**Integrated Ramp Metering and Variable Speed Limit Control of Motorway Traffic Flow**

Ioannis Papamichail*, Katerina Kampitaki*, Markos Papageorgiou* and Albert Messmer**

*Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania 73100, Greece
(Tel: +30-28210-37289; e-mails: ipapa @ dssl.tuc.gr; katerina @ dssl.tuc.gr; markos @ dssl.tuc.gr)

**Groebenseeweg 2, D-82402 Seeshaupt, Germany (Tel: +49-8801-95101; e-mail: albert.messmer @ t-online.de)

Abstract: The impact of variable speed limits (VSL) on the aggregate traffic flow behaviour is reflected in the quantitative model proposed in this paper. VSL are incorporated in a general second-order traffic flow model as an additional control component. The integrated motorway network traffic control problem is formulated as a constrained discrete-time optimal control problem which is solved very efficiently even for large-scale networks by a suitable feasible-direction algorithm. An illustrative example is presented under different control scenarios and it is shown that traffic flow efficiency can be substantially improved when VSL control measures are used, particularly in integration with coordinated ramp metering.

1. INTRODUCTION

In the last decades, motorways have become notorious sites of extensive traffic congestion, particularly in and around metropolitan areas. Traffic congestion degrades the available infrastructure in the sense of reducing the motorway throughput. Thus, the expensive motorway infrastructure is underutilised, ironically exactly at the times (peak hours) it is most urgently needed. The efficient, safe, and less polluting transportation of persons and goods on motorways calls for an optimal utilisation of the available infrastructure via suitable application of a variety of traffic control measures such as ramp metering, driver information, route guidance and variable speed limits (VSL). A number of approaches, including optimal control, expert systems, fuzzy systems, neural networks and feedback control, have been developed in the past for the design of related control strategies. In the present paper the optimal control approach is applied for optimal integrated ramp metering and VSL measures.

Ramp metering aims at improving the traffic conditions by appropriately regulating the inflow from the on-ramps to the motorway mainstream. As the ramp storage space may be limited, but also due to equity considerations, ramp metering should be applied at multiple ramps (coordinated ramp metering) for maximum efficiency. Despite the induced ramp queue delays, the higher motorway throughput (and reduced mainstream delays) due to congestion avoidance may lead to shorter average travel times. Coordinated ramp metering with the use of optimal control has been extensively studied in the past (see Papageorgiou and Kotsialos, 2002; Papamichail et al., 2007, for related overviews).

VSL installations are encountered in many countries around the world and their number is increasing at an accelerated pace. VSL are displayed on appropriate variable message signs in response to the prevailing traffic conditions.

The main impact of VSL on traffic flow is deemed to be the reduction of the mean speed at undercritical densities; and the homogenisation of speeds, i.e., reduction of speed differences among vehicles and of mean speed differences among lanes.

A main targeted result of VSL is enhanced traffic safety and, indeed, the selection of motorway stretches for VSL installations in several countries is guided by the frequency of registered accidents. Multi-year evaluations of the VSL impact on traffic safety indicate a reduction in accident numbers by as much as 20-30% after VSL installation. On the other hand, to the best of the authors' knowledge, there is no evaluation of the VSL impact of available installations that would demonstrate a consistent and measurable improvement of traffic flow efficiency, e.g., in the sense of reduced travel times.

The design of pertinent control strategies that may increase traffic flow efficiency, calls for a sufficiently accurate description of the VSL impact on the aggregate (macroscopic) traffic conditions. There were very few investigations in the past regarding the precise impact of VSL on aggregate traffic flow behaviour, e.g., on the fundamental diagram (flow-density curve). The results reported by Zackor (1991) and summarised by Zackor (1991) were the basis for Cremer (1979) to propose a quantitative model for the VSL-induced fundamental diagram change. However, it is quite likely that the increase of flow capacity suggested by Cremer (1979) is rather exaggerated. This VSL model was incorporated in a general dynamic model leading to an optimal control formulation (Cremer 1979; Alessandri et al., 1998). However, a heuristically fixed control law was eventually used, due to the size of the problem, and its parameters were optimised. In other, more recent, research work (Hegyi et al., 2003), the assumed VSL impact was to merely replace the left part of the fundamental diagram by a straight line with slope corresponding to the displayed VSL.
Recently, the effect of VSL on the aggregate traffic flow behaviour (in form of the flow-density diagram) was investigated by Papageorgiou et al. (2007) on the basis of traffic data from a European motorway where a flow/speed threshold-based VSL control algorithm is currently used. The main findings of this investigation were:

