

## Efficient freeway MPC by parameterization of ALINEA and a speed-limited area

van de Weg, Goof Sterk; Hegyi, Andreas; Hoogendoorn, Serge Paul; De Schutter, Bart

**DOI**

[10.1109/TITS.2018.2790167](https://doi.org/10.1109/TITS.2018.2790167)

**Publication date**

2019

**Document Version**

Accepted author manuscript

**Published in**

IEEE Transactions on Intelligent Transportation Systems

**Citation (APA)**

van de Weg, G. S., Hegyi, A., Hoogendoorn, S. P., & De Schutter, B. (2019). Efficient freeway MPC by parameterization of ALINEA and a speed-limited area. *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 16-29. Article 8283509. <https://doi.org/10.1109/TITS.2018.2790167>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Efficient freeway MPC by parameterization of ALINEA and a speed-limited area

Goof Sterk van de Weg, Andreas Hegyi, Serge Paul Hoogendoorn, Bart De Schutter

**Abstract**—Freeway congestion can reduce the freeway throughput due to the capacity drop or due to blocking caused by spillback to upstream ramps. Research has shown that congestion can be reduced by the application of ramp metering and variable speed limits. Model predictive control is a promising strategy for the optimization of the ramp metering rates and variable speed limits to improve the freeway throughput. However, several challenges have to be addressed before it can be applied for the control of freeway traffic. This paper focuses on the challenge of reducing the computation time of MPC strategies for the integration of variable speed limits and ramp metering. This is realized via a parameterized control strategy that optimizes the upstream and downstream boundaries of a speed-limited area and the parameters of the ALINEA ramp metering strategy. Due to the parameterization, the solution space reduces substantially, leading to an improved computation time. More specifically, the number of optimization variables for the variable speed limit strategy becomes independent of the number of variable message signs, and the number of optimization variables for the ramp metering strategy becomes independent of the prediction horizon. The control strategy is evaluated with a macroscopic model of a two-lane freeway with two on-ramps and off-ramps. It is shown that parameterization realizes improved throughput when compared to a non-parameterized strategy when using the same amount of computation time.

**Index Terms**—Variable speed limits, ramp metering, freeway management, throughput improvement, model predictive control.

## I. INTRODUCTION

**F**REEWAY congestion can reduce the freeway throughput causing societal, economical, and environmental costs. Two main reasons exist why congestion reduces throughput. First of all, congestion causes a capacity drop, i.e. the flow downstream of congestion is lower than the capacity flow that can be achieved under free-flow conditions [1], [2]. Secondly, congestion can spill back in the upstream direction and cause blocking of traffic bound for off-ramps.

Congestion can be mitigated by dynamic traffic management measures. Two popular dynamic traffic management measures on which this paper focuses are ramp metering (RM) and variable speed limits (VSLs). RM is typically used to limit the number of vehicles that want to enter the freeway from an on-ramp using a traffic light. In this way, the flow into a downstream bottleneck can be reduced so that congestion can be prevented, postponed, or resolved. VSLs are speed limits that can be varied over time and are displayed using

variable message signs. VSLs can be used to reduce the speed of freeway traffic and they are typically applied for safety reasons. However, several approaches have been designed to reduce freeway congestion using VSLs. In this paper we study the application of RM and VSLs to improve freeway throughput by reducing congestion with the aim of developing an optimization-based control strategy for the integration of VSLs and RM.

### A. Review of RM and VSL strategies

The development of RM and VSL strategies – i.e. control algorithms – is an active research area. In this brief overview we discuss several VSL and RM strategies that aim at freeway throughput improvement. We focus here on discussing the mechanisms in traffic flow exploited by the controllers, the controller properties, and investigate challenges and opportunities for further controller development. After concluding this section, we review the literature on model predictive control strategies for the integration of RM and VSLs in the next section.

1) *VSL*: According to Hegyi *et al.* [3], two main categories of VSL strategies for the improvement of freeway throughput exist, namely, the homogenizing types and the flow-limiting types. The idea behind the homogenizing types is that by displaying VSLs that are similar to the average speed of the traffic, speed differences between vehicles will be reduced but no significant reduction of the average speed will result [4], [5], [6]. In this way, the traffic flow is homogenized, resulting in a reduction of the probability of a traffic breakdown, and thus, leading to an improved freeway throughput. However, while field tests did show a reduction in speed differences, implying a more homogeneous traffic flow, no evidence was found for an improved freeway throughput [5].

The main idea behind VSL strategies of the flow-limiting type is that by imposing VSLs the flow on the freeway can be controlled. Several approaches can be found in the literature that are of the flow-limiting type. Carlson *et al.* [7] proposed a VSL strategy called mainstream traffic flow control (MTFC) for controlling freeway traffic entering a bottleneck. This strategy adjusts the VSL value at a location upstream of a bottleneck in order to create a controlled congestion upstream of the bottleneck so that the bottleneck inflow matches the bottleneck capacity. Several simulation studies were performed showing improved freeway throughput. Challenges of this approach are that very low VSL values may have to be displayed and that the application of the strategy is limited to specific locations in a road network.

Goof Sterk van de Weg, Andreas Hegyi, and Serge Paul Hoogendoorn are with the Transport & Planning department, TU Delft, the Netherlands

Bart De Schutter is with the Delft Center for Systems and Control, TU Delft, the Netherlands

Manuscript received July 5, 2017;

Hegyí *et al.* [3] proposed a VSL strategy called SPECIALIST based on shock wave theory against jam waves – i.e. congestion with a length of roughly 1 to 2 km that propagates in the upstream direction of the freeway. The SPECIALIST algorithm detects a jam wave and when it assesses this jam wave as resolvable it first applies a pre-defined VSL value instantaneously over a freeway stretch directly upstream of the jam wave. Next, VSLs are imposed upstream of the speed-limited area to stabilize the traffic flow – by creating a stable combination of speed and density – that is approaching the speed-limited area. This causes a reduction of the flow into the jam wave so that it can resolve without triggering an upstream congestion. After the jam wave is resolved, the traffic in the speed-limited area can be released and a higher freeway flow can be achieved since the capacity drop is no longer present. The density and flow in (and downstream of) the speed-limited area can be controlled by adjusting the speed with which the upstream (and downstream) boundary of the speed-limited area propagates. SPECIALIST was tested on the A12 freeway in the Netherlands and it was found that it is capable of resolving jam waves and stabilizing traffic, resulting in improved freeway throughput [3]. Recently, Mahajan *et al.* [8] proposed a reformulation of SPECIALIST called COSCAL v2. In contrast to the SPECIALIST algorithm which has a feed-forward structure, the COSCAL v2 algorithm has a feedback structure so that it can better adjust its control action to disturbances.

Chen *et al.* [9] proposed an alternative approach to resolve congestion at a bottleneck location. In their approach, VSLs are imposed upstream of the bottleneck first so that the congestion head moves away from the bottleneck and the impact of the capacity drop is decreased. After that, by adjusting the VSL values, the outflow of the speed-limited area is adjusted so that it matches the bottleneck capacity. To the best knowledge of the authors, no simulation studies have been carried out yet with this algorithm.

Recently, Zhang *et al.* [10] proposed a VSL control strategy integrated with a lane change control strategy to reduce bottleneck congestion caused by incidents. In their approach, lane change control is used to remove the capacity drop and VSL control is used upstream of the incident location to realize target densities that maximize the bottleneck flow.

2) *RM*: Similar to VSL strategies of the flow-limiting type, RM is primarily used to limit the freeway flow. The most well-known RM algorithm is ALINEA [11]. This feedback control strategy for a single on-ramp uses measurements downstream of the on-ramp and regulates the on-ramp flow with the objective of keeping the freeway flow near its critical density. In this way, congestion caused by excessive on-ramp flows can be prevented or postponed and in this way, the impact of the capacity drop is reduced, resulting in improved freeway throughput. Several other control strategies for single on-ramps exist. Middelham *et al.* [12] discusses a demand-capacity RM strategy that uses upstream freeway flow measurements in order to maximize the freeway flow. Due to its feed-forward nature its performance may deteriorate due to disturbances in the traffic flow. A major challenge of these local RM strategies is that the on-ramp queue may spill back to the upstream urban

network. Queue management may help to limit the on-ramp queue but also reduces the time that RM can be effective [13], [14].

Coordination of RM at multiple on-ramps can help to extend the RM time. HERO is an algorithm that coordinates the ALINEA-based RM actions of different on-ramps [13]. Whenever the queue caused by RM at a downstream on-ramp exceeds a threshold, the upstream RM installation starts an RM algorithm that aims at controlling the upstream queue towards a set-point determined by the downstream on-ramp. This prevents the queue at the downstream on-ramp from exceeding the maximum length and allows a longer RM time. Difficulties of coordination are that there exist time delays between the interactions of on-ramps and that not all traffic of upstream on-ramps might be headed to the bottleneck. Not including these effects may cause unnecessary delays for traffic that is not headed to the bottleneck, which may not be fair [15]. One way to include these effects is by predicting the (near) future impact of the control signal on the system performance. Model-based optimal control approaches are typically suited to include such effects and will be discussed in the next section.

3) *Integrated approaches to RM and VSL*: Integrating RM and VSL strategies is expected to lead to further freeway performance improvements. From a control engineering point of view this can be explained by the fact that the control freedom is increased. From a traffic-flow-theoretical point of view this can be explained by the possibility to distribute the flow-limiting task over freeway traffic and on-ramp traffic. Schelling *et al.* [16] proposed an extension of SPECIALIST so that it can cope with a metered on-ramp. Van de Weg *et al.* [17] extended the in-car algorithm COSCAL v1 – which is similar to SPECIALIST – with RM. Mahajan *et al.* [8] extended a macroscopic version of COSCAL v1, named COSCAL v2 with RM. In these approaches, it is computed at what time RM is switched on in order to assist the VSL system that resolves jam waves. These studies show that it is possible to integrate the VSL and RM task to resolve jam wave using limited computation time when considering only a single on-ramp. However, a challenge may be the extension to multiple on-ramps, which may lead to a complex control problem due to the time delays between the effects of different actuators.

Carlson *et al.* [14] integrated the MTFC approach with RM. They apply ALINEA RM in order to prevent congestion from forming at the bottleneck location. When the on-ramp is full or when the RM rate is near its minimum allowed rate, MTFC control is switched on in order to prolong the RM time. The authors showed that the approach outperforms non-integrated algorithms and realizes a performance that is near the performance realized with optimal control for a bottleneck scenario simulated using a macroscopic traffic flow model. An advantage of this approach is that it is based on a simple feedback control structure. Iordanidou *et al.* [?] extended this approach to coordinate RM and VSL actions at different locations by balancing the travel delays caused by the different actuators.

