Delft University of Technology

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

Technical report CSE02-010

Model predictive control with repeated model fitting for ramp metering*

T. Bellemans, B. De Schutter, and B. De Moor

If you want to cite this report, please use the following reference instead:
T. Bellemans, B. De Schutter, and B. De Moor, "Model predictive control with repeated model fitting for ramp metering," *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems (ITSC'02)*, Singapore, pp. 236–241, Sept. 2002.

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

Model predictive control with repeated model fitting for ramp metering

Tom Bellemans, Bart De Schutter, Bart De Moor

***Abstract*—In this paper we deal with model predictive control for ramp metering on motorways. A discussion of the way ramp metering tries to prevent, or postpone, congestion on a motorway is presented. As an example, a real-life motorway in Belgium is presented. We discuss a traffic flow model that is used in a receding horizon framework and we argue that the quality of the fit of the traffic flow model to the measurement data has an impact on the quality of the optimised metering signals. Therefore, we suggest to re-fit the most sensitive parameters of the traffic flow model on a regular basis. We conclude this paper with some simulation results.**

***Index Terms*—Model predictive control, traffic control, ramp metering, identification**

I. Introduction

In this paper we discuss optimal ramp metering as a means to postpone or, ideally, prevent congestion on motorways. As an example, we look at a real-life situation on the E17 motorway Ghent–Antwerp in Belgium. We present how ramp metering tries to address the problems that can occur in this network. Next, we discuss a model predictive approach for ramp metering. After a description of the model and the cost criterion used, we discuss model identification. Since the traffic situation on motorways is influenced by various factors that are not all incorporated into the model, we argue that a regular re-fitting of the motorway model improves the accuracy of the control. We illustrate this with some simulation results.

II. Problem description

Recurrent congestion during rush hours is a common problem on motorways around the globe. In this section we discuss how ramp metering can help to improve the traffic situation on the motorways [1]. When no ramp metering is applied, cars can try to enter the motorway as they wish. During rush hours, cars that want to enter the motorway quite often form small platoons (due to traffic lights, slower cars, ...). These platoons cause important disturbances of the traffic on the motorway because they have to merge in the traffic flow on the mainline. Ramp metering is implemented by placing a traffic light at the on-ramp that allows the vehicles to enter the motorway in a controlled way and thus reduces the disturbance of the traffic in the mainline.

Tom Bellemans and Bart De Moor are with ESAT-SCD (Signals, Identification, System Theory and Automation), Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. email: {Tom.Bellemans, Bart.DeMoor}@esat.kuleuven.ac.be

Bart De Schutter is with the Control Systems Engineering group, Faculty of Information Technology and Systems, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands. email: b.deschutter@its.tudelft.nl

The idea behind ramp metering follows from observations of the average traffic flow (throughput) and the average traffic density in a motorway section. The observed relationship between average traffic density and average traffic flow is plotted in Figure 1. While the average density in the section is smaller than a critical value c_{cr} the flow increases with increasing density. At c_{cr} the flow is maximal and with further increasing density, the traffic flow starts to decrease again. As the maximal average density c_{jam} is reached, the traffic comes to a halt and the flow becomes zero. This relationship between traffic density and traffic flow is known in traffic literature as the fundamental diagram [2].

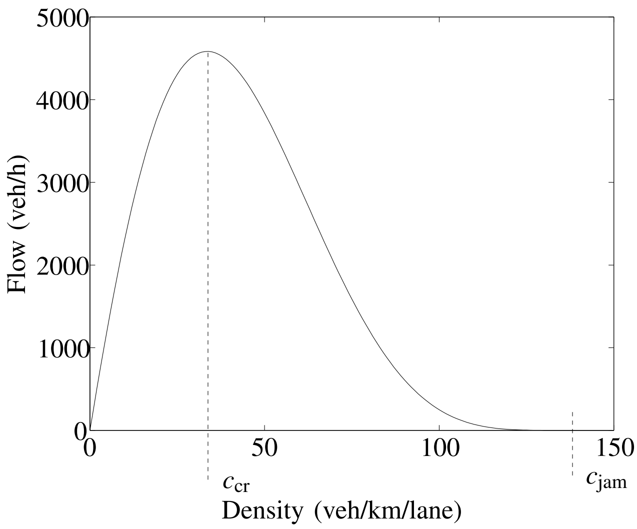

Fig. 1. The flow–density relation of the traffic in a motorway section which is also known as the fundamental diagram.