- Speed limits – when applied at undercritical densities – have the effect of decreasing the slope of the flow-density diagram. Moreover, the smaller the imposed speed limit, the larger the decrease in the slope of the flow-density diagram. This impact may be exploited in order to hold back traffic flow in order to retard the onset of congestion at downstream bottlenecks, as practiced, e.g., in Hegyi et al., 2003.

- The VSL-affect ed flow-density curve crosses (at least for some VSL) the non-VSL curve, shifting the critical density to higher values in the flow-density diagram. This impact may be exploited in order to hold more vehicles in the motorway without falling in congestion. It may sound paradoxical, but these cross points may imply that the mean speed at overcritical densities is higher when a speed limit is imposed; this may happen due to the homogenisation effects mentioned earlier.

- Regarding the potential increase of flow capacity, the data analysis was rather inconclusive, as a slight increase is indeed visible at some locations while at other locations no increase could be observed. In locations where VSL yield indeed a capacity increase, this may be exploited by a suitably designed control strategy for throughput increase.

The identified impact of VSL on the aggregate traffic flow behaviour is reflected in the quantitative model proposed in the present paper. VSL are incorporated in a general second-order traffic flow model (Section 2) as an additional control component leading to an accordingly extended optimal control formulation (Section 3). An illustrative example is discussed in Section 4 under different control scenarios and conclusions are summarised in Section 5.

2. TRAFFIC FLOW MODELLING

A macroscopic second-order traffic flow model is used in this study. The model was validated against real traffic data at several instances (Papageorgiou et al., 1990; Kotsialos et al., 2002a) and was found to reproduce the real traffic conditions with remarkable accuracy for all traffic conditions. The model is included in the METANET motorway traffic flow simulator (Messmer and Papageorgiou, 1990) and is extended here to incorporate VSL control measures.

The motorway network is represented by a directed graph whereby the links of the graph represent motorway stretches. Each motorway stretch has uniform character stics, i.e., no on-off-ramps and no major changes in geometry. The nodes of the graph are placed at locations where a major change in road geometry occurs, as well as at junctions, on-ramps and off-ramps.

The macroscopic description of traffic flow implies the definition of adequate variables expressing the aggregate behaviour of traffic at certain times and locations. The time and space arguments are discretised. The discrete time step is denoted by $T$ (typically $T = 10\text{s}$). A motorway link $m$ is divided into $N_m$ segments of equal length $L_m$ (typically $L_m = 500\text{m}$) (Fig. 1). The traffic in each segment $i$ of link $m$ at discrete time $t = kT$, $k = 0, 1, \ldots, K$, where $K$ is the time horizon, is macroscopically characterised via the following variables: the traffic density $\rho_{m,i}(k)$ (veh/km/lane) is the number of vehicles in segment $i$ of link $m$ at time $t = kT$ divided by $L_m$ and by the number of lanes $\lambda_m$; the mean speed $v_{m,i}(k)$ (km/h) is the mean speed of the vehicles included in segment $i$ of link $m$ at time $t = kT$; and the traffic volume or flow $q_{m,i}(k)$ (veh/h) is the number of vehicles leaving segment $i$ of link $m$ during the time period $[kT, (k+1)T]$, divided by $T$.

The previously defined traffic variables are calculated for each segment $i$ of link $m$ at each time step $k$ by the following equations:

$$\rho_{m,i}(k+1) = \rho_{m,i}(k) + \frac{T}{L_m} \left( q_{m,i-1}(k) - q_{m,i}(k) \right)$$  (1)

$$q_{m,i}(k) = \rho_{m,i}(k) v_{m,i}(k) \lambda_m$$  (2)

$$v_{m,i}(k+1) = v_{m,i}(k) + \frac{T}{\tau} \left[ v \left( \rho_{m,i}(k), b_{m,i}(k) \right) - v_{m,i}(k) \right] + \frac{T}{L_m} \left[ v_{m,i-1}(k) - v_{m,i}(k) \right] v_{m,i}(k)$$  (3)