4) *Conclusions from the literature*: In conclusion, RM and VSLs can both limit the freeway flow. These flow reductions can be used to prevent, postpone, or resolve congestion,

resulting in improved freeway throughput, since the impact of the capacity drop is reduced. Various algorithms have been developed for RM and VSLs. These algorithms differ in the traffic-flow-theoretical mechanisms that they exploit and their control-theoretical structure. Studies have shown that integrating RM and VSLs can lead to a better performance when compared to isolated systems. However, the control of multiple RM and VSL gantries is a complex problem due to the time delay in the impact of elements on each other.

### B. Review of model-based optimization strategies for freeway traffic control

A promising approach to account for the time delays of control actions on the network-wide performance is model predictive control (MPC) [18]. MPC uses a prediction model to predict the state of a process over a period of time – called the prediction window – given the current state, a prediction of the disturbances – i.e. inputs that cannot be controlled –, and a candidate control signal. Based on this prediction the performance of the process is expressed using an objective function. Using an optimization technique the control signal is found that leads to the minimum (or maximum) of the objective function. The first step of the control signal is applied to the process, and at the next time step, when new measurements are available, the control signal is optimized again. This is called the receding horizon principle.

Despite the advantages of MPC there also exist several open problems when it is applied to freeway traffic control as discussed in detail in [19]. Some key problems are that an accurate prediction of the traffic demand should be available, that the controller should be able to deal with uncertainties, and that the computation time used by the controller should be short enough for real-time application. In this paper we will focus on reducing the computation time of an MPC strategy.

Several authors have applied MPC to the freeway traffic control problem. Kotsialos *et al.* [20] and Hegyi *et al.* [21] used the second-order METANET model as a prediction model to optimize RM and integrated RM and VSL settings respectively. An advantage of using second-order models is that they can model more complex traffic dynamics. However, a major challenge is that the nonlinear optimization problem is computationally hard so that real-time application to large freeway networks is not feasible.

Roughly three main approaches exist to limit the computation time required by an MPC strategy. The first is to use computationally efficient traffic flow models. To this end, Gomes *et al.* [22] and Hajiahmadi *et al.* [23] use first-order traffic flow models to formulate linear and mixed integer linear optimization problems respectively. The disadvantage of using first-order traffic flow models is that some characteristics of the traffic dynamics may be lost. This may cause a performance loss when applied to a more complex traffic process.

The second strategy is to divide the optimization problem in multiple, possibly overlapping, sub-problems. One such strategy is distributed MPC as in [24]. In such approaches, the freeway network is divided into smaller sub-networks. The sub-problems that need to be solved involve optimization

of the sub-network performance and the impact on the total network performance. In some cases this might lead to reduced computation times and similar performance as centralized MPC.

The third strategy is to reduce the number of control parameters that need to be optimized by parameterizing existing control strategies. For instance, Zegeye *et al.* [25] integrated the ALINEA algorithm and a feedback algorithm for VSLs so that only the gains of the feedback strategies had to be optimized. The approach was only applied to cases where the same strategy was used for every actuator type – i.e. VSL or RM – in the network at every time step. Lu *et al.* [26] first designed the VSL signal after which the RM rates could be computed using a linear optimization problem. Recently, Van de Weg *et al.* [27] proposed a parameterization based on SPECIALIST to resolve jam waves using VSLs so that the size of the optimization problem becomes independent of the number of VSL gantries. It is shown using simulations that this approach is able to realize similar performance as the MPC proposed by Hegyi *et al.* [21] in significantly less CPU time while outperforming the approach of Zegeye *et al.* [25]. A limitation of the approach of [27] is that it is not yet suited to account for RM and that the performance is only tested in a scenario where throughput is improved by resolving a jam wave.

### C. Research approach and contributions

This paper presents a parameterized MPC strategy for integrated RM and VSLs to improve the freeway throughput. In this way, a better trade-off between the realized throughput improvement and the utilized computation time for integrated optimization of RM and VSL is obtained. The method generalizes the previous work of Van de Weg *et al.* [27]. Compared to that work, two main contributions are made. First of all, the parameterized VSL approach is extended with a parameterized RM control strategy. Secondly, an extensive qualitative analysis into the controller behavior is carried out when applying the strategy to a jam wave and a bottleneck scenario. Also, the qualitative behavior of the different combinations of RM and VSL is studied. In contrast to the work of Zegeye *et al.* [25], per RM installation the RM gain and set-point, and switching times are added to the optimization problem. The switching times are used to change the feedback policy when the traffic situation changes. The parameterization of VSLs and RM rates in METANET is formulated in such a way that the optimization problem can be solved using gradient-based solvers, which are generally faster compared to gradient-free solvers when the problem size is not too large. The third contribution of this paper is to provide insight into the impact of the available computation time budget on the controller performance.

## II. CONTROLLER DESIGN

The parameterized MPC strategy proposed in this paper is able to optimize both RM rates and VSL values with the aim of improving the freeway throughput. In the approach proposed in this paper the head and tail of a speed-limited



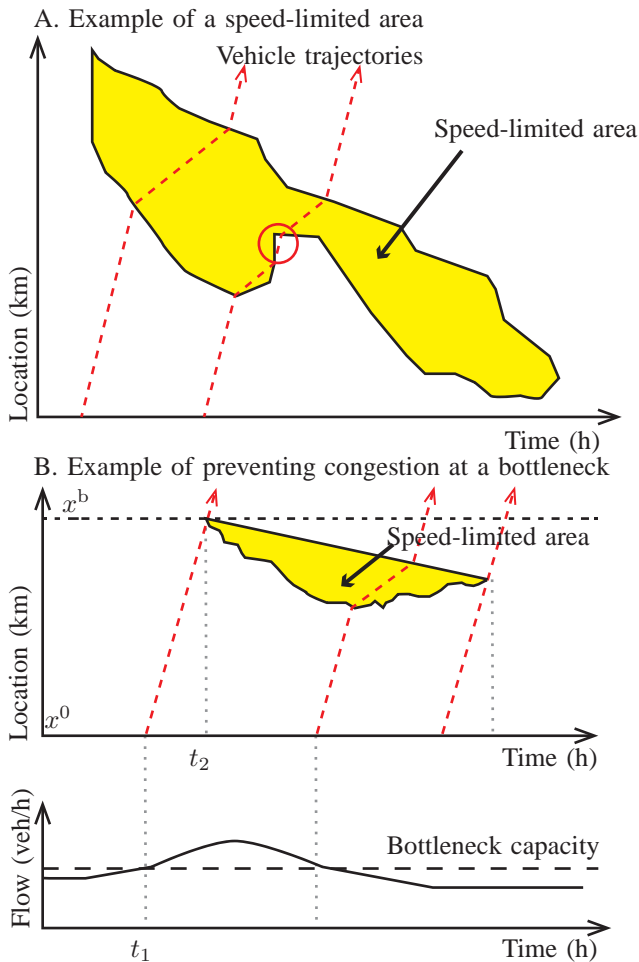


Fig. 1. A: Example of a speed-limited area that can be used to influence the traffic flow. The red-dashed lines indicate examples of vehicle trajectories. The second vehicle trajectory illustrates a vehicle experiencing a speed limit drop twice – as indicated with the red circle –, which should not occur. B: Top figure: example of a speed-limited area that can be used to prevent congestion at the bottleneck location  $x^b$ . Bottom figure: the demand entering the freeway at location  $x^0$  [27].

area are parameterized. In this way the number of optimization parameters becomes independent of the freeway length, which would be the case when using non-parameterized optimization approaches. Additionally, we optimize the parameters of the ALINEA strategy and we optimize the switching times when the controllers should change the parameters of the ALINEA strategy or when they should switch RM off. In this way, the number of optimization parameters for every RM installation becomes independent of the prediction horizon.

#### A. Design considerations

Several design considerations are taken into account when developing the parameterized MPC strategy. Special attention is paid to satisfy the requirements for applying RM or VSLs for freeway traffic control. While the primary objective of this paper is to design a control strategy of which the computation time required by the controller is lower than the controller sampling time, (which is in the range of (several) minutes), some design requirements are taken into account as well,

which are also important for the practical applicability of this method, namely:

- 1) Only a limited number of VSL values can be displayed. For instance, in the Netherlands it is only possible to show 50, 60, 70, 80, 90, and 100 km/h.
- 2) A VSL or RM system should not cause unsafe situations.
- 3) An RM system typically causes a queue on the on-ramp. The queue length should be bounded by a maximum value to avoid spillback to the upstream road network.
- 4) The RM rate is typically bounded by a minimum and maximum value.

Below, first the design considerations of the VSLs are introduced, followed by the considerations for implementing RM.

1) *VSL control design considerations:* As indicated by Van de Weg *et al.* [28], a speed-limited area – as shown in Figure 1 A – can be created by imposing VSLs. It follows from shock-wave theory that there is a relation between the slope of the boundaries of the speed-limited area and the resulting flow and density downstream of that slope [3], [29]. If the slope is steeper (more negative) then the resulting density and flow are higher. By adjusting the speed with which the upstream boundary – i.e. the tail – propagates over time, a stable combination of density and flow can be realized in the speed-limited area. Similarly, by adjusting the speed with which the downstream boundary – i.e. the head – propagates over time, the outflow of the speed-limited area can be controlled so that it is just below or at the freeway capacity. SPECIALIST is an example of an algorithm that uses a speed-limited area to resolve a jam wave [3].

Figure 1 B presents an example of using a speed-limited area in order to prevent congestion at a bottleneck. At time  $t_1$  (h) an excess demand – as illustrated in the bottom figure – enters the freeway at location  $x^0$  (km). The time-space plot in the top figure shows that this demand reaches the bottleneck location  $x^b$  (km) at time  $t_2$  (h). At this time, congestion would appear in a no control situation. However, by imposing a speed-limited area as illustrated in the top figure, congestion may be prevented.

Note that the effectiveness of this control method is limited by the length of the stretch over which speed limits are available. If the necessary queue storage space exceeds this length then the approach becomes ineffective. An analysis method to evaluate the expected effectiveness based on measurement data is presented in [28].

