the no control case (top). In the coordinated control case, the shock wave disappears after approximately 2 hours (bottom).

control 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) rounds them 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 a performance that is comparable to that of 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 and applying of 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 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 case with continuous speed limits without constraints are displayed in Figure 4. 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; the speed limit in segment 11 has higher values than the critical speed and is not limiting the flow (see Figure 5). 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 4). 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 5 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 a function of N_p and N_c is shown in Figure 6. 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 $N_\mathrm{p} = 11$ min and $N_\mathrm{c} = 8$ min.

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 performance degradation in case of "floor" can be explained by the slow dynamics of the traffic process. Initially, the flow is correctly restricted by low speed limits (here 50 km/h). However, when the congestion starts resolving, the (optimized, continuous-valued) speed limits increase to just above the natural evolution of the speed (such that they are not limiting anymore) but by "floor" this value is rounded downward. By the slow dynamics of the traffic process the speed usually does 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. For more information see [6].
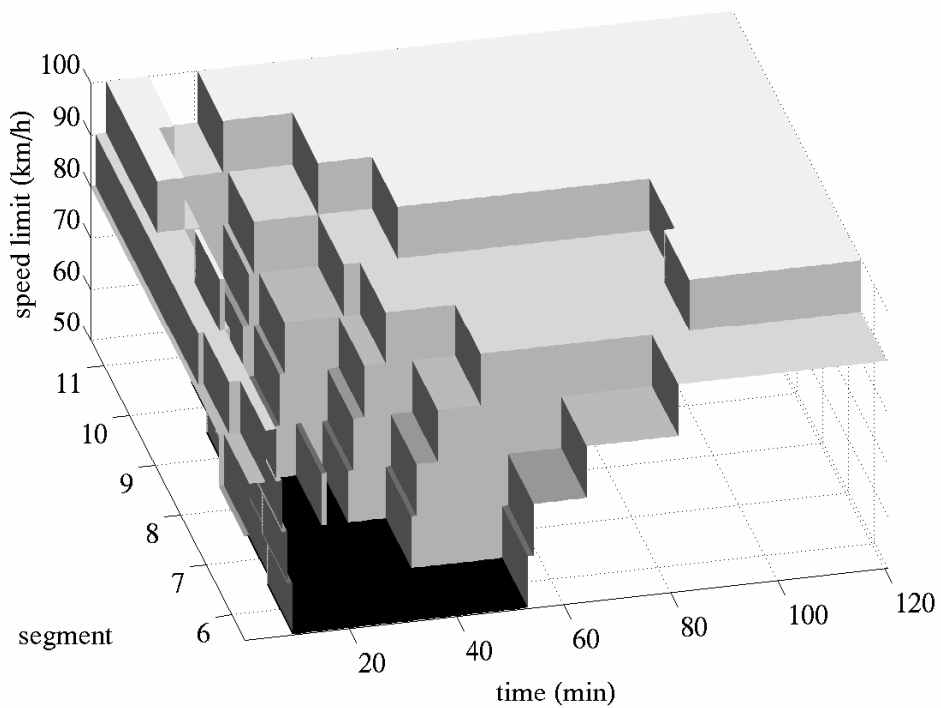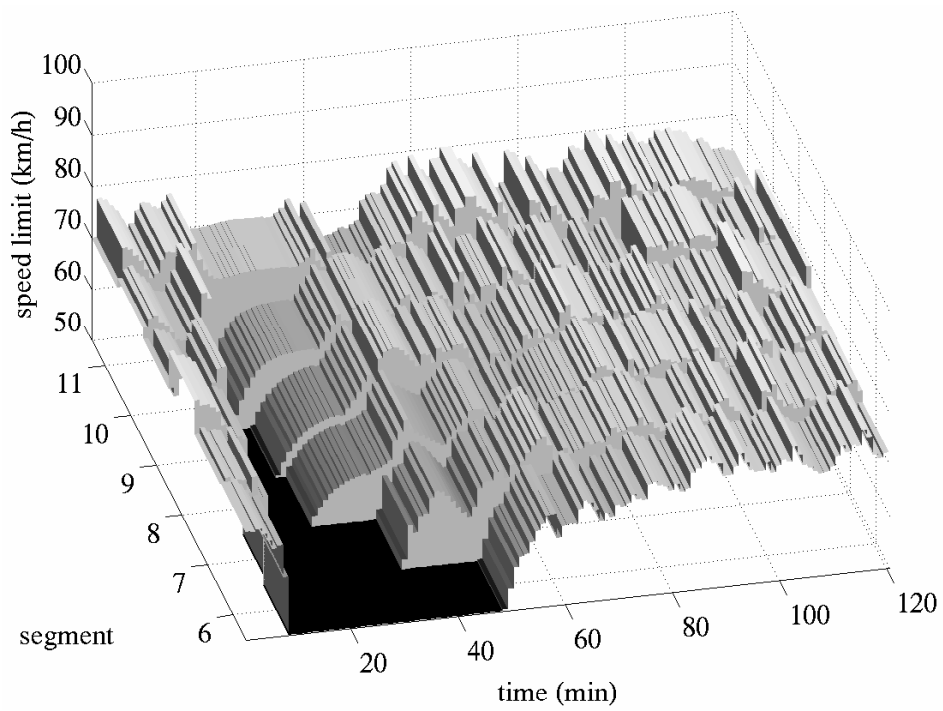
11