$$V \left[ \rho_{m,i}(k), b_{m}(k) \right] = v_{f,m} \left[ b_{m}(k) \right] \exp \left[ -\frac{1}{\alpha_m \left[ b_{m}(k) \right]} \frac{\rho_{m,i}(k)}{\rho_{m,i}(k)} \exp \left[ -\frac{1}{\alpha_m \left[ b_{m}(k) \right]} \frac{\rho_{m,i}(k)}{\rho_{m,i}(k)} \right] \right]$$  (4)

$$v_{f,m} \left[ b_{m}(k) \right] = v_{f,m} \left[ b_{m}(k) \right]$$  (5)

$$\rho_{c,r,m} \left[ b_{m}(k) \right] = \rho_{c,r,m} \left[ 1 + 2A_n \left[ 1 - b_{m}(k) \right] \right]$$  (6)

$$\alpha_m \left[ b_{m}(k) \right] = \alpha_m \left[ E_n - (E_n - 1) b_{m}(k) \right]$$  (7)

where (1) is the conservation equation; (2) is the transport equation to be replaced in (1); (3) is an empirical dynamic mean speed equation where (4) must be replaced; $\tau$ (a time constant), $\nu$ (an anticipation constant), and $\kappa$ are model parameters which are equal for all the network links. Two further terms may be added to (3) for higher accuracy under certain conditions (Papageorgiou et al., 1990).

![Fig. 1. Discretised motorway link.](image-url)
The original (non-VSL) model includes three link-specific constant parameters in the speed-density curve (4): the free speed \( v^*_f,m \), encountered at zero density (\( \rho_m = 0 \)); the critical density \( \rho^{cr,m} \) at which traffic flow is close to capacity, \( q_{ap,m} \); and \( \alpha_m \). To incorporate the VSL impact, these parameters are now rendered \( b_m \)-dependent functions, where \( b_m(k) \in [b_{min,m},1] \) is the VSL rate at link \( m \) and period \( k \), i.e., a control variable. As (5) reveals, \( b_m \) is equal to the VSL-induced \( v^*_f,m \), divided by the non-VSL \( v^*_f,m \); or, approximately, equal to the displayed VSL divided by the legal speed limit without VSL. If \( b_m(k) = 1 \), no VSL is applied, else \( b_m(k) < 1 \), while \( b_{max,m} \) is a minimum admissible VSL rate. Equations (6)-(7) suggest that \( \rho^{cr,m} \) and \( \alpha_m \) are linear functions of \( b_m \), attaining their usual non-VSL values for \( b_m(k) = 1 \). The extended speed-density curve (4)-(7) was validated by use of traffic data (Kampitaki, 2008) and the related flow-density curves (resulting from a combination of (2) and (4) for specific VSL rates) are displayed in Fig. 2. Clearly, a capacity increase is observed at this specific location. More comprehensive investigations including locations without flow capacity increase are left for future work. All the other model parameter values used in this study are taken from a previous model calibration for a real motorway (Kotsialos et al., 2002a) and this, of course, may quantitatively affect the obtained results; however, the principal impact of VSL is the main subject of this work.

For origin links, i.e., links that receive traffic demand \( d_m \) and forward it into the motorway network, a simple queue model is used (Fig. 3). The outflow \( q_o \) of an origin link \( o \) depends on the arriving demand, on the traffic conditions of the corresponding mainstream segment (\( \mu,1 \)) and on the existence of ramp metering control measures. If ramp metering is applied, then the outflow \( q_o(k) \) that leaves origin \( o \) during period \( k \), is a portion \( r_o(k) \) of the outflow \( \hat{q}_o(k) \) that would leave in absence of ramp metering. Thus, \( r_o(k) \in [r_{min,o},1] \) is the metering rate for the origin link \( o \), i.e., a control variable, where \( r_{min,o} \) is a minimum admissible value. If \( r_o(k) = 1 \), no ramp metering is applied, else \( r_o(k) < 1 \). The queuing model is described by the following conservation equation:

\[
w_o(k+1) = w_o(k) + T[d_o(k) - q_o(k)]
\]