Several design considerations are taken into account when implementing a speed-limited area. First of all, it is assumed that the value of the speed-limits in the speed-limited area is constant over time. This implies that a segment between two variable message signs is either speed-limited or not. Additionally, it is assumed that only one speed-limited area can be active at a time. Apart from that, the dynamics of the head and tail of the speed-limited area should be such that the individual vehicles can only enter and exit the speed-limited area once. If an individual vehicle observes multiple fluctuations of the speed limits, this can lead to unsafe situations, annoyance, or poor compliance. As an example, the second vehicle in Figure 1 A experiences such fluctuations. In order to prevent such behavior, the positions  $x^{H,sl}$  (km) and  $x^{T,sl}$  (km) of

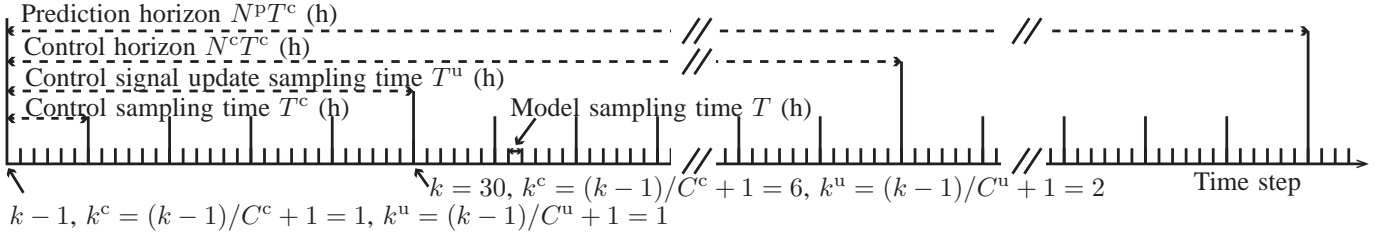


Fig. 2. Overview of the timing used in the paper for  $T$  is 10 s,  $T^c$  is 60 s,  $T^u$  is 300 s.

respectively the head and the tail of the speed-limited area are allowed to propagate in the downstream direction with a speed that is lower or equal to the effective speed  $v^{\text{eff}}$  (km/h). In the upstream direction they can propagate with any speed.

The speed in the speed-limited area is equal to the effective speed  $v^{\text{eff}}$  corresponding to the imposed VSLs. The effective speed is defined as the speed with which vehicles drive in the speed-limited area which includes possible non-compliance. This can be estimated e.g. from field tests as presented in [3].

The proposed parameterization reduces the number of optimization variables for VSLs to two per control time step. Note that the number of optimization variables at every control time step used in a nominal MPC strategy is equal to the number of VSL actuators. Hence, the advantage of this parameterization is that the number of optimization variables is reduced, and that the number of optimization variables is independent of the number of VSL actuators.

2) *RM control design considerations*: A feedback RM algorithm is used in this paper to control the on-ramp flow that has to satisfy the following properties:

- The RM rate  $r_o(k)$  (-) of an origin  $o$  should be between the minimum allowed RM rate  $r^{\min} \geq 0$  (-) and 1.
- The on-ramp queue length  $w_o(k)$  (veh) should not exceed its maximum value  $w_o^{\max}$  (veh).

Different RM strategies could be applied depending on the traffic situation. For instance, when preventing congestion at a bottleneck location, the most sensible control strategy would be to control the on-ramp flows in such a way that the flow into a bottleneck is at or just below its capacity. The ALINEA algorithm is specifically designed to realize this objective. The ALINEA algorithm has the following form [30]:

$$r_o(k+1) = r_o(k) + K_o \frac{\rho_m^{\text{crit}} - \rho_{m,1}(k)}{\rho_m^{\text{crit}}}, \quad (1)$$

where  $\rho_m^{\text{crit}}$  (veh/km/lane) is the critical density of the link directly downstream of the on-ramp,  $\rho_{m,1}(k)$  (veh/km/lane) is the current density in the most upstream segment of the downstream link, and  $K_o$  (-) is the feedback gain.

When resolving a jam, the flow into the jam should be reduced as much as possible. The standard ALINEA RM algorithm is not suited to realize this, since it tries to fit as much traffic onto the freeway without exceeding the critical density. This can be solved by adapting the set-point  $\rho_{o,1}^{\text{set}}(k)$  (veh/km/lane) of the ALINEA strategy [31], [25]:

$$r_o(k+1) = r_o(k) + K_o \frac{\rho_{o,1}^{\text{set}}(k) - \rho_{m,1}(k)}{\rho_{o,1}^{\text{set}}(k)}. \quad (2)$$

Another advantage of including such a set-point is that coordination of on-ramps becomes possible. In the case of a downstream bottleneck or congestion, the set-points of the controllers of different on-ramps can be coordinated in order to distribute the RM task over the RM installations.

Finally, it might be necessary to switch set-points a certain number of times. For instance, when resolving a jam, the preferred strategy might be to reduce the on-ramp inflow as much as possible until the moment when the jam has been resolved and afterwards the freeway flow can be increased to capacity so that the on-ramp outflow can also be increased. These two different tasks require different set-points. Therefore, we propose the following feedback control algorithm:

- Initially, RM is off until switching time  $t_{o,1}^{\text{switch}}$  (h).
- From switching time  $t_{o,1}^{\text{switch}}$  until switching time  $t_{o,2}^{\text{switch}}$  (h), the feedback law (2) with feedback gain  $K_{o,1}^s$  (-) and set-point  $\rho_{o,1}^{\text{set}}$  (veh/km/lane) is used.
- From switching time  $t_{o,2}^{\text{switch}}$  until switching time  $t_{o,3}^{\text{switch}}$  (h), the feedback law (2) with feedback gain  $K_{o,2}^s$  (-) and set-point  $\rho_{o,2}^{\text{set}}$  (veh/km/lane) is used.
- After time  $t_{o,3}^{\text{switch}}$  the RM installation is switched off.

This parameterization requires 5 parameters per RM installation, namely, three switching times, and two set-points. If needed, the approach can be extended by adding more switching time instants or to optimize the feedback gains, which are now manually tuned.

### C. Traffic flow modelling: the extended METANET model

An extended version of the METANET model is adopted to predict the evolution of the traffic in the MPC controller. The METANET model presented in [32] along with the extensions proposed in [21] is adopted since it provides a detailed description of the traffic dynamics and it can reproduce relevant traffic characteristics such as jam waves and the capacity drop. Note that in the description below only the elements relevant for this paper are discussed. For a full description of the model see [32] and [21].

In the METANET model, a freeway is divided into homogeneous – i.e. having a constant number of lanes, no on-ramps and off-ramps, and constant characteristics – links  $m$  that are connected by nodes [32]. Each link  $m$  consists of  $N_m$  (-) segments of length  $L_m$  (km) with a number of  $\lambda_m$  (-) lanes. The flow  $q_{m,i}(k)$  (veh/h), density  $\rho_{m,i}(k)$  (veh/km/lane) and

$$V(\rho_{m,i}(k)) = \min \left[ v_m^{\text{free}} \exp \left( - \frac{1}{a_m} \left( \frac{\rho_{m,i}(k)}{\rho_m^{\text{crit}}(k)} \right)^{a_m} \right), v_{m,i}^{\text{ctrl}}(k) \right]. \quad (6)$$

$$q_{\mu,1}^{\text{lim}}(k) = \begin{cases} \lambda_{\mu} v_{\mu,1}^{\text{lim}}(k) \rho_{\mu}^{\text{crit}} \left[ - a_{\mu} \ln \left( \frac{v_{\mu,1}^{\text{lim}}(k)}{v_{\mu}^{\text{free}}(k)} \right)^{1/a_{\mu}} \right] & \text{if } v_{\mu,1}^{\text{lim}}(k) < V(\rho_{\mu}^{\text{crit}}(k)) \\ q_{\mu}^{\text{cap}} & \text{if } v_{\mu,1}^{\text{lim}}(k) \geq V(\rho_{\mu}^{\text{crit}}(k)) \end{cases} \quad (9)$$

### B. Timing

Before continuing, the timing of the approach is introduced and is illustrated in Figure 2. The discrete-time second-order traffic model METANET is used to describe the evolution of the traffic [32]. The time step of the model is indicated with  $k$  (-) and the corresponding sampling time with  $T$  (h). The time step  $k$  refers to the period  $[Tk, T(k+1))$ . The control signal sampling time is  $T^c = C^c T$  (h) with  $C^c \in \mathbb{N}^+$  (-), meaning that the value of the control signal can change at time instants  $k^c T^c$  (h). The control signal is updated at time instant  $k^u T^u$  (h) for which it holds that the control signal update time  $T^u = C^u T$  (h) with  $C^u \in \mathbb{N}^+$  (-). Note that it holds that  $t = Tk = T^c k^c = T^u k^u$  (h). The controller predicts the evolution of the traffic from control time step  $k^c + 1$  until control time step  $k^c + N^p$  where  $N^p$  (-) is the prediction horizon. The control input from control time step  $k^c$  until control time step  $k^c + N^c$  is optimized by the controller where  $N^c$  (-) is the control horizon and  $N^c \leq N^p$ . After the control horizon the control signal is taken to be constant.

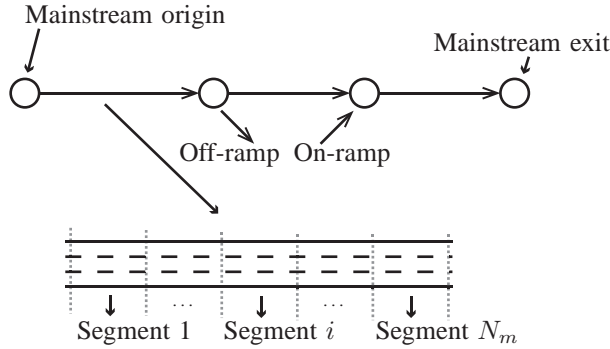


Fig. 3. Example of the METANET elements used in this paper. A freeway consist of mainstream origins, links, segments, off-ramps, on-ramps, and mainstream exits.

speed  $v_{m,i}(k)$  (km/h) in a link are updated according to:

$$q_{m,i}(k) = \rho_{m,i}(k) v_{m,i}(k) \lambda_m, \quad (3)$$

$$\rho_{m,i}(k+1) = \rho_{m,i}(k) + \frac{T}{L_m \lambda_m} (q_{m,i-1}(k) - q_{m,i}(k)), \quad (4)$$

$$\begin{aligned} v_{m,i}(k+1) &= v_{m,i}(k) + \frac{T}{\tau} (V(\rho_{m,i}(k)) - v_{m,i}(k)) \\ &+ \frac{T}{L_m} v_{m,i}(k) (v_{m,i-1}(k) - v_{m,i}(k)) \\ &- \frac{\eta T}{\tau L_m} \frac{\rho_{m,i+1}(k) - \rho_{m,i}(k)}{\rho_{m,i}(k) + \kappa}, \end{aligned} \quad (5)$$

In the latter equation,  $\tau$  and  $\kappa$  are model parameters. The parameter  $\eta$  (-) is set to  $\eta^{\text{high}}$  when the downstream density is higher than the density  $\rho_{m,i+1}(k)$  in segment  $i$ , and it is set to  $\eta^{\text{low}}$  when the downstream density is lower. This adjustment is adopted from [21] to reproduce the capacity drop. The speed  $V(\rho_{m,i}(k))$  (km/h) is given in (6) where  $a_m$  (-) is a model parameter, the speed  $v_m^{\text{free}}$  (km/h) is the free-flow speed of link  $m$ , and the density  $\rho_m^{\text{crit}}$  (veh/km) is the critical density, and

where the speed  $v_{m,i}^{\text{ctrl}}(k)$  (km/h) is the effective speed of the imposed speed limits that is corrected for the compliance of the road-users.