Figure 5: The speed for the continuous case without safety constraints and $N_p = 11$, $N_c = 8$ (top). The speed limits for the discrete (ceil) case with safety constraints and $N_p = 11$, $N_c = 8$. For the purpose of visibility, the travel direction is opposite to that in Figure 4.
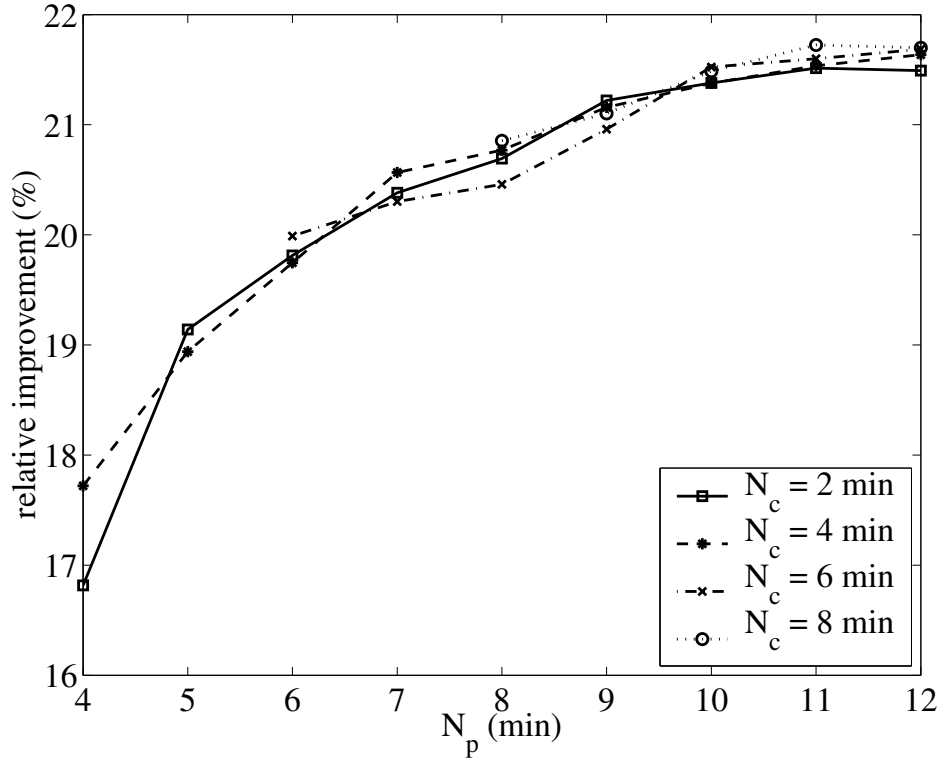
Figure 6: The relative improvement of the performance (Total Time Spent) in the continuous-valued, unconstrained case compared to the no-control case as a function of N_p for several values of N_c. The sensitivity to N_p is much higher than that to N_c.