where \( w_o(k) \) (veh/km/lane) is the density at time \( kT \) and \( d_o(k) \) (veh/h) is the demand flow at \( o \). The outflow \( q_o(k) \) is determined as follows:

\[
q_o(k) = r_o(k)\hat{q}_o(k)
\]

with

\[
\hat{q}_o(k) = \min\{\hat{q}_{o,1}(k),\hat{q}_{o,2}(k)\}
\]

and

\[
\hat{q}_{o,1}(k) = d_o(k) + w_o(k)/T
\]

\[
\hat{q}_{o,2}(k) = Q_o \min\left\{1,\frac{\rho^{max}_m - \rho^{cr,m}}{\rho^{max}_m - \rho^{cr,m}}\right\}
\]

where \( Q_o \) (veh/h) is the on-ramp's flow capacity, i.e., the on-ramp's maximum possible outflow under free-flow traffic conditions in the mainstream, and \( \rho^{max}_m \) (veh/km/lane) is the maximum density in the network. According to (10)-(12), the uncontrolled outflow \( \hat{q}_o(k) \) is determined by the current origin demand if \( \hat{q}_{o,1}(k) \leq \hat{q}_{o,2}(k) \); else by the geometrical capacity \( Q_o \) if the mainstream density is undercritical, i.e., \( \rho^{cr,m} < \rho^{cr,m} \), or by the reduced capacity due to congestion of the mainstream if \( \rho^{cr,m} > \rho^{cr,m} \).

Motorway bifurcations and junctions (including on-ramps and off-ramps) are represented by nodes. Traffic enters a node \( n \) through a number of input links and is distributed to the output links according to the following equations:

\[
Q_n(k) = \sum_{\mu=1}^{\mu_n} q_{\mu,n}(k)
\]

\[
q_{n,o}(k) = \beta^o_n Q_n(k) \quad \forall m \in O_n
\]

where \( I_n \) is the set of links entering node \( n \); \( Q_n(k) \) is the total traffic volume entering \( n \) at period \( k \); \( q_{n,o}(k) \) is the traffic volume that leaves \( n \) via link \( \mu \); and \( \beta^o_n \in [0,1] \) is the portion of \( Q_n(k) \) that leaves \( n \) through link \( \mu \) (turning rates).

At a network node \( n \), the upstream influence of the downstream-link density (i.e., in case of congestion spillback) has to be taken into account in the last segment of the incoming links (see (3) for \( i = N_n \)). This is provided via

\[
\rho_{n,N_{n-1}}(k) = \sum_{\mu=1}^{\mu_n} \rho^2_{\mu,k}(k)/\sum_{\mu=1}^{\mu_n} \rho_{\mu,k}(k)
\]

where \( \rho_{n,N_{n-1}}(k) \) is the virtual density downstream of any entering link \( \mu \) to be used in (3) for \( i = N_n \). This is the density of the first segment of the leaving link \( \mu \). The quadratic form is used to account for the fact that congestion...
on one leaving link may spill back into the entering link even if there is free flow in the other leaving links.

Similarly, at a network node \( n \) the downstream influence of the upstream-link speed has to be taken into account according to (3) for \( i = 1 \). The required upstream mean speed value is calculated from the flow-weighted average

\[
v_{n,i}(k) = \sum_{\mu \in I_n} v_{\mu,n}(k) q_{\mu,n}(k) / \sum_{\mu \in I_n} q_{\mu,n}(k)
\]

where \( v_{n,i}(k) \) is the virtual speed upstream of any leaving link \( m \) that is needed in (3) for \( i = 1 \).

Combining the equations developed above, a nonlinear macroscopic discrete-time state-space model

\[
x(k+1) = f[x(k), u(k), d(k)], \quad x(0) = x_0
\]

is obtained for the entire motorway network, where \( x \) is the state vector, \( u \) is the control vector and \( d \) is the disturbance vector. The state vector consists of the densities \( \rho_{m,i} \) and the mean speeds \( v_{n,i} \) of every segment \( i \) of every link \( m \) and the queues \( w_o \) of every origin \( o \). The control vector consists of the VSL rates \( b_o \) of every link \( m \) where VSL is applied and of the ramp metering rates \( r_o \) of every origin \( o \) that is metered. The disturbance vector consists of the demand \( d_o \) at every origin \( o \) and the turning rates \( \beta_o \) at every bifurcation node \( n \).