An origin is modeled using a simple queuing model describing the number of vehicles  $w_o(k)$  (veh) in the origin queue as a function of the demand  $d_o(k)$  (veh) and the outflow  $q_o(k)$  (veh/h):

$$w_o(k+1) = w_o(k) + T(d_o(k) - q_o(k)). \quad (7)$$

When an origin acts as the mainstream origin, the outflow is given by:

$$q_o(k) = \min \left[ d_o(k) + \frac{w_o(k)}{T}, q_{\mu,1}^{\text{lim}}(k) \right], \quad (8)$$

where the flow  $q_{\mu,1}^{\text{lim}}(k)$ , see (9), is determined by the traffic condition in the first link and the speed  $v_{\mu,1}^{\text{lim}}(k) = \min[v_{\mu,1}^{\text{ctrl}}(k), v_{\mu,1}(k)]$ . When an origin acts as a metered on-ramp, the outflow is given by:

$$q_o(k) = \min \left[ d_o(k) + \frac{w_o(k)}{T}, Q_0 r_o(k), Q_0 \frac{\rho_m^{\text{max}} - \rho_{m,1}(k)}{\rho_m^{\text{max}} - \rho_m^{\text{crit}}} \right], \quad (10)$$

with  $Q_0$  (veh/h) the on-ramp capacity, and  $r_o(k) \in [0, 1]$  the RM rate.

In the case that there is an on-ramp upstream of link  $m$ , then the term

$$- \frac{\delta T q_o(k) v_{m,1}(k)}{L_m \lambda_m (\rho_{m,1}(k) + \kappa)}, \quad (11)$$

is added to (5) for the first segment of link  $m$  with  $\delta$  (-) a model parameter.

Finally, when a link  $m$  has no leaving link – i.e. it is the most downstream link – the density  $\rho_{m,i_m^{\text{last}}+1}$  downstream of the last segment  $i_m^{\text{last}}$  is equal to:

$$\rho_{m,i_m^{\text{last}}+1} = \max \left[ \rho^{\text{DS}}(k), \min[\rho_{m,i_m^{\text{last}}}(k), \rho_m^{\text{crit}}] \right], \quad (12)$$

where the density  $\rho^{\text{DS}}(k)$  (veh/km/lane) is the destination density, which can be used as a boundary condition to the model.

$$v_{m,i}^{\text{ctrl}}(k) = \begin{cases} v^{\text{eff}} & \text{if } x^{\text{H,sl}}(k) > x_{m,i} \text{ and } x^{\text{T,sl}}(k) < x_{m,i} + L_m \text{ and } x^{\text{H,sl}}(k) > x^{\text{T,sl}}(k) \\ v^{\text{free}} & \text{otherwise,} \end{cases} \quad (13)$$

$$\gamma_{m,i}(k) = \max \left[ \frac{L_m - \max[x^{\text{T,sl}}(k) - x_{m,i}, 0] - \max[x_{m,i} + L_m - x^{\text{H,sl}}(k), 0]}{L_m}, 0 \right] \quad (14)$$

$$\tilde{r}_{o,i^{\text{P}}}(k) = \begin{cases} 1 & \text{if } i^{\text{P}} = 1 \text{ or } i^{\text{P}} = n^{\text{pol}} \\ \max \left( \min \left( \tilde{r}_o(k-1) + K_{o,i^{\text{P}}}^{\text{s}} \frac{\rho_{o,i^{\text{P}}}^{\text{set}} - \rho_{m,1}(k-1)}{\rho_{o,i^{\text{P}}}^{\text{set}}}, 1 \right), 0 \right) & \text{otherwise} \end{cases} \quad (17)$$

$$f_{i^{\text{P}}}^{\text{P}}(k) = \begin{cases} \frac{\max[0, T + \min[t_{o,i^{\text{P}}}^{\text{switch}} - kT]]}{T} & \text{if } i^{\text{P}} = 1 \\ \frac{\max[0, T - \max[t_{o,i^{\text{P}}-1}^{\text{switch}} - (k-1)T, 0]]}{T} & \text{if } i^{\text{P}} = n^{\text{pol}} \\ \frac{\max[0, T - \max[t_{o,i^{\text{P}}-1}^{\text{switch}} - (k-1)T, 0]] + \min[t_{o,i^{\text{P}}}^{\text{switch}} - kT]}{T} & \text{otherwise} \end{cases} \quad (18)$$

#### D. Extensions for parameterized MPC

This section details extensions that are included in order to use the model for parameterized MPC. These extensions do not affect the dynamic equations of the traffic states but rather the equations that relate the parameterized control signals to the dynamic equations to the control signals. Although the paper focuses on the use of METANET, the extensions may also be used in combination with other macroscopic traffic flow models.

1) *Extension with a speed-limited area:* In this paper, the VSLs  $v_{m,i}^{\text{ctrl}}(k)$  are determined by the head  $x^{\text{H,sl}}(k)$  (km) and tail  $x^{\text{T,sl}}(k)$  (km) of the speed-limited area as defined in (13), where  $x_{m,i}$  (km) is the most upstream location of segment  $i$  of link  $m$ .

In practice, the speed-limited area can either cover an entire segment or not cover it at all. This implies that the gradient of the objective function is not a continuous function of the location of the speed-limited area. In order to realize a gradient of the VSL signal that is differentiable everywhere, a parameter  $\gamma_{m,i}(k)$  (-) is introduced. The parameter  $\gamma_{m,i}(k)$  denotes the fraction of the segment that is covered by speed limits as defined in (14).

In the optimization, the speed  $v_{m,i}^{\text{ctrl}}(k)$  in (6) is replaced by  $\hat{v}_{m,i}^{\text{ctrl}}(k)$  by taking the weighted average of the effective speed  $v^{\text{eff}}$  and the equilibrium speed  $v^{\text{FD}}(\rho_{m,i}(k))$ :

$$\hat{v}_{m,i}^{\text{ctrl}}(k) = \gamma_{m,i}(k)v^{\text{eff}} + (1 - \gamma_{m,i}(k))v^{\text{FD}}(\rho_{m,i}(k)). \quad (15)$$

2) *Extension with feedback ramp metering:* The feedback RM control strategy results in a flow reduction factor  $\tilde{r}_o(k)$  (-) that limits the on-ramp flow [20]. The overall RM control strategy is as follows: until time  $t_{o,1}^{\text{switch}}$  RM is off and the RM rate is equal to 1; this policy is indicated with policy index  $i^{\text{P}} = 1$  (-). After that time until time  $t_{o,2}^{\text{switch}}$  the ALINEA algorithm is used to meter the on-ramp traffic with the gain  $K_{o,2}^{\text{s}}$  to reach the set point  $\rho_{o,2}^{\text{set}}$ ; this corresponds to policy  $i^{\text{P}} = 2$ . After time  $t_{o,2}^{\text{switch}}$  until time  $t_{o,3}^{\text{switch}}$  the maximum queue length strategy is used with gain  $K_{o,3}^{\text{s}}$  to reach the set-point  $\rho_{o,3}^{\text{set}}$ ; this corresponds to policy  $i^{\text{P}} = 3$ . After time  $t_{o,3}^{\text{switch}}$  the RM rate is switched to 1; this corresponds to policy  $i^{\text{P}} =$

4. In total a number of  $n^{\text{pol}} = 4$  (-) policies per ramp are available.

The switching time instants  $t_{o,i^{\text{P}}}^{\text{switch}}$  are real-valued while the actual model timing is discrete. This leads to a discontinuous gradient. In order to prevent this, the RM rates of the different policies  $\tilde{r}_{o,i^{\text{P}}}(k)$  are linearly interpolated giving the potential RM rate  $\tilde{r}_o(k)$  when a switching time lies in a time interval:

$$\tilde{r}_o(k) = \sum_{i^{\text{P}}=1}^{n^{\text{pol}}} f_{i^{\text{P}}}^{\text{P}}(k) \tilde{r}_{o,i^{\text{P}}}(k), \quad (16)$$

where the RM rates  $\tilde{r}_{o,i^{\text{P}}}(k)$  of the policies  $i^{\text{P}}$  are given in (17) and the fraction  $f_{i^{\text{P}}}^{\text{P}}(k)$  represents the fraction of the time step that is covered by policy  $i^{\text{P}}$  and which is computed using (18).

The next step is translating the RM rate  $\tilde{r}_o(k)$  to the actual applied RM rate  $r_o(k)$ :

$$r_o(k) = \frac{(1 - \tilde{r}_o(k))q_o^{\text{R,min}}(k) + \tilde{r}_o(k)q_o^{\text{R,max}}(k)}{Q_0}, \quad (19)$$

with the minimum on-ramp flow  $q_o^{\text{R,min}}(k)$  (veh/h) defined by the minimum allowed RM rate and the minimum required RM rate to prevent the on-ramp queue required to prevent the on-ramp queue from exceeding its maximum:

$$q_o^{\text{R,min}}(k) = \max \left[ r^{\text{min}}Q_0, \frac{w_o(k) + d_o(k)T - w_o^{\text{max}}}{T} \right]. \quad (20)$$

The maximum on-ramp flow  $q_o^{\text{R,max}}(k)$  (veh/h) is defined similarly as in (10):

$$q_o^{\text{R,max}}(k) = \min \left[ d_o(k) + \frac{w_o(k)}{T}, Q_0, Q_0 \frac{\rho_m^{\text{max}} - \rho_{m,1}(k)}{\rho_m^{\text{max}} - \rho_m^{\text{crit}}} \right] \quad (21)$$

#### E. Objective function and constraints

The objective of the controller is to minimize the Total Time Spent (TTS) by all the vehicles on the freeway by changing the VSLs and RM rates over the time steps  $k^c = k^{\text{u}}C^{\text{t}} +$



$1, \dots, k^u C^t + N^p$ . The following objective function  $J(k^u)$  expresses the TTS:

$$J(k^u) = T \sum_{\hat{k}=k^u C^u + 1}^{k^u C^u + N^p C^c} \left\{ \sum_{(m,i) \in I^{\text{links}}} \rho_{m,i}(\hat{k}) L_m \lambda_m + \sum_{o \in I^{\text{orig}}} w_o(\hat{k}) \right\}. \quad (22)$$

Here, the set  $I^{\text{links}}$  (-) is the set of indices of all pairs of segments and links, the set  $I^{\text{orig}}$  (-) is the set of all origin indices, and the set  $I^{\text{ramps}}$  is the set of all on-ramp indices.

