

Traffic flow optimisation in presence of vehicle automation and communication systems - Part I: A first-order multi-lane model for motorway traffic

Claudio Roncoli, Markos Papageorgiou, Ioannis Papamichail

*Dynamic Systems and Simulation Laboratory,
Technical University of Crete, Chania 73100 Greece
(e-mail: <croncoli, markos, ipapa>@dssl.tuc.gr).*

Abstract

Proposed or emerging vehicle automation and communication systems (VACS) may contribute to the mitigation of motorway traffic congestion on the basis of appropriate traffic control strategies. In this context, this paper presents a novel first-order multi-lane macroscopic traffic flow model for motorways which is mainly intended for use within a related optimal control problem formulation. The model's starting point is close to the well-known CTM (cell-transmission model), which is modified and extended to consider additional aspects of the traffic dynamics, such as lane changing and the capacity drop, via appropriate procedures for computing lateral and longitudinal flows. The model has been derived with a view to combine realistic traffic flow description with a simple (linear or piecewise linear) mathematical form, which can be exploited for efficient optimal control problem formulations, as described in a companion (Part II) paper. Although the model has been primarily derived for use in future traffic conditions including VACS, it may also be used for conventional traffic flow representation. In fact, the accuracy of the proposed modelling approach is demonstrated through calibration and validation procedures using real data from an urban motorway located in Melbourne, Australia.

1 Introduction

Traffic congestion is a major problem of modern motorway systems, causing serious infrastructure degradation in and around metropolitan areas. The European Commission estimates that the yearly cost of road traffic congestion in Europe exceeds 120 billion €, and similar figures apply to USA as well. Despite a multitude of practised traffic control measures, improvements in combating traffic congestion and its detrimental consequences have been relatively moderate. On the other hand, in the last two decades there has been a significant and increasing interdisciplinary effort by the automotive industry as well as by numerous research institutions around the world to plan, develop, test and start deploying a variety of Vehicle Automation and Communication Systems (VACS) that are expected to revolutionise the features and capabilities of individual vehicles within the next decades. A wide description of VACS may be found in Bishop (2005).

Although most VACS will have no direct impact on traffic flow, as they are aiming at improving safety or driver convenience, quite a few of them will change the vehicle behaviour in the longitudinal or lateral directions and will offer new opportunities for innovative traffic control actions, such as individual vehicle speed and lane-change advices, among others. An arising difficulty in exploiting these new opportunities is due to the gradual and uncertain future evolution path of VACS deployment, corresponding deployment scenarios ranging from already available ACC (Adaptive Cruise Control) systems to fully automated and connected vehicles and AHS (Automated Highway Systems) (Rao and Varaiya, 1994). This uncertainty calls for the development of modelling and control approaches that are robust, to the extent possible, to evolving VACS and corresponding penetration rates. Among the possible novel actuators, we have considered that, in case some vehicles are enabled to communicate with the infrastructure (V2I), it is possible to recommend to them variable speed limits, or even dictate them their maximum driving speed (in case they are also equipped with an ACC system), according to the real-time decisions of an external Decision Maker (DM). Of course, the optimal control decisions must be computed for the aggregate traffic flow in specific space-time windows, not for some individual vehicles only. Similarly, V2I-equipped vehicles may

receive from the DM lane-changing advices, so as to implement corresponding lateral flow decisions. To enable a rational design for these actions, an appropriate traffic flow model is required.

To address this need, it is the main purpose of this paper to develop a macroscopic traffic flow model to be used within a model-based optimal control strategy. In order to permit a flexible use of variable speed limits and to implement a lane-changing control strategy, the proposed model is defined considering multiple lanes. Moreover, due to this intended utilisation, the simplicity of the proposed model is a primary requirement, so as to enable efficient optimal control calculations even for large-scale traffic networks. Nevertheless, the model accuracy is demonstrated in this paper by use of real data to be sufficient for other, more conventional, model uses as well. This paper is followed by a second part (Roncoli et al., 2015b), where the optimal control formulation and solution approaches, that take advantage of the simple (specifically: piecewise linear) form of the proposed model.

The paper is structured as follows: in Section 2 some available models are reviewed and the proposed multi-lane macroscopic traffic flow model for motorways is described, highlighting its novel aspects. In Section 3, the proposed model is calibrated and validated using real data from a motorway network; whereas Section 4 concludes the paper.

2 A multiple-lane traffic flow model for motorways

2.1 Literature review of multiple-lane traffic flow models

The vast majority of the proposed macroscopic motorway traffic flow models consider all the variables aggregated across all the lanes (for a wide description and classification of existing models, see Treiber and Kesting, 2013); only a limited number of studies consider the different lanes in a motorway as independent entities. One of the first important works that addressed this topic was by Gazis et al. (1962), where lane densities on a multi-lane highway were assumed to oscillate around an equilibrium density; the authors developed a methodology to attenuate the disturbances and to increase the stability of the system. This work inspired

Michalopoulos et al. (1984), who proposed three models for capturing the lane-changing behaviour. The first model is a continuum model, based on the assumption that vehicles change lanes according to the difference of the deviations of their densities from equilibrium values. The second model extends the first one, taking into account also acceleration and inertia effects, incorporating them in a second-order model. A third extension was also proposed, considering, in addition, the street width. However, these models were formulated in a continuous space-time domain, without applying any discretisation scheme. In a more recent work, Laval and Daganzo (2006) exploited the kinematic wave (KW) theory, proposing a multi-lane KW-based model as a first module of a more complex model that considers also moving blockages, treated as particles and characterised by bounded acceleration rates; lane changes are assigned according to the difference of mean speed between two neighbouring lanes. Yet another approach is presented by Jin (2013), in which the author developed a multi-commodity model based on the LWR theory (Lighthill and Whitham, 1955); in this case the lane-changing flow is defined based on the concept of lane-changing intensity (that is introduced as a new variable which affects the speed-density relationship), defining a new fundamental diagram (FD) and introducing a so-called entropy condition. Other approaches are based on the gas-kinetic traffic flow model, such as Helbing (1997), where the author derived macroscopic traffic flow equations for interacting lanes taking explicitly into account queuing effects; and Hoogendoorn and Bovy (1999), where traffic is described as a collection of vehicle platoons governed by continuum (smooth changes in the traffic flow variables) and non-continuum processes (deceleration of vehicles and lane changing). Some recent works are treating the problem of modelling the lane flow distribution at specific locations of the network; Knoop et al. (2010) studied a relation between the total density and the lane densities during free-flow and congested regime; Knoop et al. (2012) examined the number of lane changes as function of several incentives, finding that the most critical are the densities in the origin and in the target lanes; whereas Duret et al. (2012) analysed real data (in free-flow conditions) for a three-lanes motorway, deriving a simple linear model considering the lane distribution of traffic flow.

2.2 The generic modelling framework

A multiple-lane motorway is considered, which is subdivided into segments, while each segment comprises a number of lanes. The index $i = 1, \dots, I$ is introduced for segments and the index $j = 1, \dots, J$ for lanes. The model is formulated in discrete time, considering the discrete time step T for a simulation horizon K indexed by $k = 1, \dots, K$, where the simulation time is $t = kT$. The motorway is discretised in space by defining the segment-lane entities, which are characterised by the following variables (a graphical representation of the segment-lane variables is presented in Figure 1):

- the density $\rho_{i,j}(k)$ [veh/km], i.e. the number of vehicles in the segment i , lane j , at time step k , divided by the segment length L_i ;
- the longitudinal flow $q_{i,j}(k)$ [veh/h], i.e. the traffic volume leaving segment i and entering segment $i + 1$ during time interval $(k, k + 1]$, thus remaining in lane j ;
- the lateral flow $f_{i,j,\bar{j}}(k)$ [veh/h] ($\bar{j} = j \pm 1$), i.e. the traffic volume moving from lane j to lane \bar{j} (vehicles changing lane remain in the same segment during the current time interval); and
- the on-ramp flow $r_{i,j}(k)$ [veh/h], i.e. the traffic volume entering from the on-ramp located at segment i , lane j , during the time interval $(k, k + 1]$.

The off-ramp flow is determined according to given turning rates $\gamma_{i,j}(k)$, as a percentage of the total flow passing through all the lanes of the segment:

$$q_{i,j}^{off}(k) = \gamma_{i,j}(k) \sum_{j=1}^J q_{i,j}(k). \quad (1)$$

According to the aforementioned notation, the variable $\rho_{i,j}(k)$ is updated based on the following conservation equation:

$$\begin{aligned} \rho_{i,j}(k+1) = \rho_{i,j}(k) + \frac{T}{L_i} & \left[q_{i-1,j}(k) + r_{i,j}(k) - q_{i,j}(k) - q_{i,j}^{off}(k) \right. \\ & \left. + f_{i,j+1,j}(k) + f_{i,j-1,j}(k) - f_{i,j,j-1}(k) - f_{i,j,j+1}(k) \right]. \end{aligned} \quad (2)$$

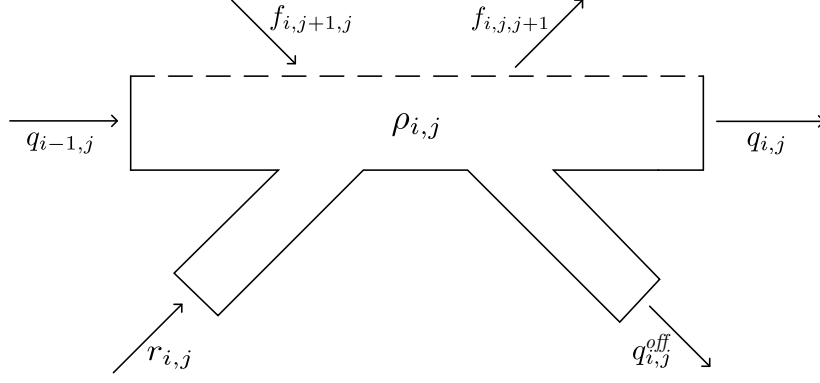


Figure 1: The segment-lane variables used in the model formulation.

It is important to highlight that, in order to ensure numerical stability, the time step T must respect the Courant-Friedrichs-Lewy (CFL) condition (Courant et al., 1928), precisely:

$$T \leq \min_{i,j} \frac{L_i}{v_{i,j}^{free}} \quad (3)$$

where $v_{i,j}^{free}$ is the maximum speed allowed in the corresponding segment-lane. Equation (2) is the only dynamic equation considered in this first-order traffic flow model.