3. THE OPTIMAL CONTROL PROBLEM

The integrated motorway network traffic control problem is formulated as a discrete-time dynamic optimal control problem with constrained control variables over a given optimisation horizon \( K_p \), which is solved very efficiently even for large-scale networks by a suitable feasible-direction algorithm (Papageorgiou and Marinaki, 1995). This extended formulation and the solution algorithm are incorporated in an accordingly extended version of the open-loop optimal control tool AMOC (Advanced Motorway Optimal Control) (Kotsialos et al., 2002b) which is able to consider coordinated ramp metering, system-optimum route guidance, variable speed limits (using the introduced extension via (4)-(7)) as well as integrated control combining all control measures.

If only ramp metering and VSL are considered, the general discrete-time formulation of the optimal control problem is the following: Given disturbance predictions \( d(k) \), \( k = 0, 1, \ldots, K_p - 1 \), and the initial state \( x_0 = x(0) \); minimise

\[
J = J(K) + \sum_{k=0}^{K_p-1} \varphi[x(k), u(k), d(k)]
\]

subject to (17) and the inequality constraints imposed on the ramp metering rates, \( r_{\min} \leq r_o(k) \leq 1 \), and the VSL rates, \( b_{\min} \leq b_o(k) \leq 1 \).

The chosen cost criterion is the Total Time Spent (TTS) of all vehicles in the network (including the waiting time experienced in the ramp queues) which is a natural objective for the traffic systems considered. The maximum ramp queue constraints may be taken into account via the introduction of penalty terms in the cost criterion penalising queue lengths larger than \( w_{\max, o} \), which is a pre-determined maximum admissible queue for origin \( o \). Another penalty term may be added in order to suppress high-frequency oscillations of the optimal control trajectories. More precisely the cost criterion used as (18) is the following

\[
J = T \sum_{k=1}^{K_p} \sum_{i} \rho_{m,i}(k) L_i + \sum_{k=1}^{K_p} \sum_{o} w_o(k)
\]

\[
+ T \sum_{k=1}^{K_p} \sum_{o} \alpha_f \left[ r_o(k) - r_o(k-1) \right]^2
\]

\[
+ T \sum_{k=1}^{K_p} \sum_{o} \alpha_b \left[ b_o(k) - b_o(k-1) \right]^2
\]

\[
+ T \sum_{k=1}^{K_p} \sum_{o} \alpha_s \left[ \max \{0, w_o(k) - w_{\max, o} \} \right]^2
\]

where \( \alpha_f \), \( \alpha_b \) and \( \alpha_s \) are weighting factors for the corresponding penalty terms.

The solution determined by AMOC consists of the optimal ramp metering and VSL rate trajectories as well as the corresponding optimal state trajectory. It is interesting to note that the solution algorithm can be readily modified to account for control variables that change their value less frequently than the state variables. Moreover, for the VSL rates, common control variables can be considered for clusters of links.

4. APPLICATION RESULTS

4.1 The Test Motorway Axis

For the purposes of this study, a hypothetical three-lane motorway axis of 4.5 km, sketched in Fig. 4, is considered. The mainstream is divided into four links (L1 to L4). There are two on-ramps (O1 and O2) on this motorway and one off-ramp (D1) in-between. The trapezoidal demand profiles, shown in Fig. 5, are used for the two on-ramps while a constant demand of 4200 veh/h is considered for the mainstream flow. The exit rate, i.e., the percentage of the mainstream flow that leaves the motorway, at the off-ramp D1 is set to 5% and the model time step used is \( T = 10 \) s. A

Fig. 4. The test three-lane motorway axis with two on-ramps.

Fig. 5. Trapezoidal demand profiles for the two on-ramps.
number of different scenarios are examined in the following, each for a time horizon of 2 hours.