Using this objective function the MPC optimization problem can be formulated:

$$\begin{aligned} & \min_{\bar{u}(k^u)} J(k^u) \\ & \text{Subject to} \\ & \text{Model: Eq. (3) – Eq. (21),} \\ & \text{Initial states and disturbances:} \\ & \rho_{m,i}(k^u C^u), v_{m,i}^{\text{ctrl}}(k^u C^u), \rho^{\text{DS}}(\hat{k}), d_o(\hat{k}), \\ & \text{Constraints:} \\ & B^L \leq A \bar{u}(k^u) \leq B^U. \end{aligned} \quad (23)$$

The matrix  $A$  and vectors  $B^L$  and  $B^U$  represent the linear inequality constraints on the VSL and RM control signal respectively as detailed in the next subsections. The control signal  $\bar{u}(k^u)$  is a vector consisting of the parameters of the head and tail of the speed-limited area – i.e. the initial location of the head  $x^{\text{H,sl}}(k^u C^u + C^c)$  (km) and tail  $x^{\text{T,sl}}(k^u C^u + C^c)$  (km), and the speed  $v^{\text{H,sl}}(k^c)$  (km/h) and  $v^{\text{T,sl}}(k^c)$  (km/h) of the head and tail over time – and the parameters of the feedback control laws of the different on-ramps – i.e. the switching times  $t_{o,1}^{\text{switch}}(k^u)$ ,  $t_{o,2}^{\text{switch}}(k^u)$ , and  $t_{o,3}^{\text{switch}}(k^u)$ , and the set-points  $\rho_{o,1}^{\text{set}}(k^u)$  and  $\rho_{o,2}^{\text{set}}(k^u)$ .

1) *VSL signal and constraints*: The evolution of the head and tail of the speed-limited area is described by the initial location of the head  $x^{\text{H,sl}}(k^u C^u + C^c)$  and tail  $x^{\text{T,sl}}(k^u C^u + C^c)$ , and the speed  $v^{\text{H,sl}}(k^c)$  and  $v^{\text{T,sl}}(k^c)$  of the head and tail over time respectively. After the control horizon  $N^c$ , until the prediction horizon  $N^p$ , the speed of the head and tail locations are assumed to remain constant:

$$v^{\text{H,sl}}(k^c) = v^{\text{H,sl}}(k^u C^t + N^c) \text{ if } k^c > k^u C^t + N^c, \quad (24)$$

$$v^{\text{T,sl}}(k^c) = v^{\text{T,sl}}(k^u C^t + N^c) \text{ if } k^c > k^u C^t + N^c. \quad (25)$$

Based on the control vector, the location of the head and the tail of the control scheme at every time step  $k$  can be computed:

$$x^{\text{H,sl}}(k) = x^{\text{H,sl}}(k^u C^u + C^c) + \sum_{j=k^u C^u + C^c + 1}^k v^{\text{H,sl}}(\lfloor (j-1)/C^c \rfloor) T, \quad (26)$$

$$x^{\text{T,sl}}(k) = x^{\text{T,sl}}(k^u C^u + C^c) + \sum_{j=k^u C^u + C^c + 1}^k v^{\text{T,sl}}(\lfloor (j-1)/C^c \rfloor) T. \quad (27)$$

Several constraints have to be respected when optimizing the VSLs. First of all, the position of the head and tail have

to lie within the upstream bounds  $x^{\text{H},0}$  (km) and  $x^{\text{T},0}$  (km) and downstream bounds  $x^{\text{H,sl}}(k^u C^u + C^c)$  (km) and  $x^{\text{T,sl}}(k^u C^u + C^c)$  (km):

$$x^{\text{H},0} \leq x^{\text{H,sl}}(k^u C^u + C^c) \leq x^{\text{H,sl}}(k^u C^u + C^c), \quad (28)$$

$$x^{\text{T},0} \leq x^{\text{T,sl}}(k^u C^u + C^c) \leq x^{\text{T,sl}}(k^u C^u + C^c). \quad (29)$$

If at time step  $k^u C^u + C^c$  the speed limits are not active or cover only 1 segment, i.e. when  $x^{\text{H,sl}}(k^u C^u + C^c | k^u - 1) - 1 \leq x^{\text{T,sl}}(k^u C^u + C^c | k^u - 1)$ , then these bounds are equal to the upstream  $x_0$  (km) and downstream end of the freeway  $x_{\text{end}}$  (km). The notation  $(\dots | k^u - 1)$  indicates the control signal that is computed at time step  $k^u - 1$ . However, when the speed limits are active at control step  $k^u C^u + C^c$ , then the location of the head  $x^{\text{H,sl}}(k^u C^u + C^c | k^u)$  and tail  $x^{\text{T,sl}}(k^u C^u + C^c | k^u)$  at control step  $k^u C^u + C^c$  should be equal to the previously computed values  $x^{\text{H,sl}}(k^u C^u + C^c | k^u - 1)$  and  $x^{\text{T,sl}}(k^u C^u + C^c | k^u - 1)$ . In that case, the constraints are set to the following:

$$x^{\text{H},0} = x^{\text{H,sl}}(k^u C^u + C^c | k^u - 1), \quad (30)$$

$$x^{\text{H,sl}}(k^u C^u + C^c | k^u - 1) = x^{\text{H,sl}}(k^u C^u + C^c | k^u - 1), \quad (31)$$

$$x^{\text{T},0} = x^{\text{T,sl}}(k^u C^u + C^c | k^u - 1), \quad (32)$$

$$x^{\text{T,sl}}(k^u C^u + C^c | k^u - 1) = x^{\text{T,sl}}(k^u C^u + C^c | k^u - 1). \quad (33)$$

Secondly, the head and tail are allowed to propagate downstream with at most  $v^{\text{eff}}$  (km/h) or to propagate upstream with any speed so that they cannot ‘overtake’ a speed-limited vehicle:

$$v^{\text{H,sl}}(k^c) \leq v^{\text{eff}}, \quad (34)$$

$$v^{\text{T,sl}}(k^c) \leq v^{\text{eff}}. \quad (35)$$

Thirdly, the position of the head should be at or more downstream than the initial position of the tail:

$$x^{\text{H,sl}}(k) \geq x^{\text{T,sl}}(k). \quad (36)$$

2) *RM constraints*: The RM control signal of an individual on-ramp consists of the switching times  $t_{o,1}^{\text{switch}}(k^u)$ ,  $t_{o,2}^{\text{switch}}(k^u)$ , and  $t_{o,3}^{\text{switch}}(k^u)$ , and the set-points  $\rho_{o,1}^{\text{set}}(k^u)$  and  $\rho_{o,2}^{\text{set}}(k^u)$ . By varying these parameters, the RM rate is affected. Several constraints on these parameters are included. First, it has to hold that the set-points  $\rho_{o,i}^{\text{set}}(k^u)$  should be between 0 and the maximum set-point  $\rho_{o,i}^{\text{set,max}}$  (veh/km/lane):

$$0 < \rho_{o,1}^{\text{set}}(k^u) \leq \rho_{o,1}^{\text{set,max}} \quad (37)$$

$$0 < \rho_{o,2}^{\text{set}}(k^u) \leq \rho_{o,2}^{\text{set,max}}. \quad (38)$$

Secondly, the switching time instants need to be constrained. Two cases are possible. The first case is that no RM is active at time step  $k^c$ . Then, it should hold that:

$$k^u T^u + T^c \leq t_{o,1}^{\text{switch}}(k^u) \leq k^u T^u + N^p T^c \quad (39)$$

$$t_{o,1}^{\text{switch}} + T^c \leq t_{o,2}^{\text{switch}}(k^u) \leq k^u T^u + N^p T^c \quad (40)$$

$$t_{o,2}^{\text{switch}} + T^c \leq t_{o,3}^{\text{switch}}(k^u) \leq k^u T^u + N^p T^c \quad (41)$$

The second case is that RM is active at time step  $k^u$ . This is the case when  $t^{\text{ini}}(k^u) = \max(t_{o,1}^{\text{switch}}(k^u - 1), k^u T^u) < k^u T^u + T^c$  and  $t_{o,3}^{\text{switch}}(k^u - 1) \geq k^u T^u + T^c$ . In this case the MPC should not be able to change  $t_{o,1}^{\text{switch}}(k^u)$  because it

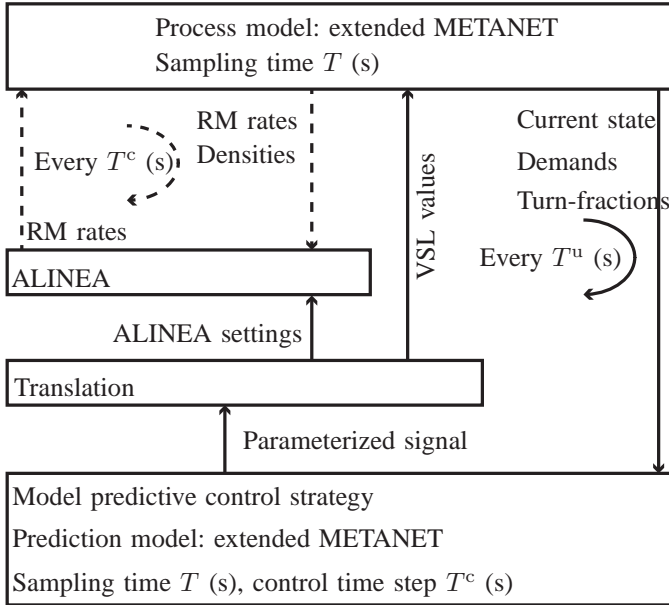


Fig. 4. Simulation set-up

lies within the current time step  $k^u C^t$  that cannot be affected. This is realized by the following constraints:

$$t^{\text{ini}}(k^u) \leq t_{o,1}^{\text{switch}}(k^u) + T^c \leq t^{\text{ini}}(k^u) \quad (42)$$

$$k^u T^u + T^c \leq t_{o,2}^{\text{switch}}(k^u) \leq k^u T^u + N^p T^c \quad (43)$$

$$t_{o,2}^{\text{switch}} + T^c \leq t_{o,3}^{\text{switch}}(k^u) \leq k^u T^u + N^p T^c \quad (44)$$

### III. SIMULATION EXPERIMENTS

Simulations are carried out in order to investigate the controller behavior and performance in terms of CPU time used and TTS improvement of the controller. To this end, several simulations are performed in which the traffic situation and controller set-up is varied.

The main topic for investigation is the trade-off between the computation time and the realized throughput improvement. To this end, the parameterized MPC (PMPC) strategy is compared with a nominal MPC (NMPC) strategy that directly optimizes the individual VSL values and RM rates. The NMPC strategy is expected to realize a similar or higher TTS when given sufficient CPU time. In order to obtain a fair comparison, both the control strategies are given the same CPU time budgets. It is expected that the PMPC strategy is able to obtain similar throughput improvement in less CPU time budget.