		Relative improvement (%)							
Horizon		unconstrained				constrained			
N_p	N_c	continuous	round	ceil	floor	continuous	round	ceil	floor
9.0	4.0	21.1	20.2	21.5	1.9	20.9	19.4	18.5	3.9
9.0	6.0	20.9	21.1	21.4	3.8	21.1	20.2	21.1	15.1
9.0	8.0	21.1	15.9	21.4	12.1	21.1	20.7	20.4	15.1
10.0	4.0	21.4	20.1	21.5	0.5	21.2	20.0	21.2	0.2
10.0	6.0	21.6	20.5	21.7	14.2	21.4	20.9	21.4	12.7
10.0	8.0	21.5	21.1	21.7	5.2	21.4	20.8	21.4	16.3
11.0	4.0	21.5	19.8	21.5	-2.8	21.3	19.5	21.3	4.2
11.0	6.0	21.6	21.1	21.7	3.7	21.4	21.1	21.4	14.2
11.0	8.0	21.7	21.1	21.7	8.1	21.5	21.0	21.7	13.0
12.0	4.0	21.6	20.3	21.6	-1.1	21.5	19.0	21.5	-0.3
12.0	6.0	21.7	21.3	21.8	7.3	21.5	21.4	21.4	14.2
12.0	8.0	21.7	21.6	21.8	8.9	21.5	21.5	21.4	15.0

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.

The results of including the safety constraints are comparable to the results without safety constraints, see Table 1. Figure 5 shows the values of the optimal speed limits discrete (ceil) case with safety constraints and $N_{\mathrm{p}} = 11$ min, $N_{\mathrm{c}} = 8$ min.

Finally, the computation time varied between 3 and 25 min, which is at least four times faster than real time. It is expected that the computation time varies linearly with the number of segments and the length of the prediction horizon, and exponentially with the number of control inputs and the control horizon.

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 segments of 1km 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 to that of the base case if the discrete-valued speed limits were 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 controlled case resulted in a network where the outflow was sooner restored to capacity, and in a decrease of the total time spent of 21 %.

Topics for further research include: 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; study of a real motorway stretch, including model calibration with real data, and the examination of the necessity of on-line calibration; simulation of other set-ups and scenarios; selection of other methods to model the effect of a speed limit; validation of the new modeling assumptions regarding the speed limits and the main stream origin; 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 unpredicted disturbances (demands); further study of the capacity drop and metastability phenomena; inclusion of extra control measures in addition to speed limits (such as ramp metering, dynamic lane assignment, route information, reversible lanes, etc.).

Acknowledgments

References

[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, no. 6, pp. 771–780, June 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] 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 (ITSC'02)*, Singapore, pp. 225–230, Sept. 2002.

[4] E.F. Camacho and C. Bordons, *Model Predictive Control in the Process Industry*. Berlin, Germany: Springer-Verlag, 1995.

[5] C.C. Chien, Y. Zhang, and P.A. Ioannou, "Traffic density control for automated highway systems," *Automatica*, vol. 33, no. 7, pp. 1273–1285, July 1997.

[6] 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.

[7] B.S. Kerner and H. Rehborn, "Experimental features and characteristics of traffic jams," *Physical Review E*, vol. 53, no. 2, pp. R1297–R1300, Feb. 1996.

[8] A. Kotsialos, M. Papageorgiou, and A. Messmer, "Optimal coordinated and integrated motorway network traffic control," in *Proceedings of the 14th International Symposium of Transportation and Traffic Theory (ISTTT)*, Jerusalem, Israel, pp. 621–644, July 1999.

[9] 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*, Washington, DC, 2001. Paper no. 01-3125.

[10] H. Lenz, R. Sollacher, and M. Lang, "Nonlinear speed-control for a continuum theory of traffic flow," in *Proceedings of the 14th IFAC World Congress (IFAC'99)*, vol. Q, Beijing, China, pp. 67–72, Jan. 1999.