For the sake of simplicity, the flow entering from an on-ramp has priority over lateral and longitudinal flow. In other words, it is not considered that mainstream congestion can spill back into merging on-ramps. As an extension, this phenomenon can be accounted for the definition of appropriate merge ratios at on-ramps; because of the constraints for the subsequent optimisation problem, only linear approaches can be considered, e.g. Papageorgiou et al. (1990) or Bar-Gera and Ahn (2010). The methodology to compute the lateral and longitudinal flows is described in the next sections.

2.3 Computation of lateral flows

When appropriate communications between equipped vehicles and the control strategy are available, efficient lane-changing advice or control may be effectuated for selected vehicles. However, if the corresponding VACS penetration rates are small or if a no-control base-case scenario is required for comparative evaluation, the model should be enabled to also reproduce the “natural” aggregate lane-changing behaviour of traffic. Lateral flows due

to “natural” lane-changing are considered among adjacent lanes of the same segment, and corresponding rules must be defined in order to properly assign and bound the lateral flows. To start with, the maximum available flow for lateral movements is calculated based on the current amount of vehicles in the segment-lane:

$$F_{i,j}(k) = \frac{L_i}{T} \rho_{i,j}(k). \quad (4)$$

The pursued approach is to compute the lateral flow from segment-lane (i, j) to (i, \bar{j}) as a rate $A_{i,j,\bar{j}} \in [0, 1]$ of the value $F_{i,j}(k)$. We call this rate the attractiveness rate.

As a matter of fact, lane-changing behaviour and flows are extremely hard to model accurately (both macroscopically and microscopically) because they depend on a very high number of partly interdependent factors. To start with, different vehicle types (e.g. cars and trucks) and related restrictions (e.g. regarding lane usage) may give rise to a variety of different lane-changing characteristics. Human driver behaviour is another source of uncertainty and variation. Furthermore, lane-changing behaviour is different in case of road curvature or grade, lane drop, as well as in tunnels, on bridges, etc.; it is also dependent on the number of lanes, environmental conditions (weather, lighting), traffic conditions (free-flow, dense, congested) and traffic signs. Last not least, lane-changing activity is quite particular in the vicinity of on-ramps and off-ramps, or at weaving sections. Given this diversity and complexity, it appears appropriate, in the present context, to provide for a simple basic lane-changing flow model, which would capture with some accuracy many “ordinary” situations; accompanied by a space-time dependent parameter that could be used to influence the model calculations appropriately whenever needed (e.g. near on- and off-ramps).

For the basic lane-changing flow model, it is assumed that drivers may consider a lane change when one of the adjacent lanes offers a higher speed or a lower traffic density. The latter option is preferred here because traffic densities are state variables of the proposed model, whereas the speed values can be computed only a posteriori. Thus, under “ordinary” conditions, the current attractiveness rate $A_{i,j,\bar{j}}$ may be deemed to increase proportionally to the current density difference $\rho_{i,j}(k) - \rho_{i,\bar{j}}(k)$ of adjacent lanes j and \bar{j} . This basic assumption, however, may be subject to variations due to various local effects, as mentioned earlier. For

example, vehicles driving on the slow lane may consider a lane change upstream of on- or off-ramps, to avoid interference with entering or exiting vehicles, respectively; similarly, vehicle lane assignment may deviate from the basic rule upstream of lane-drop locations; also, exiting vehicles have to move towards the exit lane(s) irrespective of the prevailing traffic density there. Some of these or other variations from the basic rule may additionally depend on time. To capture this variety of potential situations, the attractiveness rate is modelled to depend on the weighted density difference $P_{i,j,\bar{j}}(k)\rho_{i,j}(k) - \rho_{i,\bar{j}}(k)$, where the introduced factor $P_{i,j,\bar{j}}(k)$ is mostly equal to 1, but should be tuned to reflect particular location- or time-dependent effects where needed. This factor takes into account the same difference in density between the adjacent lanes, irrespective of the considered lane-changing direction, thus it respects the relation $P_{i,j,\bar{j}}(k) = \frac{1}{P_{i,\bar{j},j}(k)}$ for all the adjacent segment-lane couples (i, j) , (i, \bar{j}) . A similar concept, although with a different implementation based on the notion of “density equilibrium”, was proposed by Gazis et al. (1962) and utilised by Munjal and Pipes (1971) and Michalopoulos et al. (1984).

Finally, the attractiveness rate $A_{i,j,\bar{j}}$ is computed as:

$$A_{i,j,\bar{j}}(k) = \mu \max \left[0, \frac{P_{i,j,\bar{j}}(k)\rho_{i,j}(k) - \rho_{i,\bar{j}}(k)}{P_{i,j,\bar{j}}(k)\rho_{i,j}(k) + \rho_{i,\bar{j}}(k)} \right] \quad (5)$$

where the coefficient μ is a unique parameter in the range $[0, 1]$, reflecting a sort of “aggressiveness” in lane-changing. As mentioned earlier, we have $P_{i,j,\bar{j}}(k) = 1$ for the vast majority of locations; hence, the calibration of $P_{i,j,\bar{j}}(k)$ is only necessary, for good modelling accuracy, at few specific motorway locations (e.g. upstream of on- and off-ramps, lane drop locations, etc.), where lateral flow may occur in the direction from a lower density to a higher one. In general, this calibration can be performed according to empirical observation of the relative densities between adjacent segments. Analysing some available data, it was indeed confirmed that the specific locations of a motorway, where $P_{i,j,\bar{j}}(k) \neq 1$, are mainly in proximity of on- and off-ramps; some additional details are given in the case study of Section 3. The fraction within the brackets in (5) is introduced in order to obtain a value in the range $[-1, 1]$, where a value equal to 0 (i.e., no lane-changes) is obtained when the density difference is at the predefined equilibrium. Note that, due to the max-operator, only the positive values are

considered, which means that for each couple of adjacent segments-lanes, only one lateral flow has a positive value.

The modelling function in (5) is non-linear; nevertheless this is no impediment for the subsequent optimal control problem formulation because the lateral flows will be considered controllable in the control problem. Thus, (5) is intended to reflect the “natural” lane-changing flow behaviour. Since the proposed methodology is not strongly dependent on the function used here for computing $A_{i,j,\bar{j}}(k)$, other functions could be used for this computation if deemed more appropriate.

As mentioned earlier, vehicles bound for specific off-ramps should change lanes in a due manner. According to (1), the off-ramp flow is computed as a percentage of the total flow passing through all lanes of the segment and not only of the shoulder lane. If the values $P_{i,j,\bar{j}}(k)$ are properly tuned and the turning rates $\gamma_{i,j}(k)$ do not feature strong oscillations, this should not create any noticeable problem. In the general case, however, it could happen that the number of vehicles in the shoulder lane is insufficient to guarantee the necessary exiting flow in accordance with the corresponding pre-fixed exiting rate. In order to cope with this possibility, an approach is introduced, which incorporates a forecast of the entering and leaving flows in the off-ramp location. Assuming that the segment containing the off-ramp is denoted by indices \hat{i} and \hat{j} , the target of having enough vehicles in the segment to feed the exit implies that $\rho_{\hat{i},\hat{j}}(k+1) \geq 0$; since we have only one entering lateral flow (from lane \bar{j}), this condition yields, using (2):

$$\rho_{\hat{i},\hat{j}}(k) + \frac{T}{L_{\hat{i}}} \left[q_{\hat{i}-1,\hat{j}}(k) - q_{\hat{i},\hat{j}}(k) - \gamma_{\hat{i},\hat{j}}(k) \sum_{j=1}^J q_{i,j}(k) + f_{\hat{i},\bar{j},\hat{j}}(k) \right] \geq 0. \quad (6)$$

Thus, the minimum lateral flow that guarantees the condition is:

$$f_{\hat{i},\bar{j},\hat{j}}^{off}(k) = q_{\hat{i},\hat{j}}(k) + \gamma_{\hat{i},\hat{j}}(k) \sum_{j=1}^J q_{i,j}(k) - q_{\hat{i}-1,\hat{j}}(k) - \frac{L_{\hat{i}}}{T} \rho_{\hat{i},\hat{j}}(k). \quad (7)$$

Considering the assumptions of a time step $T \approx \frac{L_i}{v_{i,j}^{free}}$ and free flow conditions ($q_{\hat{i},\hat{j}} = \rho_{\hat{i},\hat{j}}(k) v_{\hat{i},\hat{i}}^{free}$, in case a triangular FD is used), it can be derived that virtually all the vehicles present in segment i at time k will be transferred into segment $i+1$ during time intervals

$(k, k + 1]$; thus it is a reasonable approximation to state that:

$$q_{i,j}(k) \approx q_{i+1,j}(k + 1). \quad (8)$$

According to (7) and (8), the following formula provides an approximation of the required amount of lateral flow needed to satisfy the off-ramp flow $q_{i,j}^{off}(k)$ in (1):

$$f_{i,\bar{j},\hat{j}}^{off}(k) = \gamma_{i,\hat{j}}(k) \sum_{j=1}^J q_{i,j}(k) - q_{i-1,\hat{j}}(k) \approx \gamma_{i,\hat{j}}(k) \sum_{j=1}^J q_{i-1,j}(k-1) - q_{i-2,\hat{j}}(k-1). \quad (9)$$

This procedure is applied only in the direction towards the lane that includes the ramp and it could be iteratively applied to further upstream segments in case the flow is not sufficient to match the exit rate. Consider also that a negative value $f_{i,\bar{j},\hat{j}}^{off}(k)$ will be rejected by a successive equation. This approach is capable of generating lateral flows sufficient to ensure an appropriate off-ramp flow for the vast majority of cases, except for some extreme scenarios where infeasible flows may still be generated.

In conclusion, the lateral demand flow, i.e. the flow that will actually materialise if there is enough space in the target segment-lane, is assigned according to the following formula:

$$D_{i,j,\bar{j}}(k) = \max \left[A_{i,j,\bar{j}}(k) F_{i,j}(k), f_{i,j,\bar{j}}^{off}(k) \right] \quad (10)$$

where $\bar{j} = j \pm 1$.