Additionally, the performance is compared when considering different controller set-ups, namely RM-only, VSL-only, and integrated RM and VSL, and when applied to different traffic situations, i.e. when resolving a jam wave – as done by the SPECIALIST algorithm – or by preventing congestion due to a high on-ramp flow. This allows to evaluate the added value of integrating the control measures in different scenarios. It is expected that integrated RM and VSL can realize the best throughput improvement because it has a larger control freedom, but that it does not necessarily lead to the best trade-off between computation time and realized throughput.

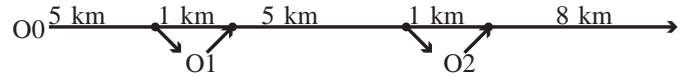


Fig. 5. Simulation network

#### A. Simulation set-up

Figure 4 provides an overview of the simulation set-up. The extended METANET model as detailed in this paper is used as the process model – i.e. the real-world – and the prediction model. When implemented as the process model 3 small changes are made. First of all, the parameter  $\gamma_{m,i}(k)$  is set to 1 in the process model if  $\gamma_{m,i}(k) > 0.1$  such that the entire segment is either speed-limited or not in order to reproduce the discrete spacing of the variable message signs. Secondly, the switching times are rounded to the nearest multiple of  $T$  that is less than or equal to the switching time. Thirdly, a lead-in procedure is introduced for the VSLs preventing too large speed drops on the freeway. To this end, the VSL value of a gantry is set to the minimum of the desired VSL and the VSL value of the downstream gantry increased with 10 km/h which is iteratively computed from downstream to upstream.

A 20 km long freeway with 2 on-ramps and 2 off-ramps is considered as shown in Figure 5. The freeway consists of three origins and 20 identical segments with a length of 1 km and 2 lanes. Every segment has the same parameters, adopted from [32], namely:  $T = 10$  s,  $\tau = 18$  s,  $\kappa = 40$  veh/km/lane,  $\rho^{\text{crit}} = 33.5$  veh/km/lane,  $a_m = 1.867$ ,  $v^{\text{free}} = 102$  km/h,  $\eta^{\text{high}} = 65$  km/h<sup>2</sup>,  $\eta^{\text{low}} = 30$  km/h<sup>2</sup>. Using these parameters, a capacity of 2000 veh/h/lane is realized and a capacity drop can be observed. The freeway traffic is simulated for scenarios of 3 hours. All the segments can be controlled by means of VSLs. The value of the effective speed limit  $v^{\text{eff}}$  is set to 50 km/h. The two on-ramps are controlled by means of RM. The minimum RM rate is set to 0.05, the feedback gains of the PMPC strategy are set to  $K_{o,ip}^s = 0.5$ , and the maximum density set-point is set to  $\rho_{o,ip}^{\text{set,max}} = 60$  veh/km/lane.

The process and prediction model sampling time steps  $T$  are set to 10 seconds. The control signal update time step  $T^u$  is set to 300 seconds, and the control time step  $T^c$  is set to 60 seconds. This means that every 300 seconds the control signal is optimized based on the current traffic state. The values of the control signals are allowed to change every 60 seconds.

It is assumed that no measurement noise affects the traffic state used by the MPC strategy. Also, it is assumed that a prediction of the demand and turn-fractions is available for the MPC strategy.

The evaluation is carried out using Matlab R2015a on a computer with a 3.6 GHz processor, 8 cores, and 16 Gb RAM. For the optimization the Sequential Quadratic Programming algorithm of the fmincon solver of the MATLAB optimization toolbox is used, the function tolerance is set to  $5 \cdot 10^{-3}$  and the step tolerance is set to  $1 \cdot 10^{-7}$ . Parallel computing on 8 cores is used to determine the numerical derivative of the objective function. In order to realize a fair comparison, both approaches are given the same amount of CPU time in which they can find the optimal solution. When this computation time is not reached, the optimization is repeated from a new,

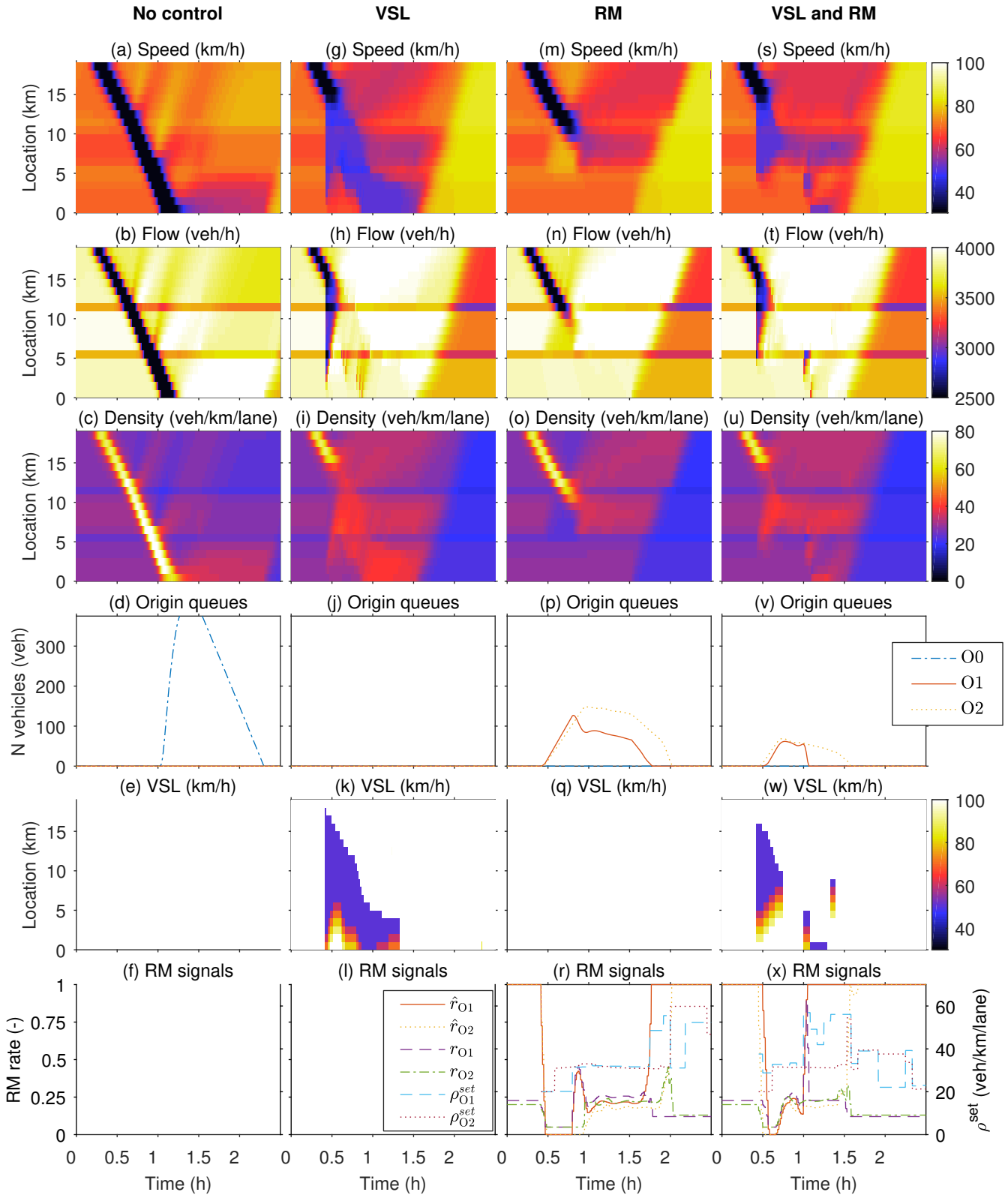


Fig. 6. Results of the jam wave case using a CPU time budget of 3600 s. Every column represent a different control strategy. The first three rows show the contour plots of the traffic dynamics, the fourth row shows the origin queues, and the bottom 2 rows show the control signals.

randomly selected starting point. When the computation time is exceeded, the optimization is stopped. The best solution out of the different starts is applied to the process. All the

simulations are repeated with three budgets, namely 300, 600, 1200, 1800, and 3600 seconds. To speed up the simulations, parallel computing is used to compute the gradient. For a fair

comparison, the CPU time budget reflects the total computation time used by all the cores. Because the computations are carried out in parallel, the actual elapsed time is smaller than the CPU time budget.

The NMPC approach is implemented as follows. Similar as in [20], the RM rate  $\tilde{r}_o(k)$  of an on-ramp is directly optimized. It is bounded between 0 and 1 and constrained in such a way that the RM can change with a maximum of 0.25 per control step. The optimized RM rate  $\tilde{r}_o(k)$  is translated to the actual applied RM rate  $r_o(k)$  using (19). The VSL strategy proposed in [21] is implemented. The VSL values are bounded so that they are larger than 50 km/h and smaller than the free flow speed. Additionally, the following constraint is included  $v_{m,i}^{\text{ctrl}}(k^c) \leq v_{m,i+1}^{\text{ctrl}}(k^c) + 10$  preventing sudden speed drops in the downstream direction.

### B. Case I: jam wave

A scenario in which a jam wave is present on the freeway is evaluated. Figure 6 (a)–(f) shows the no-control situation in which a jam wave enters the freeway at the most downstream end. This jam is created by increasing the density at the downstream end of the freeway from time 380 s to 1080 s. The demand at the origins are equal to 3800 veh/h, 455 veh/h, and 400 veh/h until time 5500 s for the mainstream origin (O0), on-ramp 1 (O1) and on-ramp 2 (O2) respectively. The percentage of traffic exiting at the off-ramps is 10% and 12% for off-ramp 1 and 2 respectively. After time 5500 s the demands decrease to 3500 veh/h, 240 veh/h, and 260 veh/h respectively. The capacity drop due to this jam wave, determined using simulation experiments, is approximately 5.6%. The total time spent of the no-control scenario is 3325.1 veh-h.

Various control set-ups are tested in the control scenario. In order to evaluate the performance and behavior of the MPC strategy when resolving a fully formed jam wave, so that we can interpret the solution, which is expected to be similar to the solution of SPECIALIST, the controller is started after 1500 seconds. Note that this represents an artificial situation, since in practice the MPC strategy is always active so that it will start controlling before the jam wave is fully formed. The maximum on-ramp queue length is set to 150 vehicles for both ramps. The prediction horizon is set to 4800 seconds and the control horizon is set to 2400 seconds. The control horizon is not applicable to the parameterized RM strategy, because a specific choice of the switching times fully determines the controller behavior over any horizon. Note that the VSL control signal is allowed to change every 60 seconds so that 40 steps are to be optimized.

The control horizon is not relevant for the parameterized RM strategy, since it optimizes the switching time instants when the feedback RM strategy is changes instead of explicitly optimizing the RM rates at the control sampling time steps.