[11] H. Lenz, R. Sollacher, and M. Lang, "Standing waves and the influence of speed limits," in *Proceedings of the European Control Conference 2001 (ECC'01)*, Porto, Portugal, pp. 1228–1232, Sept. 2001.

[12] M.J. Lighthill and G.B. Whitham, "On kinematic waves: II. A theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London*, vol. 299A, pp. 317–345, May 1955.

[13] J.M. Maciejowski, *Predictive Control with Constraints*. Harlow, UK: Prentice Hall, 2002.

[14] M. Papageorgiou, *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*. Lecture Notes in Control and Information Sciences, Berlin, Germany: Springer Verlag, 1983.

[15] 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. 24, no. 5, pp. 345–359, Sept. 1990.

[16] M. Papageorgiou, J.M. Blosseville, and H. Hadj-Salem, "La fluidification des rocades de l'Ile de France: Un projet d'importance," Tech. rep. No. 1998-17, Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania, Greece, 1998.

[17] 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. 24, no. 5, pp. 361–370, Sept. 1990.

[18] S. Smulders, "Control of freeway traffic flow by variable speed signs," *Transportation Research Part B*, vol. 24, no. 2, pp. 111–132, Apr. 1990.

[19] S. Smulders, *Control of Freeway Traffic Flow*. Amsterdam, The Netherlands: CWI (Dutch Institute for Research in Mathematics and Computer Science), 1996. CWI Tract no. 80.

[20] E. van den Hoogen and S. Smulders, "Control by variable speed signs: Results of the Dutch experiment," in *Proceedings of the 7th International Conference on Road Traffic Monitoring and Control*, London, UK, pp. 145–149, Apr. 1994.

Optimal Coordination of Variable Speed Limits to Suppress Shock Waves — Addendum[*]

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

Delft Center for Systems and Control, Delft University of Technology
Mekelweg 2, 2628 CD Delft, The Netherlands
phone: +31-15-278.20.87, fax: +31-15-278.66.79
email: {a.hegyi,b.deschutter,j.hellendoorn}@dcsc.tudelft.nl

Abstract

The purpose of this addendum is to give a more detailed explanation of the performance degradation as described in the paper "Optimal coordination of variable speed limits to suppress shock waves" (by A. Hegyi, B. De Schutter, and J. Hellendoorn, *Transportation Research Record*, no. 1852, pp. 167–174, 2004). For the simulation settings in that paper 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-valued speed limit step size is increased, the degradation occurs in all cases, but for the floor type of rounding first.

All references in this addendum that are not preceded by a capital letter A refer to tables, references, etc. of the paper [A1].

In [A1] the performance for the floor type of rounding is poor compared to the other two types of rounding for both the constrained and unconstrained cases (see Table 1). Therefore, in the remainder of this report we will consider unconstrained cases only.

A   Cause of performance degradation

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-valued 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, continuous-valued) speed limits will increase to enable the traffic to accelerate. These speed limits will be just above[1] the natural evolution[2] 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 A.1 a snapshot of the MPC procedure for speed limits rounded with floor is shown (with $N_p = 11$ min, $N_c = 8$ min). For visibility purposes we show the speed limits for one segment

---

[*]Note that this addendum is not a part of the published journal paper [A1]. However, it is available as a separate technical report [6].

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

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

ΔSL (km/h)	Relative improvement (%)		
	round	ceil	floor
0 (continuous)	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 A.1: The relative improvement of the performance (Total Time Spent) for several speed limit step sizes (ΔSL) for the continuous-valued speed limits and the three discrete-valued speed limits: round, ceil, and floor

only. The left vertical (green) 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-valued speed limits the signals between the two vertical lines are approximated by the discrete signals.

Figure A.2 is a zoom-in of Figure A.1, where we can see that the optimal continuous speed limit value in segment 11 at the current time (see the left vertical green line; this speed limit equals approximately 52 km/h) is higher than one 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-valued speed limits is reduced.

We compare several speed limit step sizes with $N_{\mathrm{p}} = 11$ min and $N_{\mathrm{c}} = 8$ min, see Table A.1. 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[3].