To complete the lateral flow modelling development, we need to account for the space available in the segment-lane that is receiving the vehicles moving laterally. To this end, the following function describing the available space in terms of flow acceptance is considered:

$$S_{i,\bar{j}}(k) = \left[\rho_{i,\bar{j}}^{jam} - \rho_{i,\bar{j}}(k) \right] \frac{L_i}{T}. \quad (11)$$

Since the available space may not be sufficient for accepting the lateral flow entering from both sides, the assigned quantity is proportionally distributed according to the following relation:

$$\begin{aligned} f_{i,j-1,j}(k) &= \min \left[1, \frac{S_{i,j}(k)}{D_{i,j-1,j}(k) + D_{i,j+1,j}(k)} \right] D_{i,j-1,j}(k) \\ f_{i,j+1,j}(k) &= \min \left[1, \frac{S_{i,j}(k)}{D_{i,j-1,j}(k) + D_{i,j+1,j}(k)} \right] D_{i,j+1,j}(k). \end{aligned} \quad (12)$$

These values may be directly used for updating the densities according to (2). It is interesting to note that the lateral flows act as “source/sink” terms for the implemented numerical scheme of the LWR model, therefore they have no influence on the CFL condition, implying that the current condition (3) is sufficient. In some exceptional cases, to avoid excessive lateral flows, an option could be to limit the lateral flow by a fixed upper bound, or by the capacity of the leaving or entering segment/lane. Note that, in case of limited available space in the target segment-lane, the described approach assigns higher priority to the lateral flows, with respect to the longitudinal flows, via (12). Naturally, these flows will be considered in the computation of the longitudinal flows in the next subsection.

2.4 Computation of longitudinal flows

As previously mentioned, the longitudinal flows are defined as the flows going from a segment to the next downstream one, while remaining in the same lane. These flows may be considered controllable via appropriate speed control of equipped vehicles, however they need to be appropriately modelled for the no-control case. The main logic for their computation is based on the conventional cell transmission model (CTM), where the longitudinal flow is computed as the minimum between an upstream demand flow and a downstream supply flow. However, one important phenomenon that regularly appears in real traffic, but is not reproduced by the classical CTM, is the capacity drop phenomenon, i.e. the reduction of discharge flow once queues start forming at a bottleneck location. The reasons for this phenomenon are not exactly clear; some research works attribute the reason of the capacity drop to the limited and varying acceleration of vehicles leaving a congested area (see, e.g. Papageorgiou et al., 2008) while trying to reach the desired speed; whereas some other works attribute the capacity drop to the voids generated due to the acceleration of merging vehicles at on-ramps (Laval and Daganzo, 2006, Cassidy and Ahn, 2005); in Treiber et al. (2006) potential explanations are also presented. In second-order models, such as METANET (Papageorgiou and Messmer, 1990), the capacity drop appears as a consequence of the spatiotemporal evolution of the speed dynamics. Since the speed is not dynamically calculated, this option is not available

for first-order LWR models; hence, several attempts have been made in order to introduce it in some other ways. Hall and Hall (1990) suggested the introduction of an inverse lambda fundamental diagram (FD), whereas Leclercq et al. (2011) proposed an analytical model that incorporates endogenously the capacity drop. Lebacque (2003) addressed the problem by imposing an upper bound to the acceleration depending on the traffic phase, distinguishing between traffic equilibrium and maximum acceleration conditions; this work was exploited, proposing a so-called “node model” by Monamy et al. (2012). Laval and Daganzo (2006) proposed the introduction of discrete particles in the traffic flow, treating them as moving temporary blockages. In a more recent and interesting approach, Srivastava and Geroliminis (2013) extended the LWR model by defining a FD with two values of capacity and providing a memory-based methodology to choose the appropriate value in the numerical solution of the problem, thus distinguishing between congested and uncongested states.

A common feature of most of the aforementioned approaches is the non-linear formulation that makes them unattractive for the purpose of this paper. For this reason, in this work, a simple model for including the capacity drop phenomenon in the model calculation is employed. The origin of the proposed approach may be found in the work by Lebacque (2003), that contains also a wider theoretical description of this model. The basic idea is to mimic the limited acceleration that usually appears at the head of a congestion, by adding a term that bounds the outflow of the segment. This way, the so-called “recovery state” replaces the usual free-flow state at the head of a congestion, causing a lower outflow, that yields also a reduced speed compared to the usual CTM.

In the remaining part of this section, we are considering the LWR model discretised according to the well-known Godunov-scheme (Godunov, 1959). Following other works (e.g. Lebacque, 2003), we are naming “demand part” the function representing the greatest possible flow that can be sent by the upstream segment; and “supply part” the function representing the greatest possible flow that can be received by the downstream segment. In addition, for the sake of simplicity, a triangular-shaped FD is used, as in the original formulation of the CTM (Daganzo, 1994). However, for the subsequent formulation of the

optimisation problem, any concave piecewise-linear function, which will be deemed to be more accurate in capturing speed at lower densities, can be considered.

In our model, since we are interested in a linear formulation, the aforementioned capacity-drop approach is achieved by redefining the demand part of the FD in the following way: in case of congestion (i.e. $\rho_{i,j}(k) > \rho_{i,j}^{cr}$, where $\rho_{i,j}^{cr}$ is the critical density for the segment-lane (i, j)), the flow is linearly decreased according to a fixed slope $-w^D$, instead of being equal to capacity flow as in the conventional CTM; this can be seen in the sketched FD of Figure 2. Note that, in order to avoid interference with possible shockwaves caused by a back-spilling congestion, the relation $w^D < w^S$ must hold. This extension calls for the definition of an additional point in the FD, $q_{i,j}^{jam}$, i.e. the flow that is allowed to leave a segment while it has entered a completely congested state (i.e. $\rho_{i,j}(k) = \rho_{i,j}^{jam}$). Note that this simple modification of the demand function is not sufficient to create a capacity drop at the head of a congestion under all circumstances; this is achieved here via the lateral and ramp flows, which act as sources for the LWR model (which are not considered explicitly in the demand and supply functions); this permits to obtain a density increase beyond the critical value in the segment-lane placed at the head of the congestion, which triggers the capacity drop. Other possibilities to account for capacity drop in a first-order model, maintaining linear constraints, are currently under investigation. Note that in the conventional CTM we have $q_{i,j}^{jam} = q_{i,j}^{max}$, i.e. no capacity drop at the head of congestion.

In conclusion, the demand part for the longitudinal flow calculation is computed as:

$$Q_{i,j}^D(k) = \min \left[v_{i,j}^{free} \rho_{i,j}(k), -w^D \rho_{i,j}(k) + g^D \right] \quad (13)$$

where:

$$\begin{aligned} w^D &= \frac{v_{i,j}^{free} \rho_{i,j}^{cr} - q_{i,j}^{jam}}{\rho_{i,j}^{jam} - \rho_{i,j}^{cr}} \\ g^D &= \frac{\rho_{i,j}^{cr} (v_{i,j}^{free} \rho_{i,j}^{jam} - q_{i,j}^{jam})}{\rho_{i,j}^{jam} - \rho_{i,j}^{cr}}. \end{aligned} \quad (14)$$

The supply function is computed based on the density of the downstream segment-lane as in the classic CTM:

$$Q_{i,j}^S(k) = \min \left[v_{i+1,j}^{free} \rho_{i+1,j}^{cr}, -w^S \rho_{i+1,j}(k) + g^S \right] \quad (15)$$

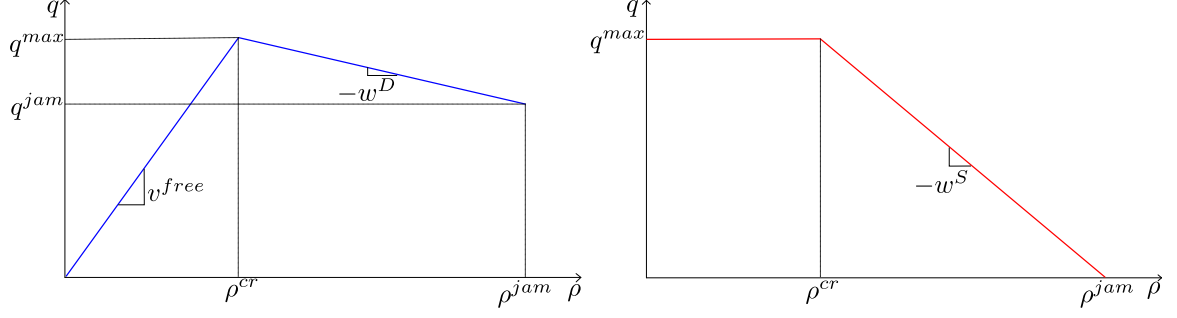


Figure 2: The proposed demand (left) and supply (right) parts of the Fundamental Diagram; the demand part includes the linear term for capacity drop.

where:

$$\begin{aligned} w^S &= \frac{v_{i+1,j}^{free} \rho_{i+1,j}^{cr}}{\rho_{i+1,j}^{jam} - \rho_{i+1,j}^{cr}} \\ g^S &= \frac{v_{i+1,j}^{free} \rho_{i+1,j}^{cr} \rho_{i+1,j}^{jam}}{\rho_{i+1,j}^{jam} - \rho_{i+1,j}^{cr}}. \end{aligned} \quad (16)$$

The calibration of the slope $-w^D$ is a non-trivial task because of the complexity in measuring the effects of a capacity drop. As mentioned earlier, the capacity drop phenomenon is an empirical fact, but the exact reasons for its occurrence have not been fully explored until now. Model calibration exercises, based on real traffic data (as in this paper), may lead to a factual quantification of its impact. However, as real traffic flow data in presence of VACS-equipped vehicles are not yet available, such calibration exercises in presence of VACS can only be produced based on corresponding simulated traffic data. Most probably, automated car-following (ACC or cooperative ACC) and lane-changing systems will alter the quantifiable impact of the capacity-drop phenomenon (see, e.g., Kesting et al., 2008) compared to the current situation; but a reliable quantification seems hardly possible at the present stage, because it will depend on the specific dynamics and parameter settings of the automated systems, as well as on the corresponding penetration rates. In view of this uncertainty, the presented modelling approach provides a flexibility in correctly reflecting the future gradual evolution of traffic flow dynamics via appropriate specification of the parameter w^D .