From Figure 1 we conclude that the traffic flow on the motorway is optimal for the critical density c_{cr}. A traffic situation with a vehicle density larger than c_{cr} is unstable, since a disturbance that temporarily increases the density will result in a reduced flow which in its turn causes a further increase of the density. Ramp metering will try to keep the point of operation in the stable region by limiting the metering rate in an attempt to prevent the density on the motorway to grow larger than c_{cr} [3], [4].

Figure 2 presents a schematic representation of the real-life motorway E17 Ghent–Antwerp in Belgium. The studied motorway is approximately 15 kilometres long and counts three lanes. There are four off-ramps and five on-ramps. In these complex configurations, ramp metering can prevent or delay grid locks by limiting the inflow into sections where congestion would occur if no ramp metering were applied. Consider e.g. congestion on the motorway that is caused by traffic from the third on-ramp. As congestion spills further and further back, it eventually blocks the third off-ramp, which in its turn causes the congestion to spill back even more. It is clear that preventing this scenario can result in important efficiency improvements. In Section V, we will simulate ramp metering and investigate the impact of a model misfit on the performance but first we will discuss model predictive control of ramp metering in the next section.

III. Model predictive control approach

In this section we describe the model predictive control [5], [6] based approach to ramp metering. The control signals (metering rates) we find using model predictive control are obtained by minimising a cost function over a prediction horizon N_p using a traffic flow model. In order to reduce the computational complexity of the optimisation, we allow the metering rates to change only during the control horizon N_c ($N_\mathrm{c} \leq N_\mathrm{p}$). After the control horizon, the metering rate is kept constant for the remainder of the prediction horizon. In a receding horizon framework, only the first sample of the calculated metering rates is implemented while the others are discarded and recalculated during the next iteration. Once the metering rate is implemented, the process starts all over again with the control and the prediction horizon shifted one sample forward.

The parameters N_p and N_c are chosen with the following trade-offs in mind. The larger N_p, the larger the time horizon we look ahead. This allows us to foresee certain events, e.g. a queue spilling back in front of an off-ramp, ... but it also increases the computational complexity. Taking into account that we want to implement optimal ramp metering in an on-line framework, we see that N_p is limited from above by the available time to do the calculations. For the length of the control horizon N_c a similar trade-off needs to be made. Since N_c determines the number of parameters (metering rates) that need to be optimised and since the computational complexity of the optimisation increases strongly with the number of parameters, we need to find a trade-off between performance and computational complexity.

It is important to note that the sampling rate of the controller will in practice be lower than the discretisation step of the discretised traffic flow model. As we will see in Section III-A, a typical discretisation step for the traffic flow model is one sample every 10 to 15 seconds ($\Delta T = 10-15$ sec). The metering rate does not need updating every 10 seconds since the average dynamics of the traffic system change much more slowly. Therefore we can choose the sampling rate of the controller to be one sample per minute or even less.

In the remainder of this section, we discuss the motorway traffic flow model and the cost function in more detail.

A. The motorway traffic flow model

The motorway traffic flow model that we discuss in this section was originally designed by Payne [7] and some additions were made by Papageorgiou [8]. The model is a second order traffic flow model that is discretised in both time and space. E.g. in the case of the motorway Ghent–Antwerp from Figure 2, the motorway is divided in 29 sections of 500 meter and the discretisation step ΔT is chosen to be 10 seconds. For more information regarding these choices we refer to [9].