Additional references

[A1] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Optimal coordination of variable speed limits to suppress shock waves," *Transportation Research Record*, no. 1852, pp. 167–174, 2004.

---

[3]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.
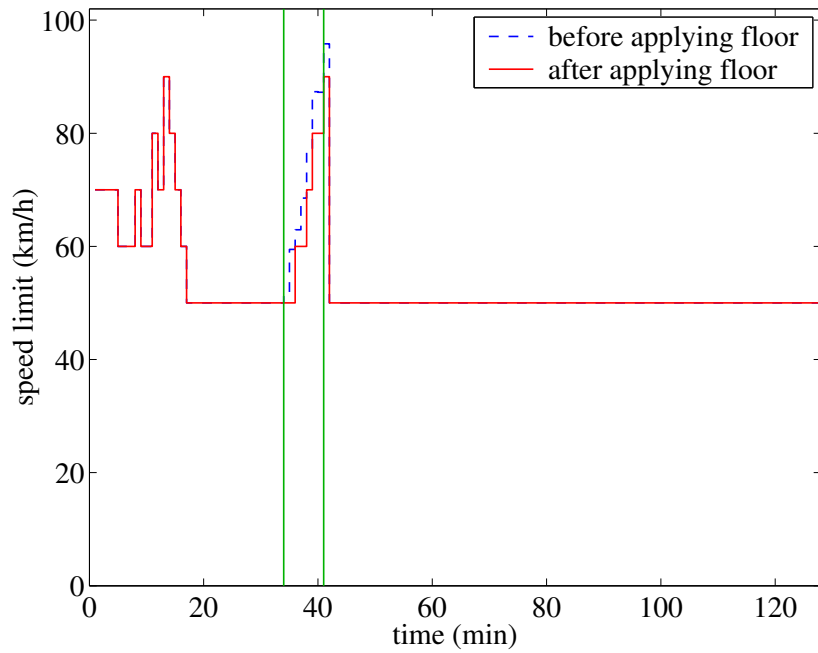
Figure A.1: Snapshot of the speed limit (segment 11) resulting from the MPC procedure rounded with floor. The signal is shown before and after rounding. The left and right vertical (green) lines indicate respectively the current time instant and the end of the prediction horizon.
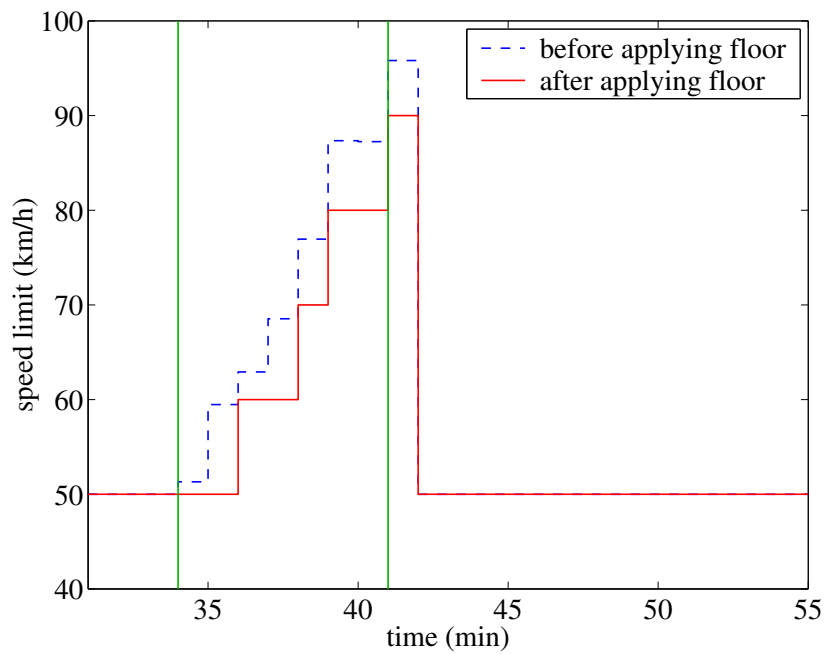

Figure A.2: Zoom-in of Figure A.1