Since the proposed model considers a higher priority associated to lateral flows, their

values must be considered while computing the demand and the supply terms, according to:

$$\begin{aligned}\hat{Q}_{i,j}^D(k) &= \min \left\{ Q_{i,j}^D(k), \frac{L_i}{T} \rho_{i,j}(k) + [f_{i,j-1,j}(k) + f_{i,j+1,j}(k)] - [f_{i,j,j-1}(k) + f_{i,j,j+1}(k)] \right\} \\ \hat{Q}_{i,j}^S(k) &= Q_{i,j}^S(k) + [f_{i+1,j,j-1}(k) + f_{i+1,j,j+1}(k)] - [f_{i+1,j-1,j}(k) + f_{i+1,j+1,j}(k)]\end{aligned}\quad (17)$$

Thus, the final longitudinal flows $q_{i,j}(k)$ are calculated according to the following equation:

$$q_{i,j}(k) = \min \left[\hat{Q}_{i,j}^D(k), \hat{Q}_{i,j}^S(k) \right]. \quad (18)$$

The model admits the possibility of introducing upper bounds $q_{i,j}^{off,max}$ for exit flows at off-ramps, which leads to the modelling of possible congestion due to limited-capacity off-ramps flows. In this case, the upper-bounds $q_{i,j}^{off,max}$ will affect the outflow in all the mainstream lanes at the off-ramp location; to reproduce this phenomenon, the longitudinal flow must be updated as follows:

$$\begin{aligned}\delta_{i,j}^{off}(k) &= \min \left[1, \frac{q_{i,j}^{off,max}}{\gamma_{i,j}(k) \sum_{j=1}^J q_{i,j}(k)} \right] \\ \hat{q}_{i,j}(k) &= \delta_{i,j}^{off}(k) q_{i,j}(k).\end{aligned}\quad (19)$$

Lateral flow into a segment-lane may affect its capacity. To reflect this possibility within our approach, we may readily render the flow capacity of a segment-lane linearly dependent on the entering lateral flows. Again, a linear relation can be considered, with the purpose of decreasing the capacity proportionally to the entering or exiting flow (and similarly, the concept can be extended also for the flow entering from on-ramps); the following general formula can be applied, that updates the computation of the demand part for the longitudinal flow used in (13):

$$\hat{g}^D = g^D - \alpha_{ef} [f_{i,j-1,j}(k) + f_{i,j+1,j}(k)] - \alpha_{lf} [f_{i,j,j-1}(k) + f_{i,j,j+1}(k)] - \alpha_r r_{i,j}(k) \quad (20)$$

where α_{ef} , α_{lf} , and α_r are parameters to be opportunely tuned for entering lateral flow, exiting lateral flow, and on-ramp flow respectively.



Figure 3: The geographical map of the M1 stretch used in this section. The considered direction is from the green (bottom right) to the red (top left) marks.

3 Model calibration and validation

3.1 Network description

The motorway traffic flow model introduced in Section 2 is now applied to a particular motorway stretch in order to calibrate its parameters and validate the model equations. The chosen network is a stretch of the Monash Freeway (M1) located in the area of Melbourne, Victoria, Australia. Specifically, it is an urban motorway characterised by a traffic pattern that is strongly dependent on the demand due to commuters driving to and from the city centre. In Figure 3, an aerial map of the area is shown; the highlighted road is the M1, where the direction towards the city centre will be considered (from East to West).

The considered stretch, sketched in Figure 4, is 5.26 km long and is composed by four lanes. It starts immediately upstream of the on-ramp connected to Gully Road, it includes the on- and off-ramps connecting M1 to Blackburn Road and Forster Road, and it terminates about 1 km downstream of the off-ramp connecting Huntingdale Road. As previously mentioned, there are two traffic peaks during working days: the morning peak, during which vehicles are mostly directed towards the city centre, and the afternoon peak in which vehicles are directed towards the residential districts outside the city. Data is collected through 11 sensor-stations that deliver measurements of flow and speed (per lane) every minute; sensors

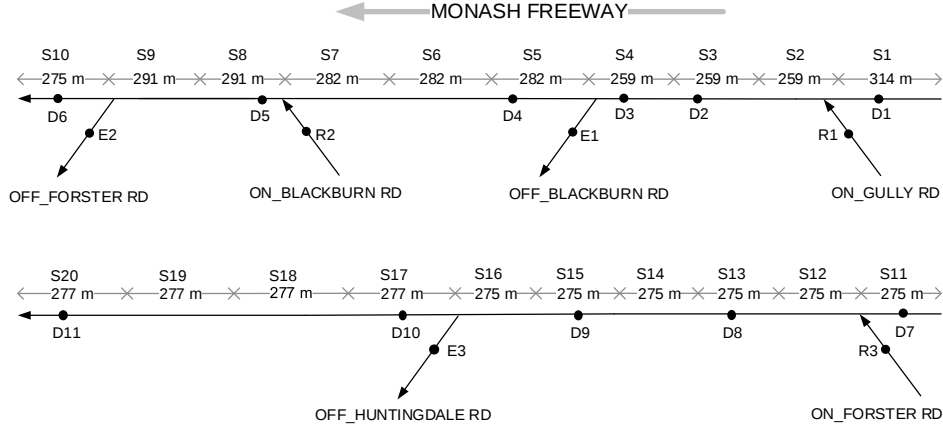


Figure 4: A graphical representation of the motorway stretch considered.

are also located at the ramps, measuring the corresponding in- and out- flows. The network is subdivided in 20 segments, with lengths ranging from 259 m to 314 m, as seen in Figure 4, that shows also the locations of the sensor-stations (black dots), on-ramps, and off-ramps. In the rest of this paper, the lanes are numbered $1, \dots, 4$ from the shoulder lane (close to the roadside), to the outer lane (close to the road median).

It was decided to take into account the morning peak, specifically from 5 AM to 9 AM; and to use the data of the 14th of August 2013 for model calibration, while morning peak data other days are reserved for model validation. Typically, major congestion is created from about 6:15 AM. The reason why a major congestion is created is quite complex; essentially, it is due to a combination of the increased demand, the high number of trucks and the slope of the motorway. As a matter of fact, in the last part of the network, the motorway slope is rising, which causes the slowing down of trucks, and, because of that, a high amount of cars leaving the shoulder lane to avoid the reduction of speed. Thus, a reduction of the capacity in the first (shoulder) lane takes place due to the slope and the high percentage of trucks, and an increase of flow in the other lanes due to escaping (lane-changing) cars. This creates a congestion that spills back and spreads all over the considered network, reaching the merging areas of on-ramps. The congestion lasts for a couple of hours, until it is resolved thanks to the reduced overall demand at the end of the peak period.

The traffic data retrieved from the detectors is adapted in order to be supplied to the

model. The Detector D1 is used for providing entering flows, modelled as on-ramps placed at the first segment. In order to obtain similar outflows at the off-ramps, the measurements E1, E2, and E3, together with the respective measurements on the mainstream D3, D5, and D9, are used to compute the turning rates $\gamma_{i,j}(k)$. No off-ramp blocking occurs in this stretch, hence (19) is not needed. Also, no capacity reduction in dependence of the lateral flows is considered in the model validation study.

3.2 Calibration results

The calibration of the model parameters to best match the measurement data has been performed with the help of an optimisation-based methodology for the validation of macroscopic traffic flow models. This methodology attempts to minimise a cost function that measures the difference between the results obtained applying the model on one hand, and the measurements taken from the detectors on the other hand, via appropriate specification of the model parameters. Because the used traffic flow model presents a strongly non-linear (and non-convex) behaviour, the chosen optimisation technique must be able to handle the high probability of being trapped in (possibly bad) local minima. In this case, the Nelder-Mead Method (Lagarias et al., 1998) has been employed for optimal parameter calibration. A detailed explanation of the applied methodology is presented by Spiliopoulou et al. (2014) and its origin goes back to the pioneering work by Cremer and Papageorgiou (1981).

The proposed model allows for different parameter values to be introduced for each segment-lane entity. This, however, would lead to a very high number of parameters to be calibrated and, most probably, to over-parametrisation; i.e. matching of details and even measurement errors which have no essential physical significance; moreover, over-parametrisation may lead to a model which seems to be very accurate in calibration with one real data set, but turns out to be much less accurate when applied to other data sets, which were not used in the optimal calibration phase. Thus, in order to reduce the number of parameters (which also simplifies the optimisation problem), segment-lanes are grouped as shown in Table 1. The displayed grouping is firstly and foremost based on the lane number; further

Lane	Segments	Detectors	ρ^{cr} (veh/km)	ρ^{jam} (veh/km)	q^{jam} (veh/h)
1	1 – 3	D1 – D2	18.57	120	1823
	4 – 7	D3 – D4	18.00	120	1500
	8 – 9	D5	18.06	120	1900
	10 – 11	D6	18.00	120	1900
	12 – 16	D7 – D9	20.01	120	1899
	17 – 20	D10 – D11	17.01	80	359
2	1 – 4	D1 – D3	20.03	120	1800
	5 – 20	D4 – D11	21.32	120	998
3	1 – 4	D1 – D3	20.00	120	1799
	5 – 20	D4 – D11	21.71	120	997
4	1 – 4	D1 – D3	24.47	120	1900
	5 – 20	D4 – D11	24.74	120	1892

Table 1: The group of segments-lanes that share the same parameters for calibration and validation.

sub-grouping was based on some observed patterns which were dependent on the network topology. Specifically, lanes 2, 3, and 4 are split in two groups, comprising segments 1-4 and 5-20, respectively. The subdivision of lane 1 into many smaller groups is mainly due to the changing of traffic pattern implied by the proximity of on- and off-ramps.