The behaviour of each of the motorway sections can be described using the following equations.

The first equation expresses the conservation of the number of vehicles on the motorway. This conservation law states that the density in section j at time $k+1$ denoted as $c_j(k+1)$ depends on the density $c_j(k)$ in section j at time k and the net inflow in section j during the time interval $[k\Delta T, (k+1)\Delta T)$:

$$c_j(k+1) = c_j(k) + \frac{\Delta T}{n_j l_j}[q_{\mathrm{in},j}(k) - q_{\mathrm{out},j}(k)] \tag{1}$$

where $q_{\mathrm{in},j}(k)$ and $q_{\mathrm{out},j}(k)$ are the inflow and the outflow of section j in the time interval $[k\Delta T, (k+1)\Delta T)$ respectively, while n_j is the number of lanes and l_j is the length of the section.

The average speed $v_j(k+1)$ in section j at time $k+1$, expressed by equation (2), is the average speed $v_j(k)$ at the previous time k altered by three terms representing the following phenomena: relaxation, convection and anticipation:

$$\begin{aligned} v_j(k+1) = {} & v_j(k) \\ & + \frac{\Delta T}{\tau}\left[V[c_j(k)] - v_j(k)\right] && \text{Relaxation} \\ & + \frac{\Delta T}{l_j} v_j(k)\left[v_{j-1}(k) - v_j(k)\right] && \text{Convection} \\ & - \frac{\nu \Delta T [c_{j+1}(k) - c_j(k)]}{\tau l_j [c_j(k) + \kappa]}. && \text{Anticipation} \end{aligned} \tag{2}$$

The relaxation term in (2) expresses that vehicles in a motorway section tend to obtain a desired average speed $V[c_j(k)]$ which depends on the traffic density $c_j(k)$ in the section. The cars adapt to this desired average speed with time constant τ. An empirical formula for the speed–density relationship is given by [2]:

$$V[c_j(k)] = v_\mathrm{f} \exp\left(-\frac{1}{a_m}\left(\frac{c_j(k)}{c_{\mathrm{cr},j}}\right)^{a_m}\right). \tag{3}$$

The convection term takes into account that vehicles that travel from one section to the next need some time to adapt their speed to the desired average speed in the new section. By consequence, vehicles entering a section bias the value of the average speed in their new section towards the average speed in the previous section. This bias is proportional to the difference in average speed between both sections and is described by the convection term in (2).

If a driver sees a higher density ahead, he will decelerate. The last term in (2) expresses this anticipation of the drivers to the density that lies ahead. The anticipation term depends on the density in the current and the first downstream section. The model parameters τ, ν and κ can be fitted as discussed in Section IV.

The flow $q_j(k)$ is expressed as the product of the density $c_j(k)$, the average speed $v_j(k)$ and the number of lanes n_j:

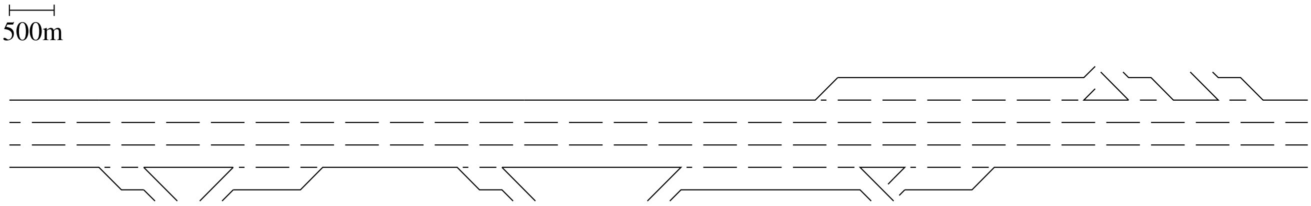

Fig. 2. Schematic representation of the E17 motorway Ghent–Antwerp in Belgium.