Table I presents the quantitative results for the different computation time budgets. It can be observed that a computation time budget of 1200 seconds is sufficient for all the parameterized strategies to realize the best throughput improvement. The average elapsed times per controller update for these budgets are all below 300 seconds. The RM-only

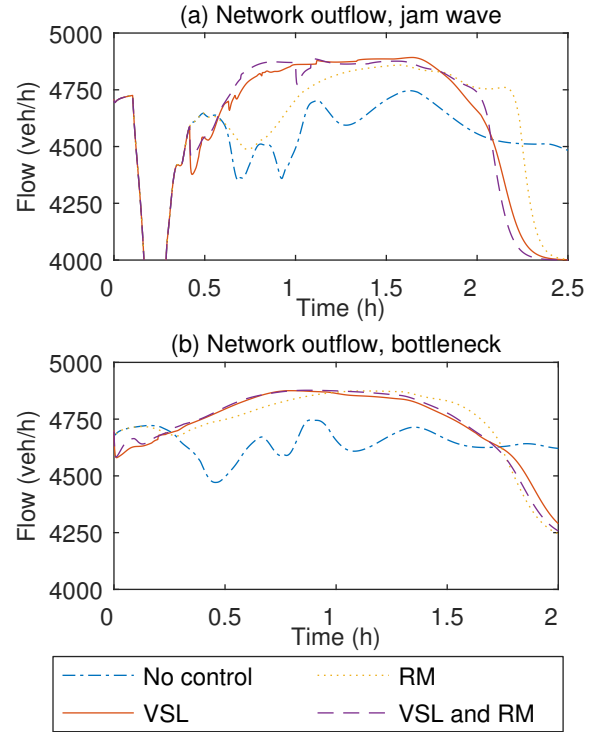


Fig. 7. Network outflow in (a) the jam wave and (b) the bottleneck case using different control strategies.

NMPC strategy achieves similar performance as the PMPC strategy for a budget of 3600 seconds. However, even the budget of 3600 seconds does not seem to be enough for VSL-only or integrated VSL and RM using the NMPC strategy.

The qualitative results of the VSL-only case shown in Figures 6 (g)–(l) show that the jam wave is resolved by imposing a speed-limited area upstream of the jam wave, similar as done by the SPECIALIST algorithm. Figures 6 (m)–(r) show the results of resolving the jam wave using the RM-only strategy. It can be seen that it takes longer for the RM-only to resolve the jam wave explaining the lower TTS gain. Figures 6 (s)–(x) show that the integration of VSLs and RM reduces the application of VSLs upstream of on-ramp 1, as well as the time over which VSLs are needed. Figure 7 (a) shows the total network outflows for the different control strategies. It can be seen that it takes longer for the RM-only strategy to improve the total network outflow. Also, it can be seen that the integration of VSL and RM reduces the initial outflow reduction and a quicker outflow increase after the jam has resolved when compared to the VSL-only case, explaining the TTS improvement.

### C. Case II: bottleneck

The second case consists of a traffic jam caused by a too high on-ramp flow. The no control situation is shown in Figure 8 (a)–(f). The origin demands were set to 3900 veh/h, 455 veh/h, and 390 veh/h for the mainstream origin (O0), on-ramp 1 (O1), and on-ramp 2 (O2) respectively, and they were taken to be constant until time 4500 s except for on-ramp 1 of which the inflow from time 1500 s to 2000 s was increased to



TABLE I

OVERVIEW OF QUANTITATIVE RESULTS FOR BOTH CASES. THE NO CONTROL TTS IS 3325.1 VEH-H FOR THE JAM WAVE CASE AND 2536.0 VEH-H FOR THE BOTTLENECK CASE. THE PERCENTAGE GAIN IN TTS FOR THE CLOSED-LOOP SIMULATION COMPARED TO THE NO CONTROL SITUATION AND THE AVERAGE ELAPSED TIME (ET) PER ITERATION ARE PRESENTED.

		CPU budget: 300 s		CPU budget: 600 s		CPU budget: 1200 s		CPU budget: 1800 s		CPU budget: 3600 s		
		% gain	ET (s)	% gain	ET (s)	% gain	ET (s)	% gain	ET (s)	% gain	ET (s)	
Jam wave	PMPC	VSL	3.9	57.7	11.1	103.3	11.1	207.8	11.1	311.7	11.2	621.5
		RM	7.2	74.6	7.3	148.8	7.3	297.9	7.3	446.6	7.3	890.4
		VSL RM	10.8	57.6	11.2	110.1	11.6	209.5	11.7	316.7	11.9	624.6
	NMPC	VSL	1.0	54.4	5.5	127.7	8.7	181.0	9.5	291.9	10.1	521.1
		RM	4.1	48.0	4.5	94.7	5.2	179.7	5.2	267.3	7.3	528.3
		VSL RM	0.8	73.4	8.1	173.1	8.1	173.0	9.8	294.5	11.2	554.8
Bottleneck	PMPC	VSL	4.7	45.9	5.0	97.3	9.3	194.7	9.2	299.4	9.2	604.8
		RM	9.7	74.2	9.7	147.9	9.7	295.5	9.7	443.2	9.7	883.0
		VSL RM	9.1	46.3	9.6	88.6	9.8	195.1	10.0	293.3	10.0	605.5
	NMPC	VSL	-3.1	54.3	3.8	125.7	5.4	191.9	6.8	285.9	9.2	531.5
		RM	9.9	48.6	9.9	91.2	9.9	171.3	9.9	260.2	9.9	508.2
		VSL RM	-3.6	73.6	2.7	172.3	2.7	180.9	5.3	316.6	6.8	550.3

1500 veh/h triggering a traffic jam. The percentage of traffic exiting at the off-ramps is 10% and 12% for off-ramp 1 and 2 respectively. After time 4500 s the demands decreased to 3500 veh/h for the mainstream origin and to 260 veh/h for on-ramp 2. The resulting TTS is 2536.0 veh-h.

Several control set-ups are evaluated in the control situation. The maximum on-ramp queue lengths were set to 75 and 20 vehicles for on-ramp 1 and 2 respectively. Due to this, coordination between the two on-ramps is required, since the capacity of on-ramp 2 is not sufficient to prevent congestion on its own. The prediction horizon is set to 4800 seconds and the control horizon is set to 2400 seconds.

Table I presents the quantitative results for the different computation time budgets. It can be observed that for VSL-only and integrated VSL and RM set-ups the PMPC realizes higher TTS gains in shorter budgets. For the RM-only case, both the NMPC and PMPC realize similar TTS improvements, namely 9.9% and 9.7% respectively for short CPU time budget of 300 seconds.

The qualitative results of the VSL-only strategy shown in Figures 8 (g)–(l) indicate the ability to prevent bottleneck congestion by imposing a speed-limited area upstream of on-ramp 2. Figures 8 (m)–(r) show that in the RM-only case on-ramp 1 starts metering immediately so that this flow reduction arrives at on-ramp 2 when the demand increases. The qualitative results of the integrated VSL and RM case in Figures 8 (s)–(x) indicate that the integration of VSL and RM reduces the extent to which VSLs are imposed upstream of on-ramp 1. When comparing the outflow plots in Figure 7 (b) it can be seen that the integrated VSL and RM strategy limits the initial flow reduction when compared to the VSL-only strategy. It also shows that integrated VSL and RM is able to quicker restore the network outflow when compared to RM-only.

#### IV. CONCLUSIONS AND RECOMMENDATIONS

In this paper the computation time of an MPC strategy for integrated RM and VSLs was improved considerably by parameterizing a control scheme based on ALINEA ramp metering and a SPECIALIST-like VSL control scheme. Due to this, the dimension of the optimization problem has become independent of the number of VSL signs. Additionally, the

number of parameters needed per on-ramp has become independent of the prediction horizon. Simulations have shown that the control approach proposed in this paper can achieve a better performance than a non-parameterized MPC strategy when using the same budget of computation time for VSL-only and integrated VSL and RM strategies. It was found that the non-parameterized strategy realizes a slightly better throughput improvement for the RM-only case.

In further research, the impact of noise and uncertainties on controller performance can be studied. When needed, a robust control design may have to be designed. Additionally, it can be studied how the approach can be extended to include multiple VSL areas when applying it to larger freeway networks. It is also recommended to compare the proposed strategy to simpler, uncoordinated or non-predictive strategies. Also, the use of in-vehicle technologies may lead to improved detection and actuation possibilities and potentially a reformulation of the control strategy. Future research can also investigate approaches to further improve the computation time, for instance, using a problem-tailored algorithm to solve the optimization problem as discussed in [33].

#### ACKNOWLEDGMENT

This work is part of the research programme ‘The Application of Operations Research in Urban Transport’, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO).

#### REFERENCES

- [1] F. L. Hall and K. Agyemang-Duah, “Freeway capacity drop and the definition of capacity,” *Transportation research record*, no. 1320, 1991.
- [2] B. S. Kerner and H. Rehborn, “Experimental features and characteristics of traffic jams,” *Physical Review E*, vol. 53, no. 2, p. R1297, 1996.
- [3] A. Hegyi, S. Hoogendoorn, M. Schreuder, and H. Stoelhorst, “Dynamic speed limit control to resolve shock waves on freeways - field test results of the SPECIALIST algorithm,” in *Proceedings of the 13th International IEEE Conference on ITS*, (Madeira, Portugal), pp. 519–524, 2010.
- [4] S. Smulders, “Control of freeway traffic flow by variable speed signs,” *Transportation Research Part B-Methodological*, vol. 24, no. 2, pp. 111–132, 1990.
- [5] 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, England), pp. 145–149, 1994.