In North-American freeways, the free speeds are similar on different lanes, while on European motorways different lanes typically feature different free speeds. The presented model allows for different free speeds to be assigned to different lanes, even to different segment-lanes. For simplicity, mainly due to the subsequent usage of the validated model within an optimal control problem formulation, the free speed is assumed equal for all segment-lanes in the calibration study. Data observation at low densities suggests that a value $v_{i,j}^{free} = 105$ km/h reflects a reasonable approximation. More information on the motivation and impact of this choice are given further below. According to the segment lengths

and this free speed value, a time step, that respects the condition (3), is $T = 5$ s. The value of the jam density is set to $\rho_{i,j}^{jam} = 120$ veh/km for all the segments-lanes except for segments 17–20 lane 1, where a value $\rho_{i,j}^{jam} = 80$ veh/km is chosen in order to reflect its aforementioned peculiarity. For the computation of lateral flows, the methodology described in (5) is applied, and the coefficients $P_{i,j,\bar{j}}(k)$ are properly tuned according to some observation of the measured densities and their proportion between neighbouring lanes at specific locations. In accordance with the discussion of Section 2.3, the measurements of traffic detectors (by lane) are used to evaluate lane-changing flows, particularly at critical segment-lanes with systematically strong lane-changing activity. Lanes 2, 3, and 4 do not feature strong differences in terms of densities, apart from some underutilisation of lane 4 in case of very low flow (that has a minor impact on traffic dynamics); therefore, for the whole stretch, the corresponding coefficients are set to $P_{i,j,\bar{j}}(k) = 1$. On the other hand, between lane 1 and lane 2, the situation is more complex, since lane 1 is directly affected by the vehicles entering or exiting through on- and off-ramps; in addition, lane 2 is also indirectly affected, because mainstream vehicles may change lane in advance in order to avoid the potential slowdown caused by vehicles entering at a lower speed from an on-ramp. Thus, in order to account for these situations, a different parameter for the lane-changing computation is used; specifically, a value of $P_{i,j,\bar{j}}(k) = 1.5$ is set in the segments upstream of the on-ramps and downstream of the off-ramps. Certainly, more complex approaches can be implemented in order to compute values $P_{i,j,\bar{j}}(k)$ (e.g. computing them endogenously in dependence of the current traffic conditions), however the obtained results appear satisfactory for the purposes of the paper.

The flows entering the network are fed in the model as the on-ramp flows $r_{i,j}(k)$, while the upstream motorway boundary flows are treated as ramp flows for the corresponding segment-lanes. All these the demand profiles shown in Figure 5.

The cost function takes into account the difference in terms of both speed and longitudinal flow, allowing indirectly to capture also the lateral flow movements. More specifically, the cost function is calculated as the root-mean-square error (using the corrected sample

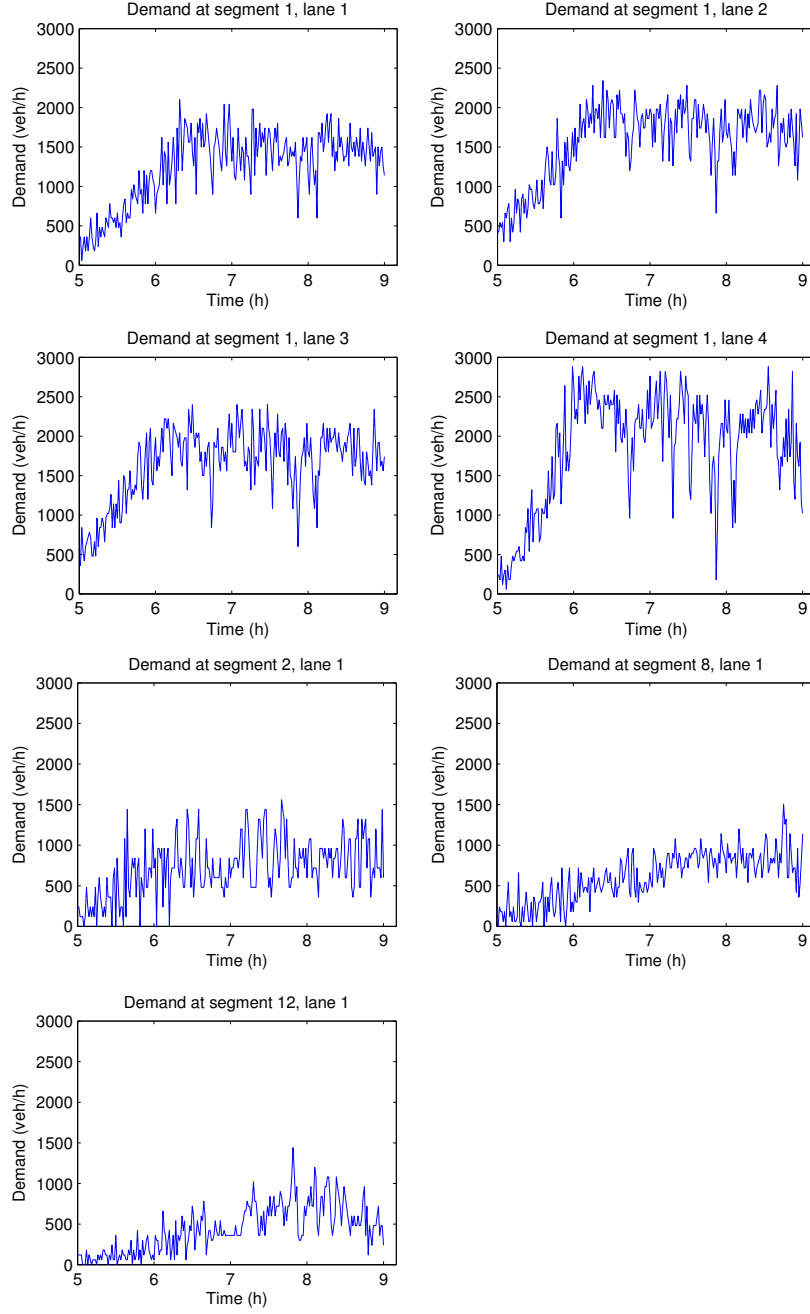


Figure 5: The demand profiles computed from detectors D1, R1, R2, and R3, that are used in the calibration phase.

variance) according to the following formula, in which $\bar{q}_j^d(k)$ and $\bar{v}_j^d(k)$ are the flow and the speed, respectively, computed by the model (the flow is calculated directly, whereas an approximation of the speed is calculated as $v_{i,j}(k) = \frac{q_{i,j}(k)}{\rho_{i,j}(k)}$) at the location of detector d ($d = 2, \dots, D$; in this case $D = 11$) and $\hat{q}_j^d(k)$ and $\hat{v}_j^d(k)$ are the corresponding values measured by the detectors:

$$\min_{\bar{q}, \bar{v}} \left\{ \alpha^q \sqrt{\frac{\sum_{k=1}^K \sum_{j=1}^J \sum_{d=2}^D [\bar{q}_j^d(k) - \hat{q}_j^d(k)]^2}{KJ(D-1) - 1}} + \alpha^v \sqrt{\frac{\sum_{k=1}^K \sum_{j=1}^J \sum_{d=2}^D [\bar{v}_j^d(k) - \hat{v}_j^d(k)]^2}{KJ(D-1) - 1}} \right\} \quad (21)$$

Weight coefficients α^q [h/veh] and α^v [h/km] are chosen in order to normalise the results; in this study, a good choice for the (normalised) weight of the speed term was found to be the double of the weight of the flow term; in this case $\alpha^q = 0.04$ h/veh and $\alpha^v = 2$ h/km. A possible useful improvement, if needed, could also be to set a different weighting factor to different parts of the network or during specific time intervals, in order to point out to the optimiser specific phenomena. The function (21) is also used as a performance index for evaluating the achieved results, particularly for the model validation.

A graphical representation of the results obtained in the model calibration may be seen in Figures 6 and 7. The tuned parameters that are obtained are given in Table 1, while the cost function and RMSE values for flow and speed are given in Table 2. Hereafter, some more detailed explanations on the real congestion pattern and the related reflection by the calibrated model are presented.

- Congestion is first created at about 6 AM because of the increase of the on-ramp flow at segment 2, and possibly because of some weaving phenomena due to the downstream exit (off-ramp at segment 4). This congestion is captured by the model (see detector 2 in Figure 7);
- As previously briefly mentioned, the main congestion starts in segment 17 (captured by Detector 10) at about 6:15 AM because of the high number of trucks reducing their velocity because of the slope of the motorway. In the calibrated model, this phenomenon is reproduced, firstly, via the relatively low calibrated $\rho_{i,j}^{cr}$ value, and

then via the decreased value $\rho_{i,j}^{jam}$ set for that specific location that causes a rapid diminishing of the outflow once the critical density is reached.

- Since lane 1 of segments 17-20 is overcrowded with trucks, most car drivers take the decision to move to the adjacent lanes; this causes an increase of density that, quite rapidly, triggers a capacity drop due to the starting of a strong congestion. While rendering the model capable of dealing with the described complex phenomenon, the lane-changing factor $P_{i,j,\bar{j}}(k)$ plays a fundamental role. In fact, for each segment-lane located in segments 18-20, lane 1, these values are set to $P_{i,j,\bar{j}}(k) = 2$ (in the direction $\bar{j} > j$) during the time period in which trucks are the predominant vehicles (until about 6:30 AM), and are decreased progressively to $P_{i,j,\bar{j}}(k) = 1$ while the traffic composition is moving back to normal. These parameters generate accordingly strong lateral flows directed from lane 1 to 2 (despite the density in lane 2 becoming higher than in lane 1), and, consequently, also to the other lanes in the direction of the road median. This generates a density increase in all the lanes and, once the density exceeds its critical value, the capacity drop is triggered. This can be observed in Figure 6, by close inspection of the flow at Detector 11 for lanes 2, 3, and 4. In fact, since no boundary conditions are set at the network exit (segment 20), the classic CTM would generate flow values equal to the capacity of the segment for all the period when $\rho \geq \rho^{cr}$. Instead, the proposed model is capable of generating a reduction of the outflow in the corresponding segments (that act as the head of the congestion), for all the congestion period, which is in excellent agreement with the observed real outflows; this is clearly visible looking at Figure 6 for Detector 11, lane 2, just after 6 AM. This reduction in the segment capacity disappears when the density is again undercritical, and the model restores the capacity flow (for Detector 11, lane 2, this happens at about 8:20 AM).
- Once the congestion in the last segments has started, it spills back, covering the whole stretch at about 6:35 AM. The back-spilling main congestion worsens further because of the increased flows entering from the on-ramps. More specifically, it may be seen that the increased ramp flow at segment 12 at about 7:30 AM contributes to strengthen

the speed reduction (this can be seen comparing the speed diagrams at Detectors D8 and D7). Another contribution to the main congestion is again due to an increased flow entering from the on-ramp R2 that increases after 7 AM, causing a further reduction of speed (see the speed diagram for Detector D5). In this situation, the model is reproducing the congestion pattern because of the proper calibration of its parameters, above all the critical density. During this period, lateral movements are present almost exclusively in proximity of on- and off-ramps, where lane-changes are assigned from and towards lane 1; therefore, at those locations, the chosen values for $P_{i,j,j+1}(k)$ are greater than one (e.g. $P_{i,1,2}(k) = 1.5$).