$$q_j(k) = c_j(k)v_j(k)n_j. \tag{4}$$

Equations (1), (2), (3) and (4) are a description of the behaviour of the motorway sections. Since we will use the model to optimise ramp metering, we also need a description of the behaviour of traffic at the on-ramps. The on-ramps can be modelled as a queue resulting in the following equation:

$$w_m(k+1) = w_m(k) + \Delta T\left(D_m(k) - q_{o,m}(k)\right). \tag{5}$$

The queue length $w_m(k)$ changes according to the difference between the traffic demand $D_m(k)$ and the service rate $q_{o,m}(k)$ of the on-ramp. The service rate of the on-ramp is the minimum of the number of cars that want to enter and the number of cars that can enter the motorway. This leads to:

$$\begin{aligned} q_{o,m}(k) \quad = \quad & \min\left[D_m(k) + \frac{w_m(k)}{\Delta T},\right. \\ & \left. Q_m \min\left(r_m(k), \frac{\rho_{\max,j} - \rho_j(k)}{\rho_{\max,j} - \rho_{\mathrm{cr},j}}\right)\right], \end{aligned} \tag{6}$$

where Q_m is the maximal capacity of the on-ramp (veh/h) and $\rho_{\max,j}$ is the maximal possible density in the section the on-ramp feeds into (here section j). Through the metering rate $r_m(k)$, we can limit the service rate of the on-ramp. The metering rate $r_m(k)$ theoretically lies in the interval $[0,1]$, but often a lower bound is imposed on the metering rate such that $r_m(k) \in [r_{\min}, 1]$.

B. Control objective and receding horizon control

Now that we have a description of the motorway system including the on-ramps, we need to define a cost function that expresses the performance of the traffic situation on the motorway.

We suggest to use the total time spent by all the vehicles in the system under study combined with a penalty term for variations of the control signal [10]. The total time spent takes the vehicles in the different sections of the motorway as well as the vehicles in the queues at the on-ramps into account. This way, we try to make a fair trade-off between the time spent by vehicles in the queues at the on-ramps and the time spent by vehicles on the motorway. Since we are using model predictive control, we work in a receding horizon framework resulting in the following definition of the cost function: the cost function at time k_0 is the total time spent by the cars in the network during the time interval $[k_0\Delta T, (k_0+N_\mathrm{p})\Delta T)$ plus a penalty term for the variations of the control signal during the same interval.

This definition of the cost function leads to the following expression:

$$\begin{aligned} J(k_0) \quad = \quad & \sum_{k=k_0}^{k_0+N_\mathrm{p}-1} \left[\sum_{j\in\mathscr{I}_\mathrm{s}} c_j(k) l_j n_j + \alpha \sum_{m\in\mathscr{I}_\mathrm{o}} w_m(k)\right. \\ & \left. + \alpha_{\mathrm{ramp}}\left(r(k) - r(k-1)\right)^2\right]\Delta T \end{aligned} \tag{7}$$

with $\mathscr{I}_\mathrm{s}$ the set of the indices of the motorway sections and $\mathscr{I}_\mathrm{o}$ the set of indices of the on-ramps. The parameter α is a weighing factor that allows to put more or less emphasis on the occurrence of queues at the on-ramps. The parameter α_{ramp} determines the relative importance of the penalty term for variations of the control signal. By increasing α_{ramp} we obtain a smoother control signal.

IV. Identification

Equations (1), (2), (3), (4), (5) and (6) completely define the traffic flow model. In the study case of the E17 motorway in Belgium we need to determine the parameters: v_f, c_cr, a_m, τ, ν and κ for every one of the 29 motorway sections. For every on-ramp, we also need to determine its capacity Q_m and the maximal possible density $c_{\max,j}$ in the section fed by the on-ramp. For the studied E17 motorway stretch this adds up to a total of 186 parameters that need to be fitted.