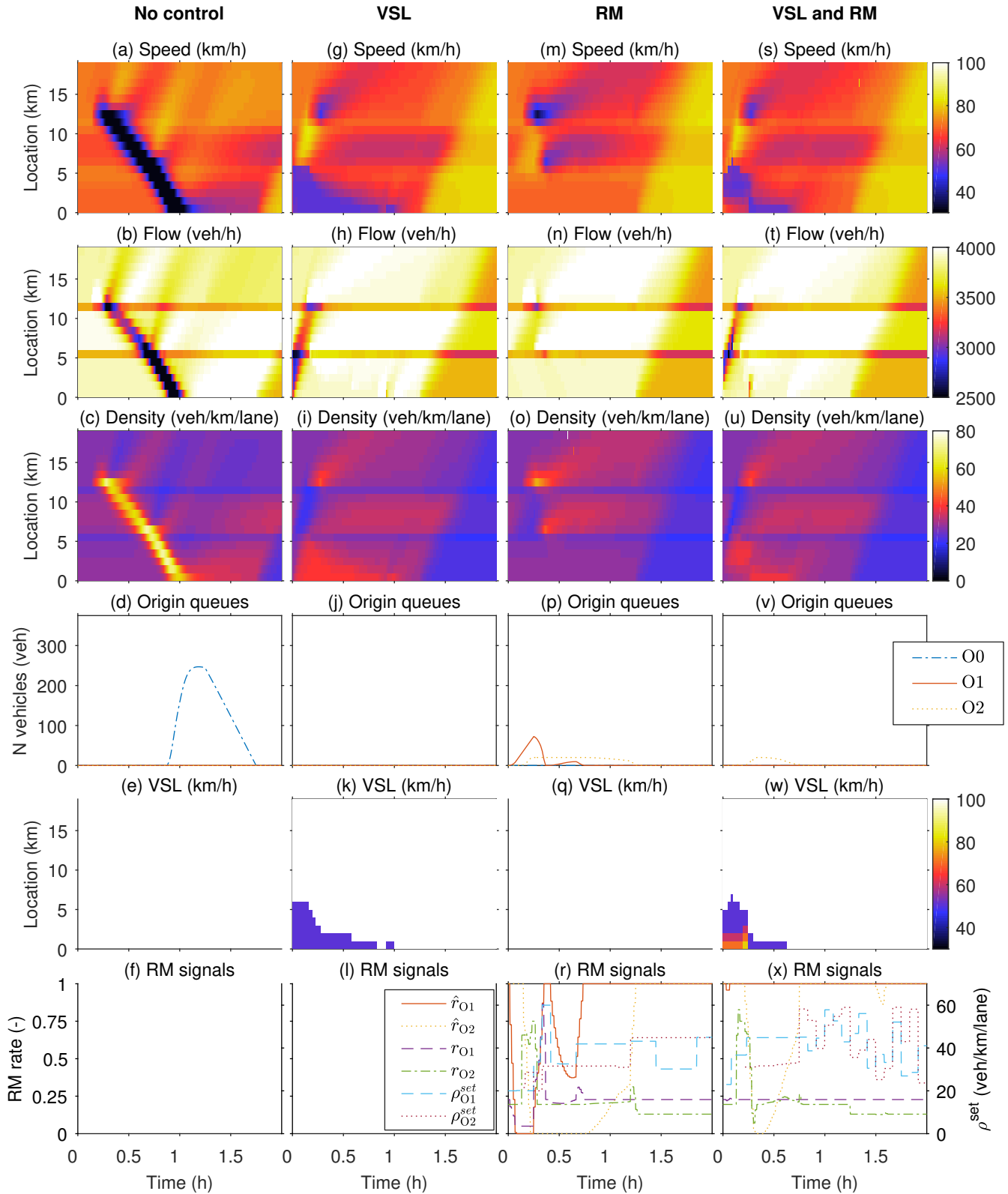


Fig. 8. Results of the bottleneck case using a CPU time budget of 3600 s. Every column represent a different control strategy. The first three rows show the contour plots of the traffic dynamics, the fourth row shows the origin queues, and the bottom 2 rows show the control signals.

[6] R. D. Kühne, "Freeway control using a dynamic traffic flow model and vehicle reidentification techniques," *Transportation Research Record*, no. 1320, pp. 251–259, 1991.

[7] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Local feedback-

based mainstream traffic flow control on motorways using variable speed limits," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1261–1276, 2011.

[8] N. Mahajan, A. Hegyi, G. S. van de Weg, and S. P. Hoogendoorn,

- “Integrated variable speed limit and ramp metering control against jam waves a COSCAL v2 based approach,” in *Proceedings of the 17th International Conference on ITS*, (Las Palmas, Spain), pp. 1156–1162, 2015.
- [9] D. Chen, S. Ahn, and A. Hegyi, “Variable speed limit control for steady and oscillatory queues at fixed freeway bottlenecks,” *Transportation Research Part B: Methodological*, vol. 70, pp. 340–358, 2014.
- [10] Y. Zhang and P. A. Ioannou, “Combined variable speed limit and lane change control for highway traffic,” *IEEE Transactions on Intelligent Transportation Systems*, 2016.
- [11] H. Hadj-Salem, J. Blossville, and M. Papageorgiou, “Alinea: A local feedback control law for on ramp metering; a real life study,” in *Proceedings of the Third International Conference on Road Traffic Control*, pp. 194–198, 1990.
- [12] F. Middelham and H. Taale, “Ramp metering in the Netherlands: an overview,” in *Proceedings of the 11th IFAC Symposium on Control in Transportation Systems*, vol. 39, pp. 267–272, 2006.
- [13] I. Papamichail and M. Papageorgiou, “Traffic-responsive linked ramp-metering control,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 111–121, 2008.
- [14] R. C. Carlson, I. Papamichail, and M. Papageorgiou, “Integrated feedback ramp metering and mainstream traffic flow control on motorways using variable speed limits,” *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 209–221, 2014.
- [15] A. Kotsialos and M. Papageorgiou, “Efficiency and equity properties of freeway network-wide ramp metering with amoc,” *Transportation Research Part C: Emerging Technologies*, vol. 12, no. 6, pp. 401–420, 2004.
- [16] I. Schelling, A. Hegyi, and S. Hoogendoorn, “SPECIALIST-RM - integrated variable speed limit control and ramp metering based on shock wave theory,” in *Proceedings of the 14th International IEEE Conference on ITS*, (New York, USA), pp. 2154–2159, 2011.
- [17] G. S. van de Weg, A. Hegyi, H. Hellendoorn, and S. E. Shladover, “Cooperative systems based control for integrating ramp metering and variable speed limits,” in *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, 2014.
- [18] J. Rawlings and D. Mayne, *Model Predictive Control: Theory and Design*. Madison, Wisconsin: Nob Hill Publishing, 2009.
- [19] M. Burger, M. Van Den Berg, A. Hegyi, B. De Schutter, and J. Hellendoorn, “Considerations for model-based traffic control,” *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 1–19, 2013.
- [20] A. Kotsialos, M. Papageorgiou, and F. Middelham, “Local and optimal coordinated ramp metering for freeway networks,” *Journal of ITS*, vol. 9, no. 4, pp. 187–203, 2005.
- [21] A. Hegyi, B. De Schutter, and H. Hellendoorn, “Optimal coordination of variable speed limits to suppress shock waves,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 102–112, 2005.
- [22] G. Gomes and R. Horowitz, “Optimal freeway ramp metering using the asymmetric cell transmission model,” *Transportation Research Part C-Emerging Technologies*, vol. 14, no. 4, pp. 244–262, 2006.
- [23] M. Hajiahmadi, G. S. van de Weg, C. Tampère, R. Corthout, A. Hegyi, B. De Schutter, and H. Hellendoorn, “Integrated predictive control of freeway networks using the extended link transmission model,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 65–78, 2015.
- [24] J. R. D. Frejo and E. F. Camacho, “Global versus local MPC algorithms in freeway traffic control with ramp metering and variable speed limits,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1556–1565, 2012.
- [25] S. K. Zegeye, B. De Schutter, H. Hellendoorn, E. A. Breunese, and A. Hegyi, “A predictive traffic controller for sustainable mobility using parameterized control policies,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1420–1429, 2012.
- [26] X.-Y. Lu, P. Varaiya, R. Horowitz, D. Su, and S. Shladover, “Novel freeway traffic control with variable speed limit and coordinated ramp metering,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2229, pp. 55–65, 2011.
- [27] G. S. van de Weg, A. Hegyi, S. P. Hoogendoorn, and B. De Schutter, “Efficient model predictive control for variable speed limits by optimizing parameterized control schemes,” in *Proceedings of the 17th International Conference on ITS*, (Las Palmas, Spain), pp. 1137–1142, 2015.
- [28] G. S. van de Weg, A. Hegyi, and S. P. Hoogendoorn, “Ex-ante data analysis approach for assessing the effect of variable speed limits,” in *Proceedings of the 17th International Conference on ITS*, (Qindao, China), pp. 1317–1322, 2014.
- [29] M. J. Lighthill and G. B. Whitham, “On kinematic waves, II. a theory of traffic flow on long crowded roads,” in *Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences*, vol. 229A, pp. 317–345, 1955.
- [30] M. Papageorgiou, J. M. Blossville, and H. Hadj-Salem, “Modeling and real-time control of traffic flow on the southern part of boulevard-peripherique in Paris part 2: Coordinated on-ramp metering,” *Transportation Research Part A-Policy and Practice*, vol. 24, no. 5, pp. 361–370, 1988.
- [31] E. Smaragdis, M. Papageorgiou, and E. Kosmatopoulos, “A flow-maximizing adaptive local ramp metering strategy,” *Transportation Research Part B: Methodological*, vol. 38, no. 3, pp. 251–270, 2004.
- [32] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, “Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool metanet,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 282–292, 2002.
- [33] A. Kotsialos, M. Papageorgiou, M. Mangeas, and H. Haj-Salem, “Coordinated and integrated control of motorway networks via non-linear optimal control,” *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 1, pp. 65–84, 2002.

**Goof Sterk van de Weg** received his Ph.D. degree from Delft University of Technology (TU Delft) in the Netherlands in 2017 and his M.Sc. degree in Systems and Control in 2013 from the TU Delft as well. His main research interest is the design of control algorithms for cooperative systems, traffic lights, variable speed limits, ramp metering, and route guidance to improve the performance of road traffic networks.



**Andreas Hegyi** is an Assistant Professor at TU Delft in the Netherlands. He received his M.Sc. degree in Electrical Engineering in 1998 and the Ph.D. in 2004, both from the TU Delft, The Netherlands. From 2004 to 2007 he was a postdoctoral researcher at TU Delft and at Ghent University. He is the author or coauthor of over 100 papers. He is a member of IEEE and IEEE-ITSS, member of IFAC-CTS, and has served as Program Chair of the IEEE-ITSC 2013 conference and as General Chair of the IXth TRISTAN symposium 2016, and IPC member of various other conferences. Dr. Hegyi is Associate Editor of IEEE Transactions on Intelligent Transportation Systems and member of the Editorial Board of Transportation Research Part C. His research interests are in the areas of Traffic Flow Modeling and Control, Connected and Cooperative Vehicles, Traffic State Estimation, and Traffic Data Analysis.



**Serge Paul Hoogendoorn** is a full professor at the TU Delft, where he holds the chair Traffic Operations and Management. He is also Distinguished Professor Smart Urban Mobility at the same university. He received his M.Sc. degree in Applied Mathematics in 1995 and his PhD in Civil Engineering in 1999. He is (co-)author of over 280 papers. He is a member of the TRB Traffic Flow Theory Committee and chairs the TRB Crowd Modeling and Management subcommittee. He is IAC member of the ISTTT. He is editor of the Journal of Advanced Transportation, EJTL and EJTR. In 2014, he received an ERC Advanced Grant. His research interest cover a variety of topics, including traffic flow theory, traffic management, ITS, and active mode mobility.



**Bart De Schutter** (IEEE member since 2008, senior member since 2010) is a full professor at the Delft Center for Systems and Control of TU Delft in Delft, The Netherlands. He is senior editor of the IEEE Transactions on Intelligent Transportation Systems. His current research interests include intelligent transportation and infrastructure systems, hybrid systems, and multi-level control.