- Around 8 AM, the demand starts to decrease, and this causes the gradual disappearance of the congestion, restoring the maximum flow and speed. Again, the model is able to follow properly this congestion dissolution pattern.

Via careful inspection of the graphs of Figure 7, it may be noticed that, as mentioned earlier, the real speeds are sometimes different among different lanes, while, in the proposed model, the same free speed is used for all the lanes. This choice was made here to account for the ultimate intended usage of the calibrated model, i.e. as the modelling part of an optimal control problem in Part II (Roncoli et al., 2015b). Specifically, using different free speeds for different lanes may result in underutilisation of the slower lanes (in low-density situations) in an attempt of the control to decrease the vehicle travel times. Some ways to overcome this limitation are mentioned in Part II. Nevertheless, in the presented example, the difference in speed among lanes is not excessive, and the calibrated model results reproduce the main traffic flow phenomena (in terms of traffic volumes and mean speeds) with remarkable accuracy in space and time. Thus the model can be employed as a solid basis for the assessment of improvements that may be obtained via appropriate control actions.

Another deliberate simplification, which is not dictated by the model structure, but by the need to use the same model with the same parameters for traffic control testing in Part II, is that, in the present model application and validation, we do not use downstream boundary data (from detector D11). More specifically, no supply functions (according to Figure 2

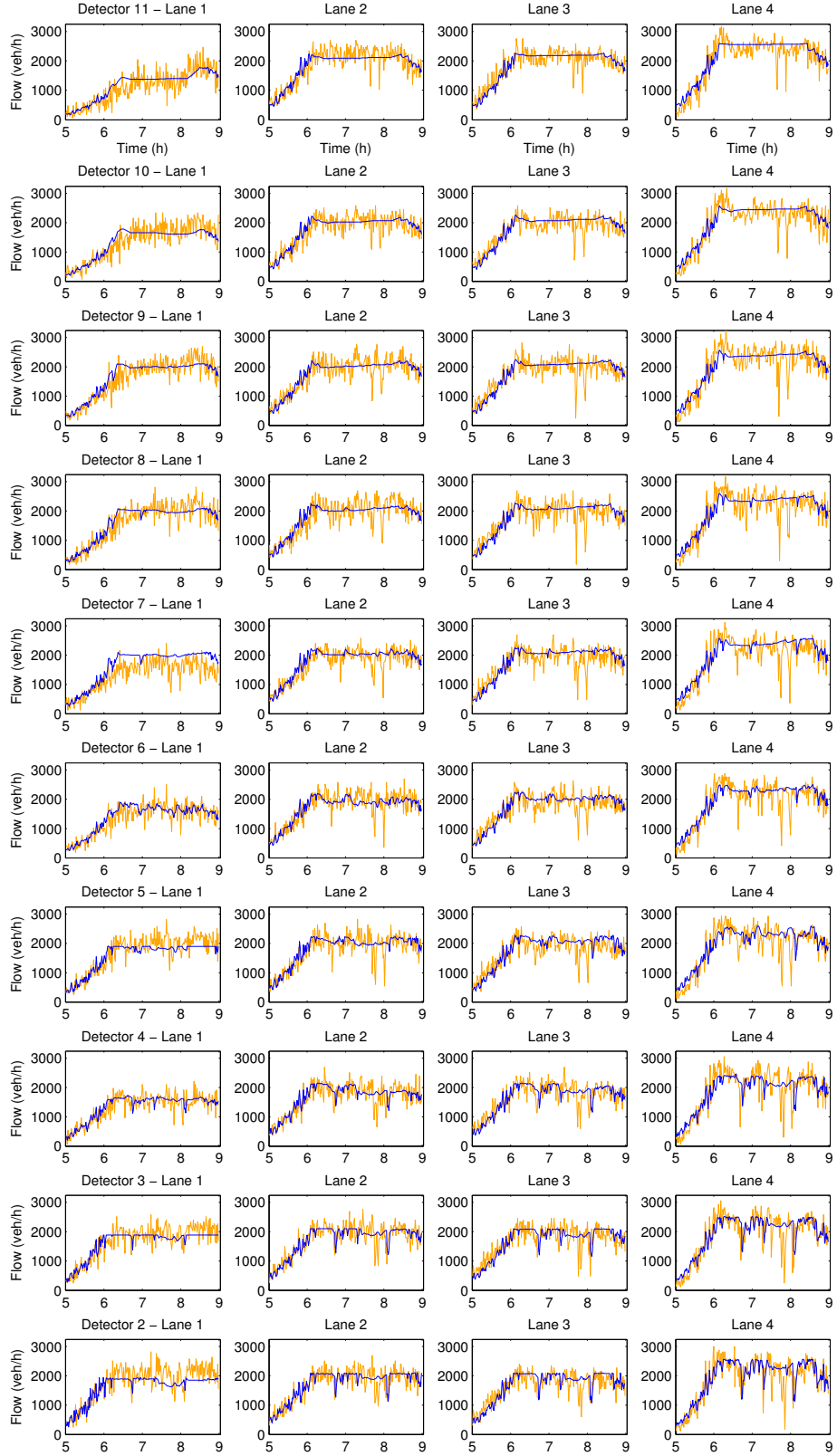


Figure 6: Comparison between the real flow data (orange/lighter line) and the calibrated model results (blue/darker line).

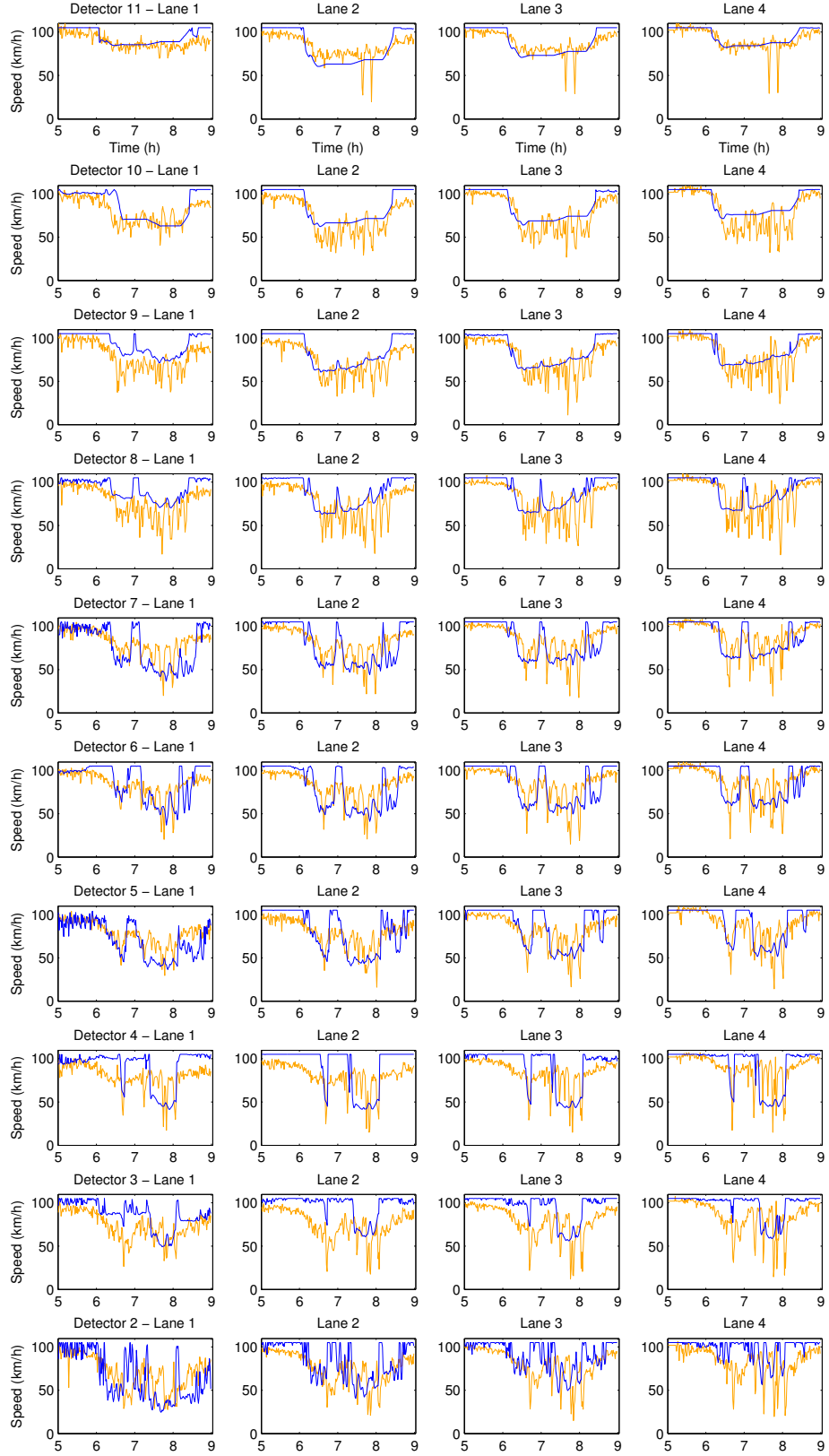


Figure 7: Comparison between the real speed data (orange/lighter line) and the calibrated model results (blue/darker line).

and (18)) are considered in the calculation of flows for the downstream-most segment-lanes of the modelled motorway stretch. As a consequence, any congestion phenomena back-spilling from downstream cannot be imported by the model, although this would improve the model results, since some few short-lived shock waves in the data are actually stemming from downstream. A calibration and validation exercise of the same model without these two simplifications (i.e. allowing for different free speeds for different lanes and using downstream boundary data) was reported in Roncoli et al. (2015a) to improve the model accuracy; see Table 2 for the corresponding improved index values; see Roncoli et al. (2015a) for the detailed diagrams and results.

Even though the model is capable of reproducing the congestion pattern, the resulting variable trajectories appear smoother than in reality. This is due to the nature of this model that, considering macroscopic variables, cannot reproduce accurately some microscopic phenomena (e.g. the shockwaves created by some slow cars changing lanes). As a final remark, no direct comparative results of lane changing flows are provided, simply because of no availability of real data. An accurate account of real lane-changing flows could be obtained, for example, by counting vehicles which change lane in each specific area of the long motorway stretch, which implies the availability of an enormous amount of data (i.e. the trajectories of all vehicles, retrieved, for instance, via video analysis). Since the only flow measurements available were longitudinal flow measurements (per lane) on the mainstream, we believe that, having a satisfactory match of flows (per lane) can be deemed as a reasonable indirect option of assessing the lane-change modelling component. In fact, we have re-run the model with the calibrated parameters, but without lane-changing. The corresponding results (not detailed here) show clearly that, without the lane-change component, the model is not capable of reproducing the traffic pattern; which highlights the pertinence of the proposed lane-changing modelling.