The available data to fit the model to is available through cameras that are installed every 500 m along the E17 motorway stretch under study. The camera images are processed into measurements of the flow and the average speed on a minutely basis. The flow and the average speed in a section vary in time as can be seen in Figure 3 where the measurements over a day are presented.

The parameter estimation problem can be formulated as a nonlinear least squares problem where the set of model parameters β that minimises the following cost criterion:

$$I(\beta) = \sum_{k=0}^{K}\left[\sum_{j\in\mathscr{I}_\mathrm{s}}\left(q_j(k) - \hat{q}_j(k)\right)^2 + \gamma\sum_{j\in\mathscr{I}_\mathrm{s}}\left(v_j(k) - \hat{v}_j(k)\right)^2\right] \tag{8}$$

is sought [11].

In the cost function for identification (8), the simulated flow $q_j(k)$ and the average speed $v_j(k)$ in every section are compared to the measured values $\hat{q}_j(k)$ and $\hat{v}_j(k)$ respectively. The squared error signals are summed over the $K+1$ samples present in the identification data set.

One approach to the identification of a model to use in the model predictive control framework could be to collect

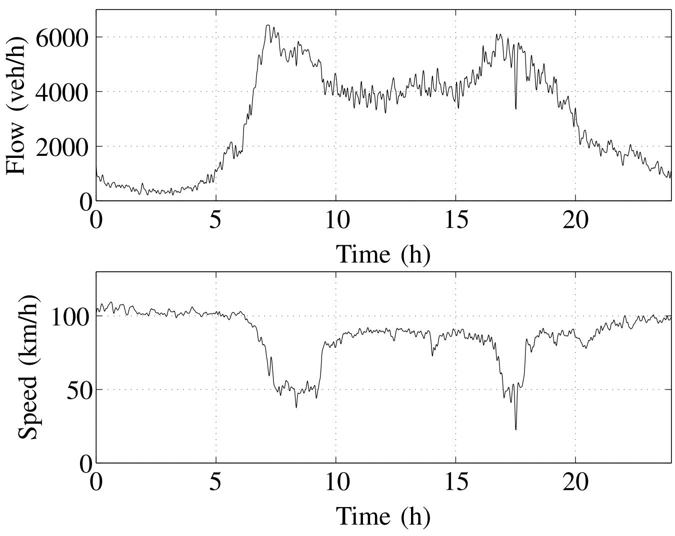

Fig. 3. Evolution of the traffic flow and the average speed (average over a minute) as measured by a camera detector on the E17 Ghent–Antwerp.

sufficient data and identify the model using this extensive dataset. The obtained model can then be used to optimise the ramp metering signals as discussed in Section III. Since not all parameters come to expression in every traffic operation mode, it is important to include traffic measurements from the different traffic operation modes in the identification dataset.

Even if we include data from the different traffic operation modes in our identification dataset, the resulting model is still not capable of mimicking the changing traffic behaviour due to external, non-modelled influences and disturbances such as the weather, an obstruction on the motorway, ... Therefore, we suggest an adaptive approach where the traffic model is re-fitted to the traffic measurement data on a regular basis. This way, changes in the model parameters due to the non-modelled influences are incorporated in the model that is used to determine the control signals. In the remainder of this section, we discuss some issues that need to be addressed when repeatedly re-fitting the nonlinear traffic flow model for use in the model predictive control framework.

As mentioned before, we need to estimate 186 parameters in the example of the E17 Ghent–Antwerp. This is a computationally very demanding task for an on-line traffic control system. This computational challenge is alleviated since we do not need to re-fit the model at the same pace as the control signals are calculated since we assume that the dynamic behaviour of the variations of the model parameters (e.g. weather changes) is slower than the dynamic behaviour of the traffic flows. By consequence, we can reduce the computational complexity by re-fitting the traffic flow model every 30 to 60 minutes. Since we can always use the old parameters as the starting point for the optimisation process when re-fitting the model, we can expect fast convergence towards the optimal parameters and thus a moderate computational complexity.