4.2 No-control Case

When no control measures are applied, the resulting ramp queue, density and flow profiles for both merge areas are shown in Fig. 6. Mainstream congestion appears after 50 min in the merge area of the O2 on-ramp due to high flows arriving there; this leads to a visible gradual mainstream flow decrease (capacity drop). The created congestion travels upstream and reaches the merge area of the O1 on-ramp at around $t = 1$ hour, leading to a visible flow decrease there as well. In this scenario, the small queue (9 veh) formed at the O1 on-ramp is due to the reduction of the on-ramp's flow capacity caused by the mainstream congestion (see (12)).

The resulting TTS is equal to 1204 veh·h. The TTS was calculated over the 2 hours of simulation depicted in the figures plus a cool-down period of 10 min with zero inflows, which was introduced in order to have equal traffic conditions on the stretch (empty network) at the end of the simulation and hence comparable TTS values for all investigated scenarios.

4.3 Coordinated ramp metering

AMOC is now applied for coordinated ramp metering with maximum admissible ramp queues equal to 50 veh. The ramp metering rates are allowed to change every 30 sec with a minimum admissible value equal to 0.05 in order to avoid ramp closure. The resulting TTS value is equal to 1172 veh·h, which is a 2.7% improvement compared to the no-control case. The related ramp queue, density and flow profiles for both merge areas are shown in Fig. 7.

The optimal solution maintains the density and flow at the O2 merge area as long as possible close to the critical and capacity value (5940 veh/h), respectively, so as to maximize the freeway exit flow (which leads to minimisation of TTS). To achieve this, ramp queues are created quasi-simultaneously in both ramps. The congestion appearing at around $t = 1.2$ hours is unavoidable in view of the high involved demands and limited ramp storage.

4.4 VSL control

For the application of VSL control, the mainstream is divided into three clusters of links, each one with its own control variable. The first cluster comprises L1, the second cluster comprises L2 and L3 and the third cluster comprises L4, i.e. one control variable is used for both L2 and L3. The VSL rates are allowed to change every 300 sec with a minimum admissible value equal to 0.5. The resulting TTS value is equal to 741 veh·h, which is a 38.5% improvement compared to the no-control case. The related ramp queue, density and flow profiles for both merge areas are shown in Fig. 8 while the optimal VSL rate trajectories are shown in Fig. 9.

The trajectory for the VSL rate of L4 is such that the highest possible flow and capacity value (around 6400 veh/h) is achieved at the merge area of the O2 on-ramp. After $t = 1.0$ hours, the VSL rate of L1 switches gradually from 0.95 to 0.5. As a result, the flow arriving in the bottleneck area downstream of the O2 on-ramp decreases temporarily, thus delaying the bottleneck activation and the resulting congestion. Note that this temporary flow decrease during the VSL-triggered traffic state transition is due to the fact that density in the new VSL-state is higher than in the previous one; thus, during this transition, the flow is temporarily reduced in order to "create" the higher density of the new VSL-state. Despite these measures, the congestion cannot be fully avoided but it is delayed compared to previous scenarios.

4.5 Integrated control

When both coordinated ramp metering and VSL control are applied, i.e. integrated control, TTS is reduced even more to 673 veh·h, which is a 44.1% improvement compared to the no-control case. The related ramp queue, density and flow profiles for both merge areas are shown in Fig. 10 while the optimal VSL rate trajectories are shown in Fig. 11.
Once more the trajectory for the VSL rate of L4 is such that the highest possible flow and capacity value (around 6400 veh/h) is achieved at the merge area of the O2 on-ramp while the ramp queues created in both ramps hold back traffic in order to avoid the congestion. In this case, no real VSL action is needed for L1 as ramp metering is sufficient to completely suppress the congestion formation.

5. CONCLUSIONS

A quantitative model for the impact of VSL on the aggregate traffic flow behaviour was proposed. VSL were incorporated in a general second-order traffic flow model as an additional control component leading to an accordingly extended optimal control formulation. An illustrative example was presented under different control scenarios. It was shown that traffic flow efficiency can be substantially improved when VSL control measures are used, particularly in integration with coordinated ramp metering.

Due to various inherent uncertainties, the open-loop optimal solution delivered by optimal control approaches becomes suboptimal when directly applied to the motorway traffic process. However, the optimal results can be used in a rolling horizon mode or can be utilised to extract useful conclusions for the development of efficient feedback control strategies. This is left for future investigations.

REFERENCES