Day	Without downstream boundary conditions			With downstream boundary conditions		
	Cost function	Flow RMSE	Speed RMSE	Cost function	Flow RMSE	Speed RMSE
14/08	47.9	319	17.6	45.1	293	16.7
12/08	46.9	308	17.3	39.7	285	14.1
15/08	50.9	344	18.6	42.3	298	15.2
16/08	44.1	297	16.1	36.4	278	12.6

Table 2: Comparison of cost function values for different days used in the model validation. The RMSE values do not include the weights α^q and α^v , and they are expressed in [veh/h] and [km/h] respectively.

3.3 Model validation

In order to test and demonstrate the validity of the proposed model, the parameters resulting from the calibration process are applied to the same motorway stretch but for different days. Specifically, the model is fed with boundary data stemming from different (incident-free) days, and its results, obtained with the same parameter values specified in the calibration procedure, are compared with corresponding real data from stretch-internal detectors. For all days, the (recurrent) traffic behaviour is similar to the one described in the previous subsection. The main congestion starts around 6:15 AM from the segments 17-20 of the stretch; it spills back covering all the other segments; and it lasts for about 2 hours, until the reduction of demand causes its dissolution. Again, the results obtained are satisfactory, since the traffic flow model is able to reproduce correctly the traffic congestion pattern. For a quantitative comparison, the cost function (21) is used for the different days, obtaining the results shown in Table 2.

For the sake of brevity, only the demand profiles (Figure 8) and the graphical comparisons of resulting flow and speed (Figures 9 and 10) related to the 15th of August 2013 are presented here.

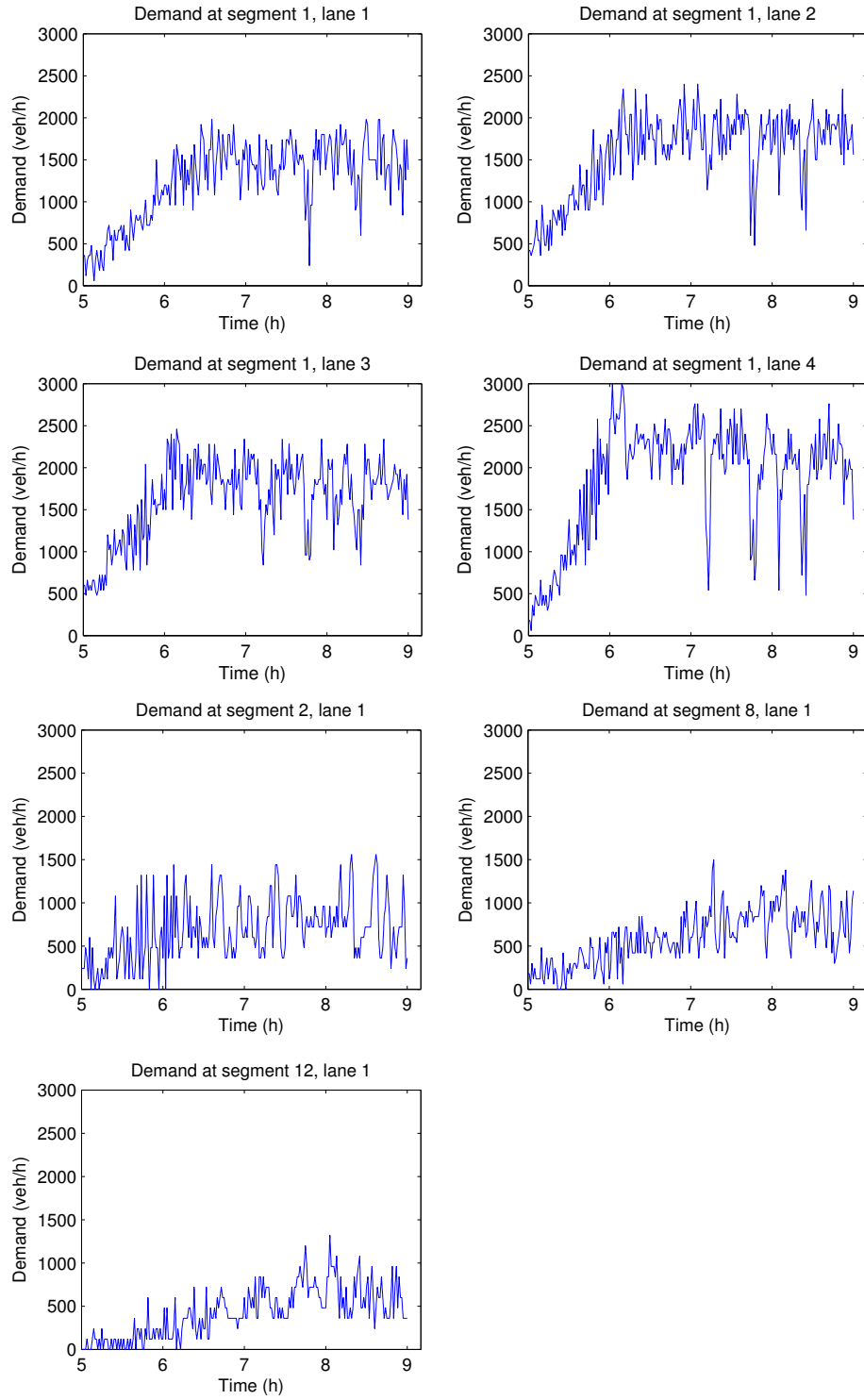


Figure 8: The demand profiles computed from detectors D1, R1, R2, and R3, that are used in the validation phase.

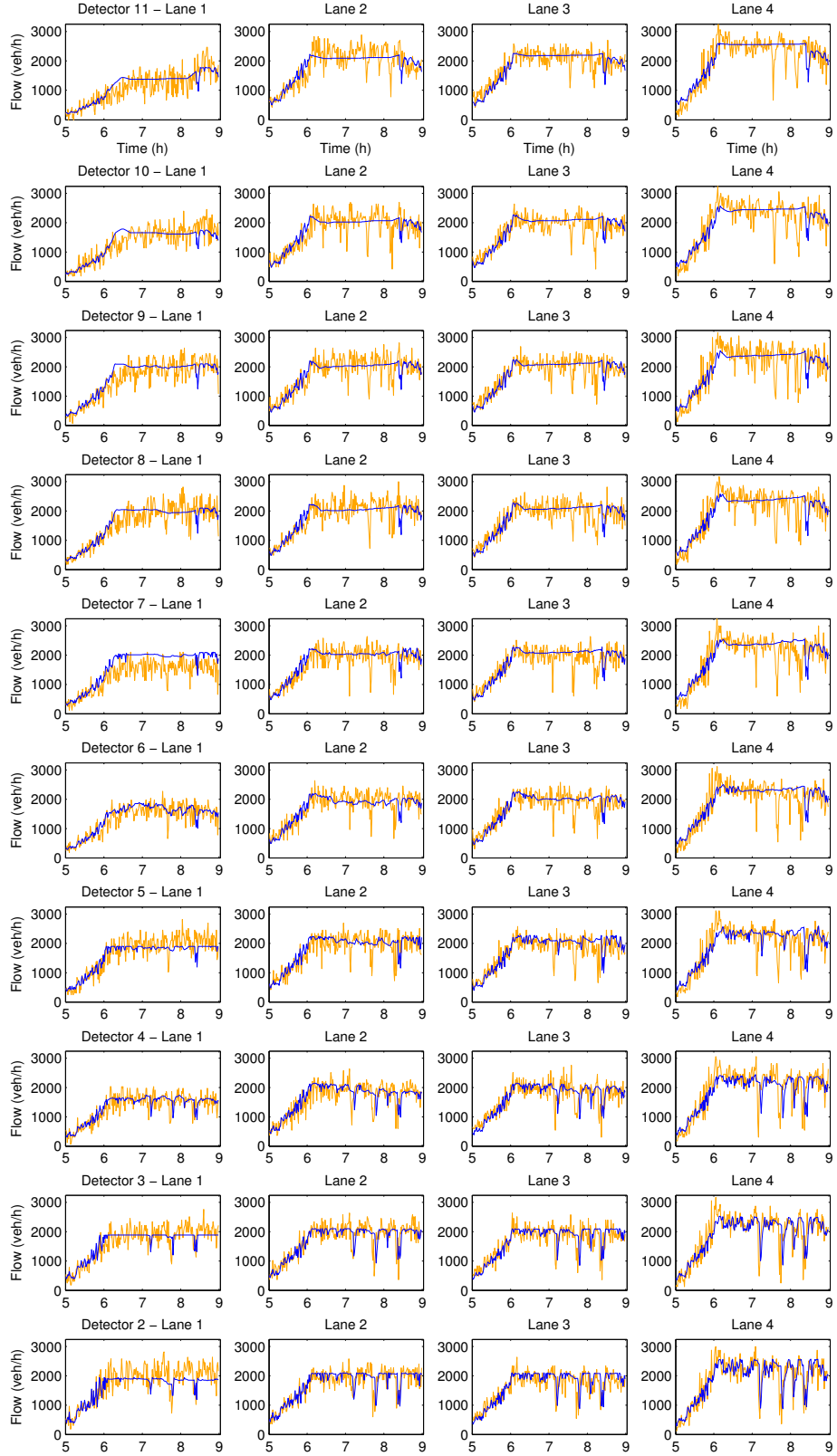


Figure 9: Comparison between the real flow data (orange/lighter line) and the validated model results (blue/darker line) for the 15/08.