In literature [12], the sensitivity of the model quality on the different parameters was already investigated. The most important parameters of the traffic flow model were found to be v_{f} and c_{cr}. By only updating the parameters with the highest sensitivity, we can reduce the computational complexity substantially. E.g. for the presented motorway E17 Ghent–Antwerp, this approach reduces the number of parameters to be estimated from 186 to 70. In [12], it was claimed that the difference in sensitivity is so high that parameter values carefully chosen from literature can suffice for the least sensitive parameters.

V. Simulation results

In this section we use computer simulation to illustrate the importance of a good fit of the model parameters.

We implemented the traffic flow model discussed in Section III-A for the motorway E17 Ghent–Antwerp and we assumed that there is a ramp metering setup present at the fourth on-ramp. We also assume that we know the parameter set β_{ref} which, combined with the traffic flow model, perfectly describes the real-life behaviour of the motorway. The traffic flow model with parameter set β_{ref} will be used to assess the performance of the developed controllers.

Two model predictive control based ramp metering controllers are developed and tested on the traffic flow model with parameter set β_{ref}. The first controller uses the traffic flow model with parameter set β_{ref} in order to make predictions of the traffic situation over the prediction horizon N_{p}. The second controller uses a perturbed parameter set β_{pert} to make predictions over the prediction horizon. We chose the free flow speeds and the critical densities in parameter set β_{pert} about ten percent larger than those in the reference set β_{ref}. This results in a model that overestimates the capacity of the motorway. The prediction horizon N_{p} is seven minutes and the control horizon N_{c} is five minutes for both controllers. The control signal is only allowed to change every minute and the parameters α and α_{ramp} in (7) are chosen to be 1 and 10 respectively.

The simulation experiment covers four hours and the traffic demands on the mainline and at the fourth on-ramp are presented in Figure 4. The mainline traffic demand is considered to be constant and equal to 5500 vehicles an hour, while the capacity of the motorway is 6000 vehicles an hour. At the fourth on-ramp we assume a demand peak with a maximal demand of 750 vehicles an hour.

Since the total traffic demand on the mainline during the peak period is larger than the capacity of the motorway, congestion or a queue (or both) will occur. In Figure 5 we see the evolution of the density in the section that is fed by the fourth on-ramp and the queue length at the fourth on-ramp for both controllers. The solid line in the upper plot represents the evolution of the density realised by the first controller (β_{ref}). The density increases gradually but once the density becomes too high, the metering rate drops to a lower value as presented in Figure 6. The controller starts metering the on-ramp, resulting in the build-up of a queue. If we look at the performance of the second controller (β_{pert}, dashed line), we see in Figure 5 that no queue is formed at the fourth on-ramp. However, the rush hour traffic density in the section fed by the fourth on-ramp is higher for the second controller, resulting in a lower traffic flow. This is due to the fact that the model used

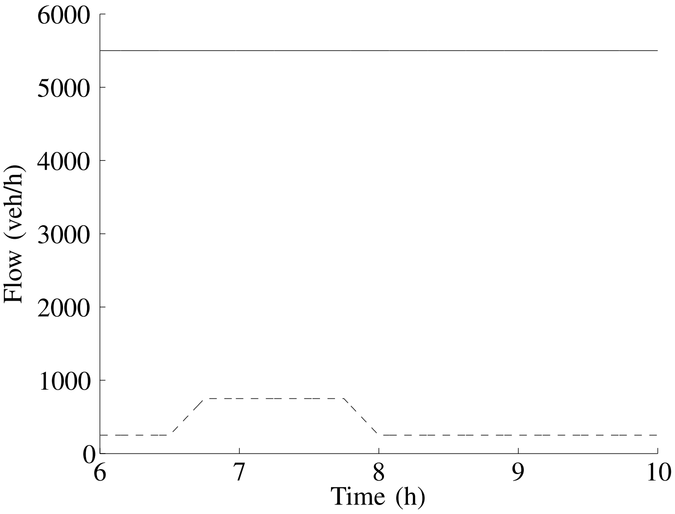

Fig. 4. Plot of the traffic demand on the motorway (solid line) and the traffic demand on the fourth on-ramp (dashed line) for the motorway of Figure 2.

by the second controller overestimates the motorway capacity what results in a metering rate that is too high for the capacity of the motorway (Figure 6).