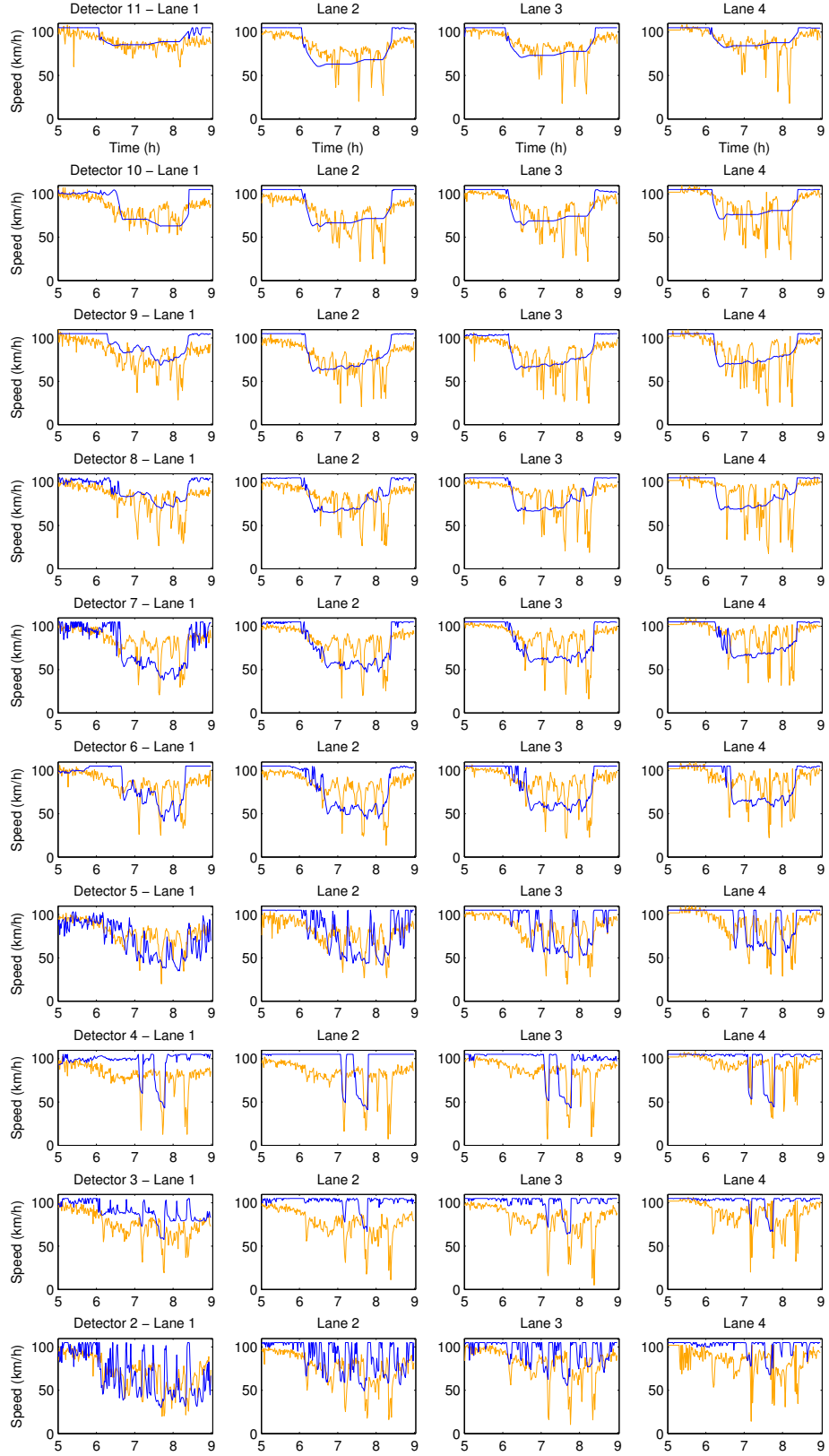


Figure 10: Comparison between the real speed data (orange/lighter line) and the validated model results (blue/darker line) for the 15/08.

4 Conclusions

This paper introduced a novel first-order multi-lane macroscopic modelling approach for motorways, with the main purpose of applying it in a model-based optimal control scheme. The model originates from the conservation equation, and methodologies for the computation of lateral and longitudinal flows are proposed; a noteworthy ingredient is the inclusion of the capacity drop phenomenon in the first-order model. In order to test and demonstrate the accuracy of the proposed model, a calibration procedure, based on real data collected by a set of detectors in an urban motorway, has been conducted. Despite the complex causes behind the occurring congestion, the model is capable to replicate the flows and speeds for the analysed scenario. Validation results, while applying the model with data from different days at the same stretch, confirm the robustness of the modelling approach.

The extension of this work, presented in Part II (Roncoli et al., 2015b), is the formulation of a linearly-constrained optimal control problem with the purpose of mitigating traffic congestion assuming that vehicles are equipped with VACS and that they are able to materialise different control actions, i.e. ramp metering, mainstream traffic flow control, and lane-changing control.

Acknowledgement

The authors would like to thank VicRoads, Melbourne (Australia), for providing the traffic data.

The research leading to these results has been conducted in the frame of the project TRAMAN21, which has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Advanced Grant Agreement n. 321132.

References

- Bar-Gera, H. and S. Ahn (2010). Empirical macroscopic evaluation of freeway merge-ratios. *Transportation Research Part C: Emerging Technologies* 18(4), 457–470.
- Bishop, R. (2005). *Intelligent vehicle technology and trends*. Artech House Publishers.
- Cassidy, M. and S. Ahn (2005). Driver turn-taking behavior in congested freeway merges. *Transportation Research Record* (1934), 140–147.
- Courant, R., K. Friedrichs, and H. Lewy (1928). Über die partiellen differenzengleichungen der mathematischen physik. *Mathematische Annalen* 100(1), 32–74.
- Cremer, M. and M. Papageorgiou (1981). Parameter identification for a traffic flow model. *Automatica* 17(6), 837–843.
- Daganzo, C. F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological* 28(4), 269–287.
- Duret, A., S. Ahn, and C. Buisson (2012). Lane flow distribution on a three-lane freeway: General features and the effects of traffic controls. *Transportation Research Part C: Emerging Technologies* 24, 157–167.
- Gazis, D. C., R. Herman, and G. H. Weiss (1962). Density oscillations between lanes of a multilane highway. *Operations Research* 10, 658–667.
- Godunov, S. K. (1959). A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb. (N.S.)* 47 (89), 271–306.
- Hall, F. L. and L. M. Hall (1990). Capacity and speed-flow analysis of the Queen Elizabeth Way in Ontario. *Transportation Research Record* 1287, 108–118.
- Helbing, D. (1997). Modeling multi-lane traffic flow with queuing effects. *Physica A: Statistical Mechanics and its Applications* 242(1-2), 175–194.

- Hoogendoorn, S. and P. Bovy (1999). Gas-Kinetic Model for Multilane Heterogeneous Traffic Flow. *Transportation Research Record* 1678(1), 150–159.
- Jin, W.-L. (2013). A multi-commodity Lighthill-Whitham-Richards model of lane-changing traffic flow. *Procedia - Social and Behavioral Sciences* 80, 658–677.
- Kesting, A., M. Treiber, M. Schönhof, and D. Helbing (2008). Adaptive cruise control design for active congestion avoidance. *Transportation Research Part C: Emerging Technologies* 16(6), 668–683.
- Knoop, V. L., A. Duret, C. Buisson, and B. van Arem (2010). Lane distribution of traffic near merging zones influence of variable speed limits. In *13th International IEEE Conference on Intelligent Transportation Systems*, pp. 485–490.
- Knoop, V. L., S. P. Hoogendoorn, Y. Shiomi, and C. Buisson (2012). Quantifying the Number of Lane Changes in Traffic. *Transportation Research Record: Journal of the Transportation Research Board* 2278, 31–41.
- Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization* 9(1), 112–147.
- Laval, J. A. and C. F. Daganzo (2006). Lane-changing in traffic streams. *Transportation Research Part B: Methodological* 40(3), 251–264.
- Lebacque, J. (2003). Two-phase bounded-acceleration traffic flow model: analytical solutions and applications. *Transportation Research Record* 1852(1), 220–230.
- Leclercq, L., J. A. Laval, and N. Chiabaut (2011). Capacity drops at merges: An endogenous model. *Transportation Research Part B: Methodological* 45(9), 1302–1313.
- Lighthill, M. J. and G. B. Whitham (1955). On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 229(1178), 317–345.

- Michalopoulos, P. G., D. E. Beskos, and Y. Yamauchi (1984). Multilane traffic flow dynamics: Some macroscopic considerations. *Transportation Research Part B: Methodological* 18(4-5), 377–395.
- Monamy, T., H. Haj-Salem, and J.-P. Lebacque (2012). A Macroscopic Node Model Related to Capacity Drop. *Procedia - Social and Behavioral Sciences* 54, 1388–1396.
- Munjal, P. and L. Pipes (1971). Propagation of on-ramp density perturbations on unidirectional two- and three-lane freeways. *Transportation Research* 5(4), 241–255.
- Papageorgiou, M., J.-M. Blosseville, and H. Hadj-Salem (1990). Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part I: Modelling. *Transportation Research Part A: General* 24(5), 345–359.
- Papageorgiou, M. and A. Messmer (1990). METANET: A macroscopic simulation program for motorway networks. *Traffic Engineering & Control* 31(9), 466–470.
- Papageorgiou, M., I. Papamichail, A. Spiliopoulou, and A. Lentzakis (2008). Real-time merging traffic control with applications to toll plaza and work zone management. *Transportation Research Part C: Emerging Technologies* 16(5), 535–553.
- Rao, B. and P. Varaiya (1994). Roadside intelligence for flow control in an intelligent vehicle and highway system. *Transportation Research Part C: Emerging Technologies* 2(1), 49–72.
- Roncoli, C., M. Papageorgiou, and I. Papamichail (2015a). An optimisation-oriented first-order multi-lane model for motorway traffic. In *Proceedings of the 94th Annual Meeting of the Transportation Research Board (TRB)*, Washington, D.C., USA.
- Roncoli, C., M. Papageorgiou, and I. Papamichail (2015b). Traffic flow optimisation in presence of vehicle automation and communication systems - Part II: Optimal control for multi-lane motorways. *Submitted to: Transportation Research Part C: Emerging Technologies*.

- Spiliopoulou, A., M. Kontorinaki, M. Papageorgiou, and P. Kopelias (2014). Macroscopic traffic flow model validation at congested freeway off-ramp areas. *Transportation Research Part C: Emerging Technologies* 41, 18–29.
- Srivastava, A. and N. Geroliminis (2013). Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model. *Transportation Research Part C: Emerging Technologies* 30, 161–177.
- Treiber, M. and A. Kesting (2013). *Traffic flow dynamics*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Treiber, M., A. Kesting, and D. Helbing (2006). Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps. *Phys. Rev. E* 74, 016123.