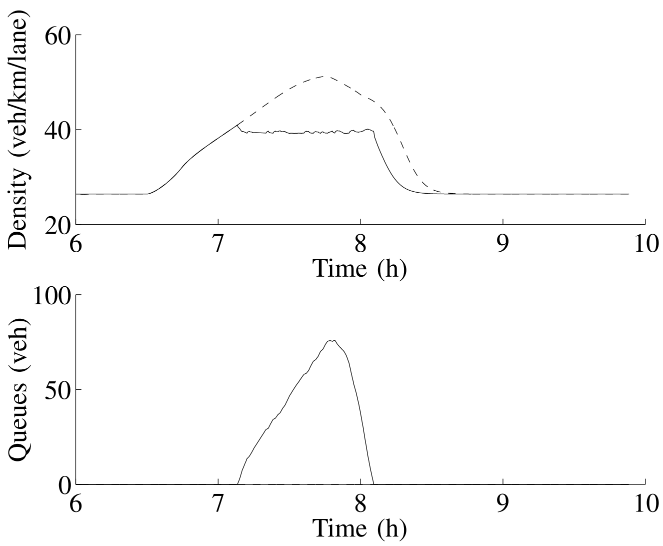

Fig. 5. Plot of the evolution of the density in the section fed by the fourth on-ramp and the queue at the first on-ramp for two model predictive control based ramp metering controllers. The solid line represents the controller using the correct model parameters (β_{ref}). The dashed line represents the controller using the perturbed parameter set (β_{pert}).

Given the traffic demands, we can calculate the total time spent associated with both controllers over the simulation interval using (7) with $\alpha_{\text{ramp}} = 0$. The first controller ($J_{\beta_{\text{ref}}} = 4102$) outperforms the second ($J_{\beta_{\text{pert}}} = 4140$). We can see from Figure 5 and Figure 6 that the mode of traffic operation realised by both controllers is totally different. The first controller keeps traffic on the mainline flowing smoothly at the cost of a queue at the fourth on-ramp while the second controller does not cause a queue at the fourth on-ramp but at the cost of a higher traffic density and delays on the mainline.

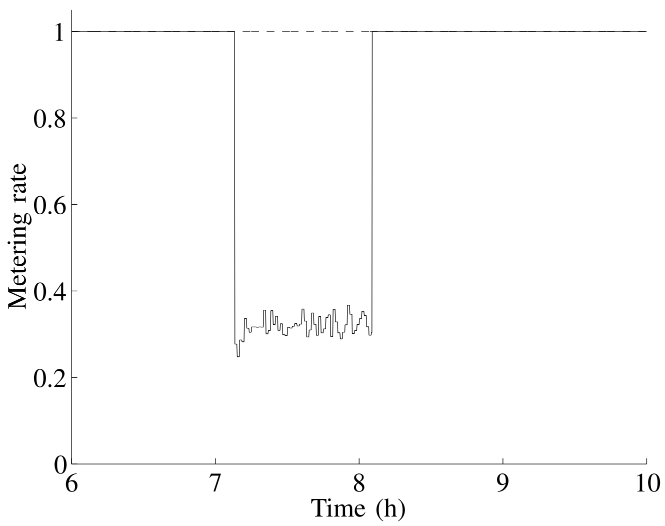

Fig. 6. Plot of the evolution of the metering rate as presented by both controllers. The solid line represents the controller using the correct model parameters (β_{ref}). The dashed line represents the controller using the perturbed parameter set (β_{pert}). Note that the controller with the perturbed parameter set never becomes active.

VI. Conclusions

We have presented ramp metering in a receding horizon framework. As an example, we studied the real-life motorway E17 Ghent-Antwerp in Belgium. We discussed a traffic flow model that needs to be fitted to measurement data for use in a model predictive control based ramp metering setup. Since unpredictable external factors influence the traffic situation on the motorway, we argued that a regular re-fitting of the model parameters improves the quality of the controller. This was illustrated with some simulation results.

Acknowledgements

Our research is supported by grants from several funding agencies and sources: Research Council KUL: Concerted Research Action GOA-Mefisto 666 (Mathematical Engineering), IDO (IOTA Oncology, Genetic networks), several PhD/postdoc & fellow grants; Flemish Government: Fund for Scientific Research Flanders (several PhD/postdoc grants, projects G.0256.97 (subspace), G.0115.01 (bio-i and microarrays), G.0240.99 (multilinear algebra), G.0197.02 (power islands), G.0407.02 (support vector machines), research communities ICCoS, ANMMM), AWI (Bil. Int. Collaboration Hungary/ Poland), IWT (Soft4s (softsensors), STWW-Genprom (gene promotor prediction), GBOU-McKnow (Knowledge management algorithms), Eureka-Impact (MPC-control), Eureka-FLiTE (flutter modeling), several PhD grants); Belgian Federal Government: OSTC (IUAP IV-02 (1996-2001) and IUAP V-10-29 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling), Program Sustainable Development PODO-II (CP-TR-18: Sustainibility effects of Traffic Management Systems); Direct contract research: Verhaert, Electrabel, Elia, Data4s, IPCOS.

References

[1] B. Hellinga and M. Van Aerde, "Examining the potential of using ramp metering as a component of an ATMS", *Transportation Research Record*, vol. 1494, pp. 75–83, 1995.

[2] A. D. May, *Traffic flow fundamentals*, Prentice Hall, Englewood Cliffs, NJ,, 1990.

[3] H. Zhang, G. Ritchie, and W. Recker, "Some general results on the optimal ramp control problem", *Transportation Research C*, vol. 4, no. 2, pp. 51–69, 1996.

[4] Cambridge Systematics, Inc., Oakland, California, *Twin Cities Ramp Meter Evaluation – Final Report*, Feb. 2001, Prepared for the Minnesota Department of Transportation.

[5] J.M. Maciejowski, *Predictive Control with Constraints*, Prentice Hall, Harlow, England, 2002.

[6] B. Kouvaritakis, M. Cannon, and J.A. Rossiter, “Non-linear model based predictive control”, *International Journal of Control*, vol. 72, no. 10, pp. 919–928, 1999.

[7] H. J. Payne, “Models of freeway traffic and control”, in *Mathematical Models of Public Systems*, G. A. Bekey, Ed., La Jolia, California, 1971, vol. 1 of *Simulation Council Proceedings Series*, pp. 51–61.

[8] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, “Modelling and real-time control of traffic flow on the southern part of boulevard périphérique in Paris: Part I: Modelling”, *Transportation Research A*, vol. 24A, no. 5, pp. 345–359, 1990.

[9] M. Papageorgiou, J. M. Blosseville, and H. Hadj-Salem, “Macroscopic modelling of traffic flow on the boulevard périphérique in Paris”, *Transportation Research B*, vol. 23B, no. 1, pp. 29–47, 1989.

[10] A. Kotsialos, M. Papageorgiou, and A. Messmer, “Optimal coordinated and integrated control motorway traffic control”, in *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, A. Ceder, Ed., 1999.

[11] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, New Jersey, 2nd edition, 1999.

[12] M. Cremer and M. Papageorgiou, “Parameter identification for a traffic flow model”, *Automatica*, vol. 17, no. 6, pp. 837–843, 1981.